Local economies amidst the COVID-19 crisis in Italy: a tale of diverging trajectories

CERQUA, AUGUSTO and LETTA, MARCO

Sapienza University of Rome

26 November 2020

Local economies amidst the COVID-19 crisis in Italy: a tale of diverging trajectories

Augusto Cerqua and Marco Letta

[a] Department of Social Sciences and Economics, Sapienza University of Rome

This version: November 26, 2020

Abstract: *Impact evaluations of the microeconomic effects of the COVID-19 upheavals are essential but nonetheless highly challenging. Data scarcity and identification issues due to the ubiquitous nature of the exogenous shock account for the current dearth of counterfactual studies. To fill this gap, we combine up-to-date quarterly local labor markets (LLMs) data, collected from the Business Register kept by the Italian Chamber of Commerce, with the machine learning control method for counterfactual building. This allows us to shed light on the pandemic impact on the local economic dynamics of one of the hardest-hit countries, Italy. We document that the shock has already caused a moderate drop in employment and firm exit and an abrupt decrease in firm entry at the country level. More importantly, these effects have been dramatically uneven across the Italian territory and spatially uncorrelated with the epidemiological pattern of the first wave. We then use the estimated individual treatment effects to investigate the main predictors of such unbalanced patterns, finding that the heterogeneity of impacts is primarily associated with interactions among the exposure of economic activities to high social aggregation risks and pre-existing labor market fragilities. These results call for immediate place- and sector-based policy responses.*

JEL Codes: C53; D22, E24; R12

Keywords: impact evaluation; counterfactual approach; machine learning; local labor markets; firms; COVID-19; Italy

E-mail addresses: augusto.cerqua@uniroma1.it (A. Cerqua), marco.letta@uniroma1.it (M. Letta)

1. Introduction

With over 52,000 deaths and almost 1,500,000 cases (as of November 26, 2020), Italy ranks among the worst-hit countries by COVID-19.[1] The Italian government was the first in Europe to declare, on March 9, an unprecedented national lockdown that paralyzed the country. From March 25, productive activities were shut down, except for those deemed 'essential' for the functioning of the country's economic system. On May 4, lockdown rules started to be lifted, and, from June 15, almost all economic activities were finally allowed to re-open, albeit under strict safety protocols. The suspension of restrictive measures continued throughout the summer until the impressive resurgence of the contagion in the fall of 2020 forced the government and regional authorities to issue new social distancing policies, including the reintroduction of restrictive measures targeted to economic activities. While we write, Italy has adopted a place-based approach to contain the second wave, with different levels of restrictions and workplace closures, in line with the pandemic's heterogeneous evolution in its territory.

The repercussions of this remarkable series of disruptive events on the Italian economy are enormous, and the Italian government tried to attenuate these impacts via the adoption of several emergency measures and fiscal packages.[2] In order to increase workers' protection, the government also issued an *ad hoc* Decree-Law on March 17, which introduced two labor market policies: a special COVID-19 short-time work retroactive compensation scheme (of the duration of 9 weeks) and a firing freeze that stopped firings[3] (Casarico & Lattanzio, 2020). The firing freeze measure has been extended several times, and it is currently in effect until March 2021.

Despite the implementation of a wide range of policy interventions, annual forecasts by the Bank of Italy (July 2020) pointed to a 9.5% GDP fall, a reduction of 11.8% in the number of hours worked and a decrease of 4.5% in the number of persons employed.[4] More recent estimates (October 2020) by the International Monetary Fund (IMF) suggest an even larger annual GDP drop at 10.6%.[5] These projections might be optimistic, given that they do not take fully into account the adverse effects of the ongoing second wave.

[1] See https://www.worldometers.info/coronavirus/country/italy/.

[2] For a database of fiscal policy responses to COVID-19 in Italy (as well as many other countries), please refer to https://www.imf.org/en/Topics/imf-and-covid19/Fiscal-Policies-Database-in-Response-to-COVID-19.

[3] The firing freeze could also be applied retroactively to firings pending from February 23.

[4] https://www.bancaditalia.it/pubblicazioni/proiezioni-macroeconomiche/2020/en-estratto-boleco-3-2020.pdf?language_id=1.

[5] https://www.imf.org/en/Publications/WEO/Issues/2020/09/30/world-economic-outlook-october-2020.

However, credible *ex-post* quantifications of these effects, especially empirical evidence on microeconomic and local impacts, are still missing. Such an empirical vacuum is hardly surprising as real-time microdata is scarcely available. On top of data scarcity, rigorous evaluation of the crisis effects is challenging because of econometric issues: the COVID-19 exogenous shock virtually left no part of the world unaffected. In econometric jargon, this means that a control group is hardly available because the treatment affected all units simultaneously or with short lags.[6] As noted by Chudik et al. (2020), this implies that in most cases, standard evaluation techniques, such as difference-in-difference or the synthetic control method (SCM), are not applicable.[7] This is probably the reason why, despite the micro literature on the pandemic is flourishing (Adams-Prassl et al., 2020; Baker et al., 2020; Bartik et al., 2020; Benedetti et al., 2020; Bick & Blandin, 2020; Blundell et al., 2020; Buchheim et al., 2020; Cajner et al., 2020; Carvalho et al., 2020; Forsythe et al., 2020; Gourinchas et al., 2020; Von Gaudecker et al., 2020), almost all these policy-relevant works are not based on counterfactual impact evaluation methodologies. A notable exception is the study by Chetty et al. (2020), who employ private real-time anonymized data and an evaluation strategy which exploits between-state heterogeneity in the reopening's timing to document the granular impact of the pandemic and the related policy responses on various economic outcomes in the United States.

Concerning Italy, Ascani et al. (2020) provide evidence of a close relationship between COVID-19 disease patterns and local economies' characteristics. Casarico and Lattanzio (2020) focus on how different categories of workers were affected by the pandemic in the short-term and carry out a first evaluation of the policy responses implemented. Using a linear probability model, they find that workers already in disadvantaged conditions before the shock (young, low-skilled, and seasonal workers) exhibit substantially higher risks of losing their jobs.

These studies underline important local and sectoral components of the impacts of the crisis in Italy.

---

[6] There are some exceptions: in countries and areas where no total lockdowns were implemented, one might exploit staggered or heterogenous policy responses to generate a counterfactual scenario (see the study by Chetty et al. (2020) mentioned below). This is not the case of Italy. Yet, one could argue that since the spread of the contagion, especially in the first wave, was highly heterogeneous and predominantly affected Northern Italian regions, it would be possible to use the Southern regions as control group or to consider the shock as 'continuous' treatment with different intensity levels. However, we disagree with the premise. The national lockdown implemented during the first wave, and the shutdown of entire sectors, involved the entire country.

[7] To make up for this, Chudik et al. (2020) develop a cross-country econometric model in which the Covid-19 shock is identified using the IMF's GDP growth forecast revisions between January and April 2020, under the assumption that Covid-19 was the main driver of these forecast revisions. In this way, they use the difference in the forecasts as a counterfactual strategy to quantify the economic impact of the shock.

Indeed, in Europe as elsewhere, the current crisis is undoubtedly a regional one, because the economic impacts are unfolding unevenly at the local level, so regional perspectives are essential to understand the unequal impacts of the pandemic (Bailey et al., 2020). At least in the Italian context, however, we are not aware of any paper showing *ex-post* counterfactual evidence on the local microeconomic effects of the COVID-19 disruption on labor and firm outcomes.

This article quantifies the heterogeneous impacts of COVID-19 on employment and business demography for all 610 Italian local labor markets (LLMs) and investigates the main territorial features of such unevenness. To this end, we leverage up-to-date quarterly LLMs data, collected from the Business Register kept by the Italian Chamber of Commerce, combined with the newly developed machine learning control method (MLCM). MLCM draws on the predictive ability of machine learning (ML) algorithms to generate a no-COVID counterfactual scenario (i.e. a 'business-as-usual' scenario) in such a peculiar econometric setting. The use of the MLCM is made possible by constructing a comprehensive time-series cross-sectional database on LLMs.

Thanks to this counterfactual approach, we document that at the end of the third quarter of 2020, the shock has not only already caused a steep decrease in firm entry and a moderate drop in employment and firm exit at the aggregate level but, more importantly, that the effects have been markedly heterogeneous across the Italian territory. In the following step, we collect data encompassing economic, mobility, and pandemic-related features of each LLM and link them to the detected geographic diversity of COVID-19 employment impacts by identifying the features that matter the most in explaining the heterogeneity of the outcome variable, i.e. the estimated treatment effect of employment change. Regression trees suggest that effect sizes stem from a series of exposure of economic activities to high social aggregation risks and pre-existing labor market fragilities. These findings have self-evident implications that call for immediate place- and sector-based policy responses.

The remaining of this paper proceeds as follows. Section 2 describes the data. Section 3 introduces the econometric methodologies. Section 4 reports the treatment effects resulting from the counterfactual analysis, while the subsequent section investigates the main predictors of the estimated impacts. Section 6 concludes.

2. Data

Our primary dependent variable is the log of overall employment. In addition, we also split employment between manufacturing and services, and investigate the impact of COVID-19 on the number of new business registrations (births) and cessations of trading (deaths). All these variables come from the Business Register kept by the Union of the Italian Chamber of Commerce (*Unioncamere*). The Business Register is based on administrative data on the Italian companies gathered by the provincial Chambers of Commerce. It contains information on the registration data of the universe of Italian private non-financial sector firms. The Business Register quarterly data on local employment are made available by the Italian Social Security Institute (INPS) since the third trimester of 2014.

To estimate the impact of COVID-19 on each LLM, we build a comprehensive, balanced panel of all 610 Italian LLMs from 2014 Q3 to 2020 Q3 and employ the random forest algorithm described in Section 3. The counterfactual is estimated by controlling for the industrial structure of each LLM. To this end, we exploit the classification by the Italian National Institute of Statistics (Istat), which splits the Italian LLMs into four classes: without specialization, non-manufacturing, made in Italy,[8] and other manufacturing. Furthermore, in light of the expected plunge in tourism-related employment, we split the non-manufacturing class into touristic and non-touristic. We then control for LLM size, geographical dummies, population density, unemployment rate, activity rate, yearly and quarterly fixed effects, and trends in employment, business births, and business deaths. For each of the latter three variables, we control for two lags of the same quarter, the lags of the four preceding quarters, and four lags of the yearly averages. The total number of features included in the counterfactual analysis is 54.

In the second phase of the empirical analysis, the association analysis uses the estimated COVID-19 impact on employment for all LLMs as the outcome of interest to uncover its primary predictors. For this analysis, we collected several variables potentially correlated with the employment change due to COVID-19. We use the dependency ratio to control for the population structure and its implications for the productive part of the population. As a measure of the spread of COVID-19, we use the excess mortality estimates provided by Cerqua et al. (2020), updated to 31 August 2020.[9] We also employ two variables which capture the criticality of the tasks performed by employees, the possibility of exposure to the virus and physical proximity to the workplace, all highlighted as relevant factors in the literature (see Barbieri et al., 2020): the share of jobs having a high risk of social aggregation and the share of jobs having a high 'integrated' risk. These variables were created on the basis of the work

[8] The 'made in Italy' manufacturing LLMs are characterized by small firms organized in industrial districts (Cainelli et al., 2013).

[9] These data are publicly available here: https://www.stimecomunalicovid19.com/.

conducted by an *ad hoc* task force,[10] which linked to each economic sector (2-digit NACE Rev.2 classification) a level of social aggregation and integrated risks from low to high.

We then build the share of short-term contracts as a metric for temporary jobs' local relative importance.[11] Additionally, in March 2020, the Italian government was forced to suspend many economic activities considered 'non-essential'. The selection of these activities was carried out on the basis of the NACE Rev.2 classification. We made use of this information to generate the share of jobs in suspended economic activities.

Other economic variables included in this phase of the analysis are income per capita, unemployment rate, Istat's economic classification (described above), the share of innovative start-ups as a proxy for local innovation, and a measure of economic fragility, i.e. the share of firms having employees in *Cassa Integrazione Guadagni Straordinaria* (CIGS), namely the most utilized Italian short-time work program providing subsidies for temporary reductions in the number of hours worked.[12]

Lastly, as mobility is one of the critical aspects linked to the epidemiological spread of COVID-19, we take it into account by using three variables:

- the number of road accidents per 10,000 inhabitants;
- the share of population living in peripheral areas;
- the index of relational intensity within the LLM (IIRFL). The higher the IIRFL, the greater the inter-municipal turbulence in terms of flows.

Table A1 in the Appendix includes a more detailed description of all the variables, while Table A2 in the Appendix provides descriptive statistics. The availability of these indicators (18, in total) will allow us to identify the LLM characteristics that matter the most in explaining the treatment effects' heterogeneity.

---

[10] In April 2020, Italy's Prime Minister Giuseppe Conte appointed Vittorio Colao, former Vodafone Group CEO, to lead a group of lawyers, economists, and experts to outline a plan on how to restart the Italian economy after the coronavirus emergency. One of the group's objectives was to reschedule the gradual reopening of economic activities based on two criteria: the risk of social aggregation and the 'integrated' risk.

[11] Even if this variable refers to 2015, we argue that this is a valid proxy for 2020, as there is evidence of a strong temporal persistence in the variation of this variable across locations (Caselli et al., 2020).

[12] CIGS targets firms experiencing economic shocks, broadly defined: it can be a demand or revenue shock, a company crisis, a need for restructuring or reorganization, a liquidity or insolvency issue, etc. CIGS is a subsidy for partial or full-time hour reductions, replacing approximately 80% of the worker's earnings due to hours not worked, up to a cap (Giupponi & Landais, 2020).

3. Methods

Our empirical exercise consists of two tasks: a counterfactual analysis and an association analysis. For both steps, we harness ML's predictive power, but with a key difference: in the counterfactual analysis, the ultimate aim is causal inference; when looking at impact predictors, instead, we tackle a purely predictive problem. The choice of the algorithm employed in each phase is in line with the different goals of the two analyses: the trade-off between accuracy and interpretability (Hastie et al., 2009; Murdoch et al., 2019) is solved in favor of the former in the counterfactual analysis, and of the latter in the association analysis. Below we separately discuss the two methodologies and their different purposes and empirical frameworks.

3.1 Counterfactual analysis: the machine learning control method

To tackle the econometric challenges related to the pandemic shock's pervasive nature, we draw on the newly developed MLCM to generate a counterfactual scenario in which the COVID-19 crisis never hit Italy. In other words, we employ the MLCM to address the fundamental problem of causal inference, i.e. the impossibility to observe the potential outcome in the no-treatment scenario, a curse that affects all LLMs.

Although ML algorithms primarily deal with out-of-sample predictions or 'prediction policy problems' (see Kleinberg et al., 2015), more recently, they have been combined with causal inference approaches (Athey & Imbens, 2016; Athey et al., 2017; Athey et al., 2019; Belloni et al., 2017; Varian, 2016; Wager & Athey, 2018). Varian (2016) was among the first to note that counterfactual building is essentially a predictive task, which is exactly the task at which ML excels. In a panel or time series setting, he noted that one could exploit pre-treatment observations to generate an artificial control group that acts as a counterfactual in the no-treatment, 'business-as-usual' scenario. This way, one could readily retrieve treatment effects as the difference between the observed outcome and the ML-generated potential outcome. Varian called this straightforward counterfactual method the 'train-test-treat-compare' process. This process is similar to the SCM developed by Abadie et al. (2010), with the key difference that it does not require the availability of untreated units, as it draws on pre-treatment information to generate a credible estimate of the 'outcome for the treated if not treated'.

Early empirical applications of this intuitive methodology for counterfactual building have recently appeared (Abrell et al., 2019; Benatia, 2020; Benatia and de Villemeur, 2020; Bijnens et al., 2019; Burlig et al., 2020; Cerqua et al., 2020; Souza, 2019). Except Burlig et al. (2020) and Souza (2019), all the other studies cannot rely on an original control group in their research design because they only observe treated units in settings with simultaneous treatment, just as in our case.

Benatia (2020) and Cerqua et al. (2020) are the most closely related to this study because they both investigate the causal effects of the COVID-19 crisis. Benatia (2020) applies a neural network model to study the impact of containment measures on the demand reduction in New York's electricity markets; Cerqua et al. (2020) employ three different ML routines (LASSO, random forest, and stochastic gradient boosting) to derive municipality-level excess mortality estimates during the COVID-19 pandemic in Italy.

In the spirit of this nascent evaluation approach, we apply the MLCM to pursue our causal inference analysis of COVID-19 local economic impacts in Italy. Our artificial control group comes from a ML predictive model developed to forecast a post-treatment counterfactual for each LLM. In this way, under the crucial assumption of stable trends in the absence of the shock, we can assess the LLM-specific causal impact of the exogenous shock by comparing the observed post-shock trajectory with the most credible trajectory the LLM unit would have followed in a no-shock scenario. A critical requirement for this approach's validity is that the predictive ML model must not include predictors that may be affected by the treatment (Varian, 2016). We avert this issue by employing only pre-2020 features in our counterfactual building. Finally, the use of the MLCM is made possible from the construction of a comprehensive time-series cross-sectional database on LLMs (see Section 2).

We apply a powerful and popular ML algorithm: the random forest.[13] The random forest is a fully non-linear technique based on the aggregation of many decision trees. In particular, random forest builds many trees (1000, in our case) based on bootstrapped training samples and, at each split of a tree, uses only a random subset of the predictors as split candidates, thus introducing a double layer of decorrelation of the trees from one another (Hastie et al., 2009).

Drawing from the routine already implemented by Cerqua et al. (2020), our counterfactual analysis is based, for each outcome variable, on the following 7-step methodological sequence:

1) We randomly split the pre-2019 quarterly dataset (2016 Q3 - 2018 Q4 for employment; 2015 Q1 - 2018 Q4 for firm outcomes) into a training sample, made up of 80% of the LLMs, and a test set, consisting of the remaining 20%;[14]
2) We train our random forest algorithm on the training set and perform 10-fold cross-validation to

[13] We also tested another well-known ML technique's predictive ability, the least absolute shrinkage and selection operator (LASSO), but it was outperformed by random forest in all specifications.

[14] We apply the random splitting of the sample at the LLM level, *not* on *LLM-year* pairs so that there is no data leakage, i.e. the same LLM only appears either in the training or the testing set.

select the best-performing tuning hyperparameter;[15]

3) We test the out-of-sample predictive performance on the corresponding pre-2019 testing sample;
4) We test model accuracy on the entire 2019 sample and compare its performance with the pre-/post- comparison method, which has become a common and intuitive metric to gauge the magnitude of the impact of the pandemic;[16]
5) We repeat the same routine on the entire pre-2020 dataset (2016 Q3 - 2019 Q4 for employment; 2015 Q1 - 2019 Q4 for firm outcomes) and finally predict, for the first three quarters of the 2020 sample, employment levels, business births, and business deaths in a 'no-COVID' ('business-as-usual') scenario;
6) We derive individual treatment effects for all LLMs as the difference between the observed 2020 outcomes and the ML-generated potential outcomes;
7) We map the individual treatment effects of the LLM-level economic impacts of COVID-19.

The critical assumption behind this MLCM routine is that the difference between our observed and counterfactual economic outcomes is due to the impact of the COVID-19 crisis. While this is not a trivial assumption, we deem it plausible given the unprecedented disruption to the economy caused by the sudden arrival of the pandemic. Finally, please note that we include lockdowns, workplace closures, and all the social distancing policies adopted to contain the spread of the contagion in our definition of shock, and not just the epidemiological spread of the virus *per se*, in line with our goal of capturing the *economic* developments of the pandemic.

3.2 Association analysis: the employment change regression tree

To estimate the relationship between the estimated employment outcomes and potentially relevant covariates linked to economic, mobility, and pandemic-related LLM features, we harness the efficacy and power of another well-known ML algorithm: the regression tree.

First and foremost, bear in mind that here we abandon the causal inference setting to go back to the

[15] We use cross-validation on the training sample to solve the bias-variance trade-off (Hastie et al., 2009) by selecting the best-performing values of the key tuning parameter *m*, i.e. the number of features randomly sampled as candidates at each split, for which we use, as alternative candidates, *p/2, p/3*, and *p/6*.

[16] In the Italian context, see, for example, Casarico and Lattanzio (2020), as well as here: https://www.lavoce.info/archives/68205/cosi-il-coronavirus-ha-contagiato-limprenditorialita/ (in Italian) and here: https://www.bancaditalia.it/media/notizie/2020/Nota-Covid-19.11.2020.pdf?language_id=1 (in Italian) for intuitive comparisons between 2020 observed data and past trends or averages in the previous year(s).

original ML habitat, i.e. the realm of pure prediction. What we want to do in this analysis is to get an idea of the factors which matter the most in predicting the heterogeneous local economic impact of the pandemic.

Regression trees are an ideal tool to fulfill this purpose for two reasons: i) differently from complex, black-box ML methods such as random forest, regression trees allow an intuitive understanding of the mechanism through which the outcome variable of interest is linked to its most relevant predictors, thus producing an easy-to-interpret output which can be particularly valuable when the model must be shared to support public decision-making (Andini et al., 2018; Lantz, 2019); ii) regression trees are extremely flexible methods that can easily capture, in the sequence of splits, the entire range of potential non-linearities and interactions between the features, without imposing any parametric functional form to the underlying data-generating process.

From a technical point of view, this ML algorithm divides the data into progressively smaller subsets to identify significant patterns that are then used to predict the continuous output. Compared to standard regression tree analyses, two necessary clarifications are in order. First, we do not divide our sample into a training and testing set. The reason is straightforward: instead of testing for the out-of-sample accuracy of our regression tree model, we want to investigate the main predictors of our outcome variable, i.e. the estimated treatment effect for employment change in 2020 Q3. Operationally, this means that, in this case, we perform an *in-sample* predictive exercise on the full sample of Italy's LLMs. Second, and related, we do not apply cross-validation to select the hyperparameter of the regression tree method (named 'complexity parameter', *cp*).

Therefore, we run a basic regression tree model of the employment effects to uncover the most relevant predictors of treatment effect unevenness at the local level. Notably, the associations emerging from the regression tree should not be interpreted in a causal sense but rather as a way to uncover significant correlations between the most important features and the outcome variable of interest.

4. Counterfactual analysis

We begin by reporting in Table 1 the random forest technique's predictive performance compared to the intuitive pre-/post- comparative method often adopted to gauge the magnitude of the COVID-19 shock.

As signaled by the Mean Squared Error (MSE) and Median Squared Error (MEDSE) of the various methods, random forest predictions substantially outperform the intuitive methodology in the out-of-sample predictive test on the 2019 sample. Using MSE as the reference metric, the predictive gain of the random forest performance is of more than 26% compared to last year's figures, and of 77% compared to the three-year (2016-2018) average of the outcome variable, two of the most commonly adopted metrics to gauge the impact of the pandemic. MEDSE performances are even more dramatically unbalanced in favor of the random forest. This test demonstrates that data-driven methodologies lead to far more accurate predictions of potential outcomes in a given, 'ordinary' year.

Table 1 – Predictive performances for 2019 (log) overall employment levels

Predictive method	MSE	MEDSE
Corresponding quarter – Last year (2018)	0.0011209	0.0005058
Corresponding quarter – 3-year average (2016-2018)	0.0036044	0.0024622
Random forest	0.0008268	0.0001938

Notes: Estimates on the 2019 full LLM sample (2440 observations; 610 per quarter). MSE stands for Mean Squared Error; MEDSE for Median Squared Error.

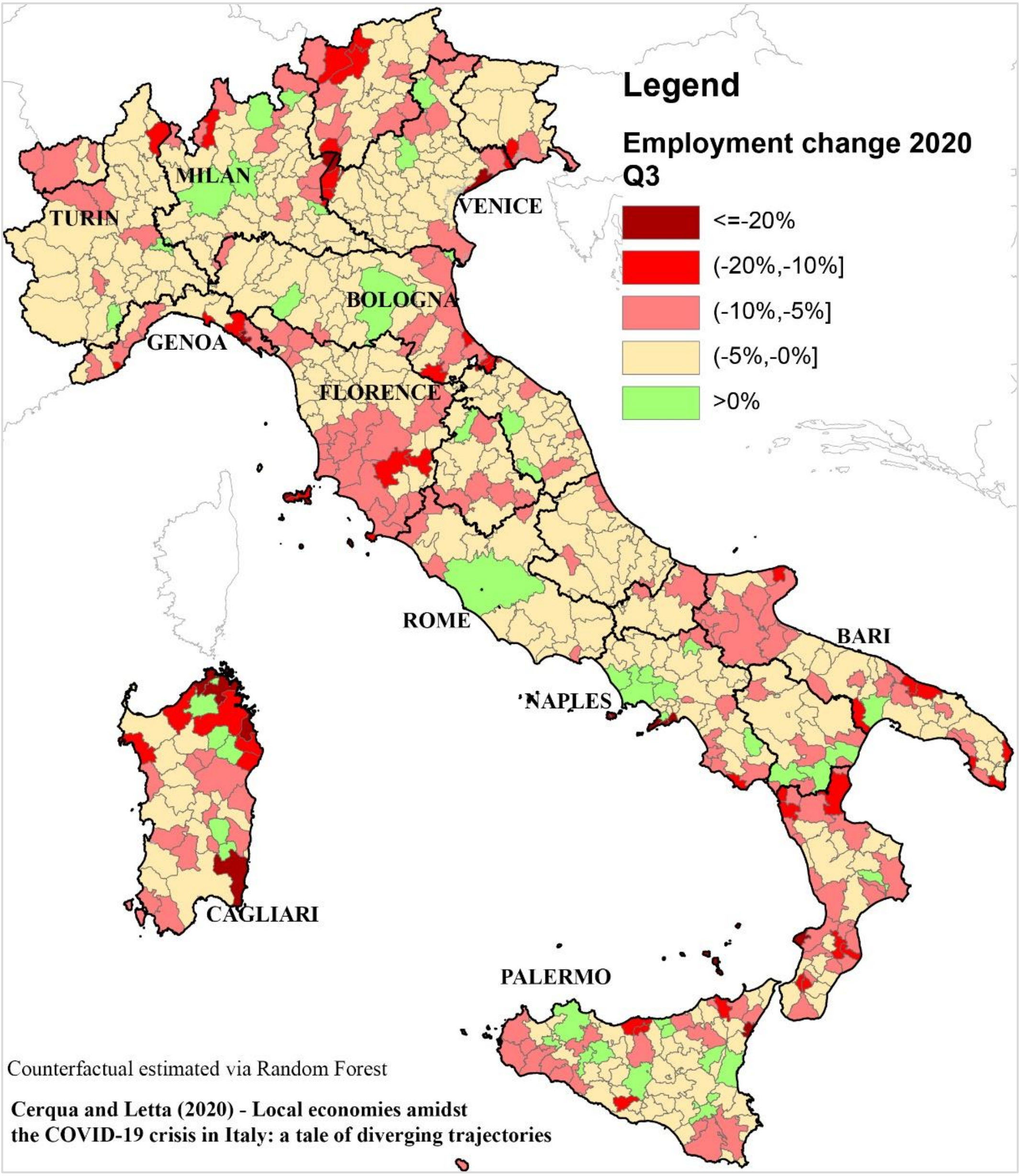
Having established that ML algorithms exploit past information to predict future outcomes much better than descriptive methods, we take a quick look at the aggregate treatment effects of the coronavirus crisis for the employment outcome. By the end of the third quarter of 2020, the pandemic has entailed a 1.86 % decrease in overall employment in Italy, compared to what employment levels would have been had the pandemic never reached the country.

As we mainly focus on the local heterogeneous impact of COVID-19, in the following sections, we first map LLM-specific treatment effects and then gauge the heterogeneity in COVID-19 impacts across local economies.

4.1 Employment

Figure 1 shows the map of the 2020 Q3 employment change at the LLM level. The degree of treatment effect heterogeneity is striking. Except for a few small clusters, the crisis does not seem to unfold along well-defined spatial dimensions or the North-South axis. Nevertheless, some local economies have been hit much harder than others, with impacts ranging from drops larger than 20% in some LLMs of Lombardy, Veneto, Liguria, Calabria, Sicily, and Sardinia to small decreases or even mildly positive effects in Piedmont, Marche, Umbria, Lazio, Abruzzo, and Molise. On top of regional-level differences, what is even more striking is the *within-region* heterogeneity, which shows how, in all Italian regions, some LLMs fared much better than others despite being geographically close and often contiguous. From an economic geography perspective, our findings suggest that the spatial dimension played a minor role as a transmission channel of the crisis's impacts and suggests a far more prominent role of LLM-specific sectoral characteristics and labor market features. Figure A1 in the Appendix displays the temporal evolution of the employment effects over the first three quarters of 2020: only in the third quarter of 2020, the impacts appear, and local trajectories start to diverge.

Figure 1 – Employment change 2020 Q3



We then compare the employment and epidemiological outcomes engendered by COVID-19. Figure A4 in the Appendix presents a visual comparison between the economic vs. epidemiological effects of COVID-19 in Italy. Looking at the maps, the geographic distribution of impacts seems at odds with the COVID-19 epidemiological spread during the first wave, which is proxied by excess mortality estimates from February 21, 2020, to August 31, 2020. To test the spatial correlation between these outcomes, we measure their overall spatial relationship across all LLMs using the bivariate Moran's I. This index ranges from -1 (perfect negative spatial correlation) to 1 (perfect positive spatial correlation), and we obtained a Moran's I coefficient close to 0 (0.108), which suggests a lack of significant spatial correlation between employment and epidemiological outcomes.

It is worth noticing that the documented employment impacts are net of the Italian government's protective measures. This means that absent these protective measures (the firing freeze and CIGS extensions in particular), local impacts would have likely been even more sizeable.[17]

4.2 Employment by sector

If LLMs' regional or spatial location is not a primary driver, where does the heterogeneous impact on overall employment originate? Sectoral specialization of LLMs is part of the answer. As shown in the maps of employment change in manufacturing and services, depicted in Figure 2 below, the tertiary sector was much more severely affected than the manufacturing one and appears to be the leading cause of the overall employment change observed in Figure 1.[18] This is not unexpected, as the workplace closures affected primarily economic activities in the tertiary sector. At the same time, a large share of manufacturing firms could avert the shutdown thanks to being comprised in the list of 'essential activities' that the government decided to keep open to guarantee the basic functioning of Italy's economic system. The tertiary sector is also notably the one with the highest prevalence of temporary jobs and seasonal workers, which could only marginally benefit from the firing freeze measure. Given these facets, it comes as no surprise that employment losses primarily affected LLMs specialized in services.

Figures A2 (for manufacturing) and A3 (for services) in the Appendix also provide the evolution of impacts by quarter: while the manufacturing sector experienced only a moderate negative trend over the year, the services sector suffered a massive blow during the third quarter, in line with the trajectory of overall employment illustrated in Figure A1.

[17] For example, a recent study by the Bank of Italy suggests that the government's protective policies avoided at least 600,000 firings in 2020: https://www.bancaditalia.it/media/notizie/2020/Nota-Covid-19.11.2020.pdf?language_id=1 (source in Italian).

[18] This is confirmed by the national-level estimates, which unveil an aggregate 0.28% decrease in manufacturing compared to a 2.13% decrease in services.

4.3 Business demography

We then look at how COVID-19 affected business demography outcomes. At the national level, by the end of the third quarter of 2020, the crisis determined a 20.99% decrease in business births and a 2.11% decrease in business deaths. Figure 3 disaggregates these country-level estimates and maps the cumulative impact of COVID-19 for business births change (i.e. firm entries) and business deaths change (firm exits) over the first three quarters of 2020.

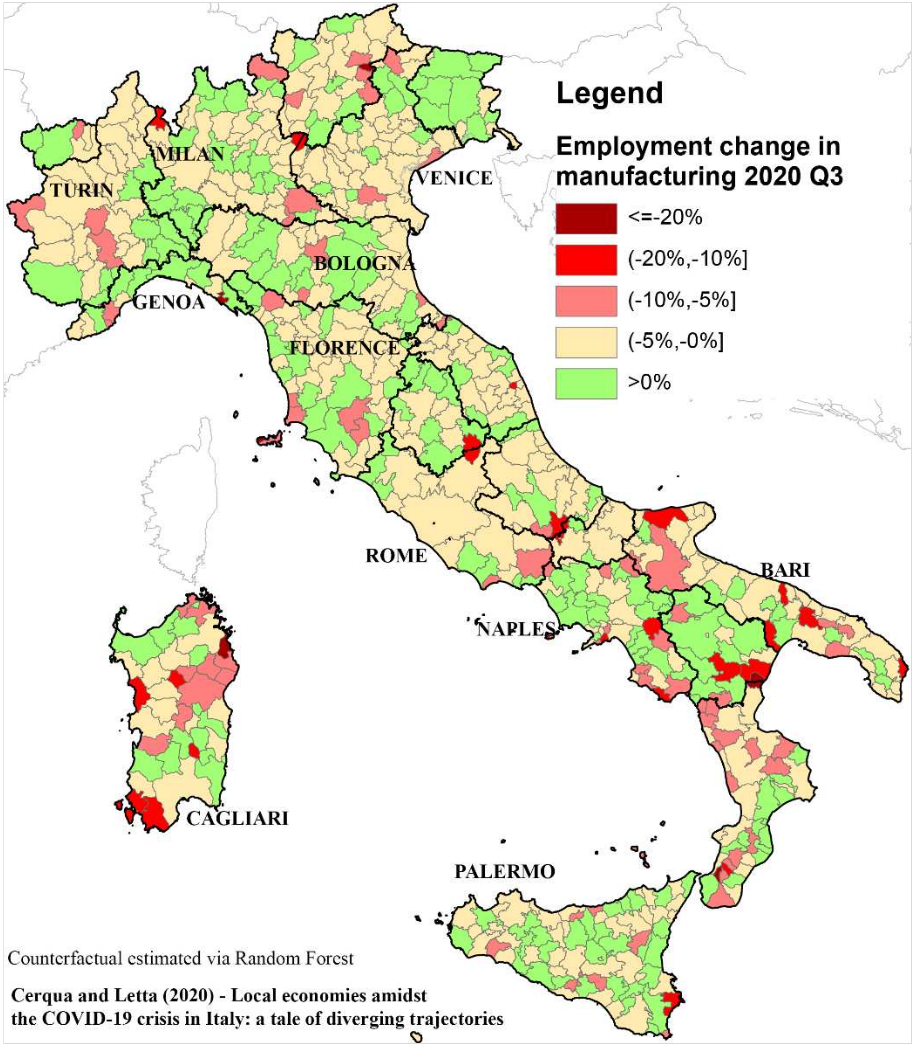
The impact on business births is particularly acute and, with almost no exception, involves the entire national territory. This anomalous plunge happened despite the so-called *Decreto Rilancio* (May 14, 2020), which included a set of protective measures intended to support investments in start-ups (Fini & Sobrero, 2020). In contrast, the impact on firm exits is more polarized and geographically dispersed, with several regions experiencing substantial reductions in cessations of trading, e.g. Emilia-Romagna and Marche, whereas others (Lazio, Abruzzi, Basilicata and, in particular, Sardinia) saw a significant increase in firm exits. Sardinia's case is emblematic as tourism, arguably the hardest-hit sector, plays a vital role in its economy.

The generalized drop in the number of newly-born firms across the country is particularly troublesome because start-ups and young firms are usually the most innovative ones, thus pointing to dire forecasts about the potentially long-lasting effects of the fall in business births in terms of aggregate productivity growth.[19] Moreover, this lost generation of firms creates a persistent dent in overall employment as subsequent years will be characterized by a lower number of firms (Sedláček, 2020). This is all the more worrying in Italy, a country whose economic dynamism – its ability and willingness to allocate resources efficiently – has been steadily declining in the last quarter of a century (Rossi & Mingardi, 2020).

[19] On this issue, see also https://www.lavoce.info/archives/68205/cosi-il-coronavirus-ha-contagiato-limprenditorialita/ (in Italian).

Figure 2 – Employment change 2020 Q3 by sector

Manufacturing



Services

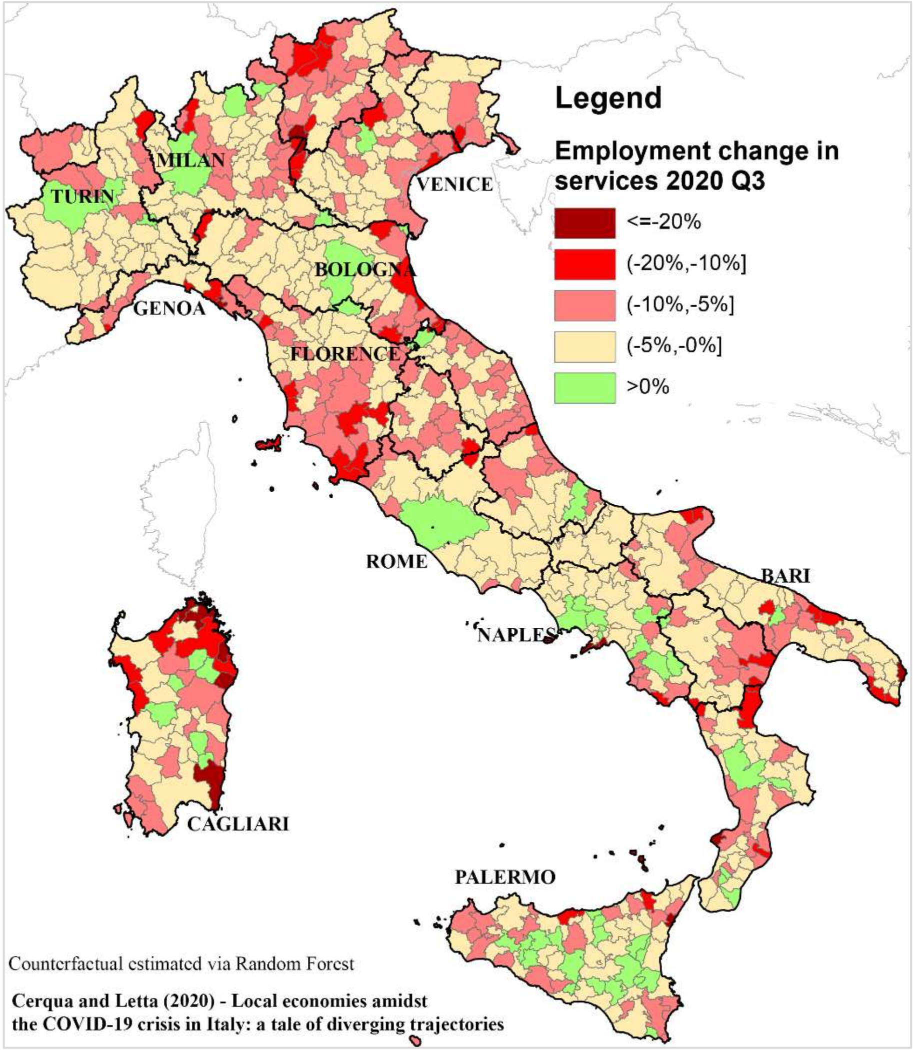
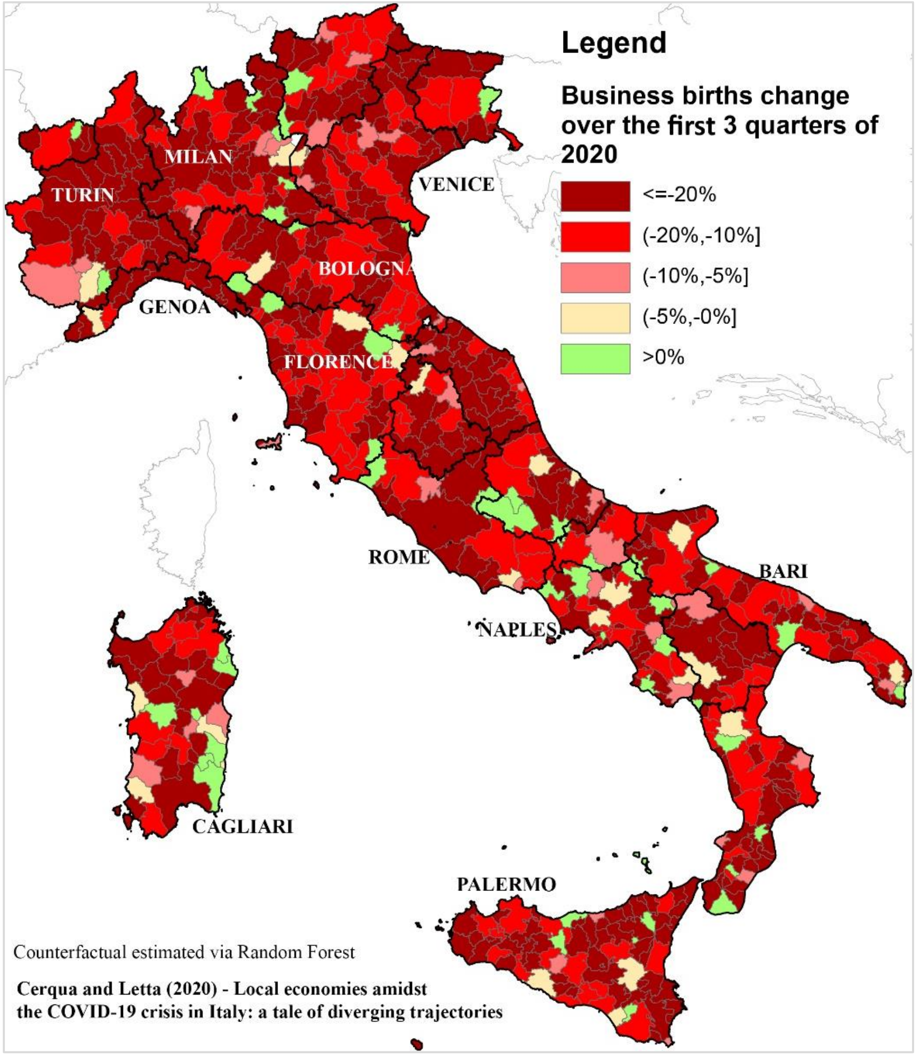
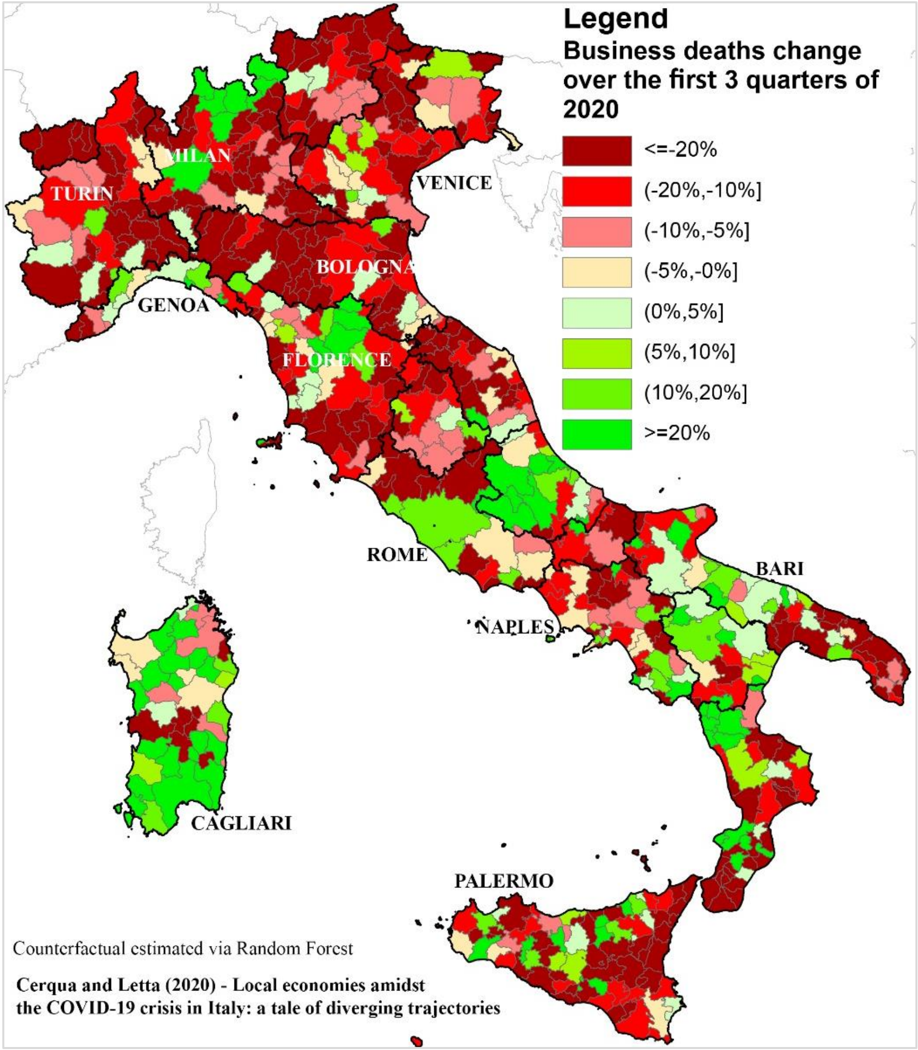


Figure 3 – Business births and deaths change 2020 Q1-Q3

Business births



Business deaths



5. Association analysis

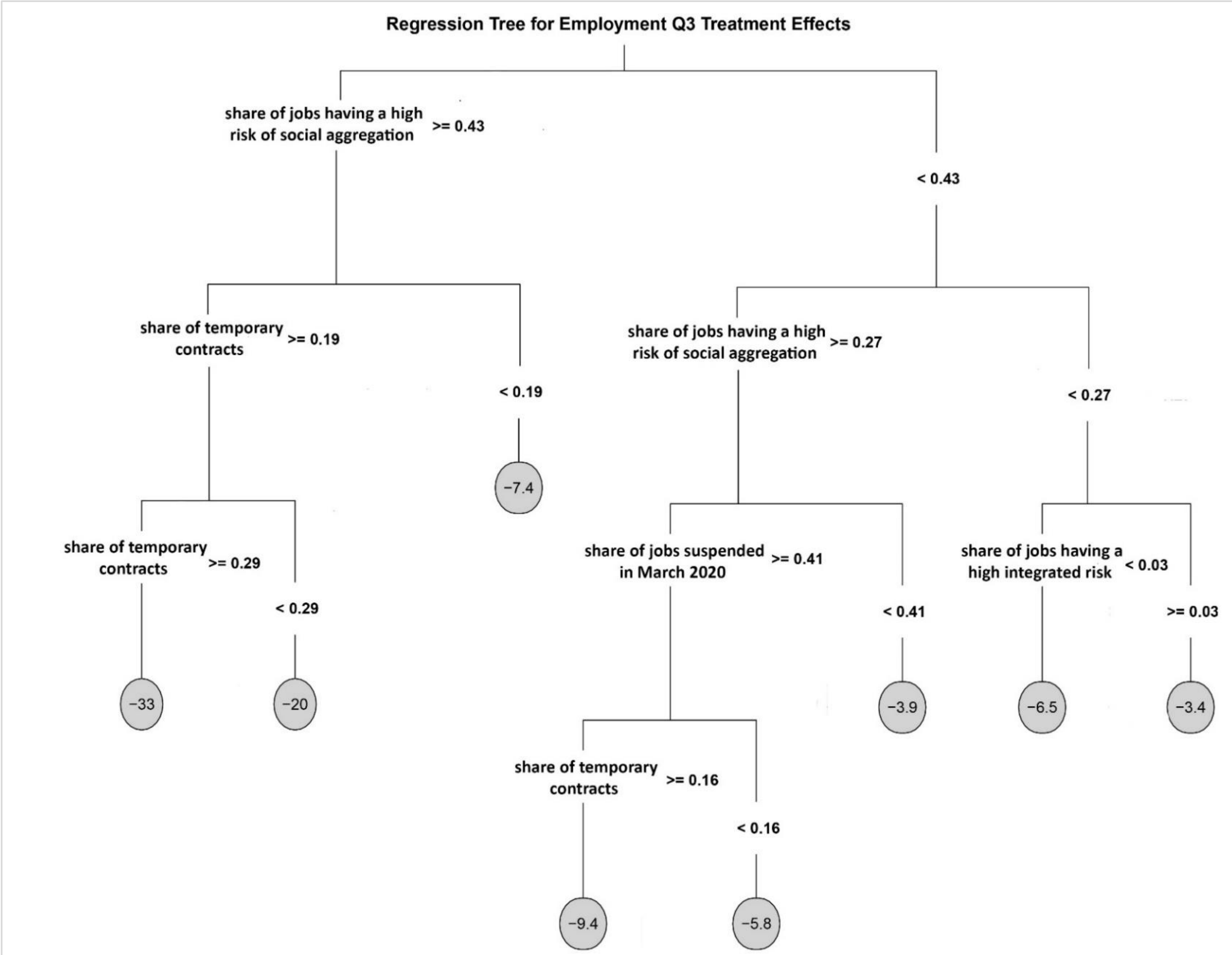
The counterfactual analysis revealed a substantial heterogeneity of the pandemic economic effects. Such heterogeneity does not stem from regional or intra-regional clusters and is partly driven by the LLMs sectoral specialization. Nevertheless, we want to go further than that and understand the factors that matter the most in generating such a fragmented landscape. Therefore, in this section, we use a regression tree to examine the main predictors of our primary variable of interest, employment.

Figure 4 illustrates the regression tree of the LLM-specific overall employment treatment effects. The tree reveals interesting patterns. First, the few variables that generate the tree belong exclusively to two variable groups: aggregation and mobility features and labor market characteristics. Second, the most severely affected LLMs are those in which there is a high share of jobs at a high risk of social aggregation and a high share of jobs suspended in March 2020, and, even more importantly, a high share of temporary contracts. For instance, the tree predicts that LLMs with a share of jobs having a risk of aggregation equal to or higher than 43% and a share of temporary contracts equal to or higher than 29% will experience a 33% drop in employment.

Exposure to high aggregation and proximity risk seems to be a primary discriminant of impacts across LLMs with different shares or 'workers at risk' (Barbieri et al., 2020). The relevance of these labor market attributes in generating the regression tree provides empirical support for the above discussion on the unequal exposure of different workers' categories and types of contracts in the face of the crisis, in line with the heterogeneous findings of Casarico and Lattanzio (2020) for Italy and Blundell et al. (2020) for the UK. This analysis also suggests that emergency measures and fiscal packages were by design effective only for specific categories of workers and types of contracts. In contrast, more fragile categories (think of seasonal workers and occasional jobs) proved to be more vulnerable to the crisis's labor market consequences.

In sum, the interactions between economic sectors having high social aggregation risks and fragile labor markets are associated with sharp drops in overall employment at the local level.

Figure 4 – Regression tree on employment change 2020 Q3



6. Conclusions

In a first and preliminary *ex-post* impact evaluation analysis, based on the use of up-to-date quarterly data and a predictive ML method in a causal inference setting, we have documented the striking level of inequality of the economic impacts of the coronavirus crisis across the Italian territory. This heterogeneity is associated with LLM-specific features such as sectoral specialization, exposure of economic activities to high social aggregation risks, and pre-existing labor market vulnerabilities. In contrast, there is no discernible spatial correlation between the economic and epidemiological geographical patterns of the pandemic.

We deem the local and sectoral dimensions of the crisis to be policy-relevant, especially in light of the current political debate on the allocation of the forthcoming resources from the aid mechanisms developed by the European Union, namely the *Recovery Plan* and the *NextGenerationEU* initiative. A broad glance at the national level can capture the generalized sharp drop in firm entries but overlooks the high degree of unevenness in the effects on employment levels and business deaths across the Italian territory.

Coupled with the relevant role played by labor markets' insecurity emerged from the association analysis, these findings call for more research to untangle the local economic impacts of the pandemic and a place- and sector-based approach in the policy response to the crisis. National policies and top-down plans will be insufficient to lead the recovery (Bailey et al., 2020). Therefore, to inequality and heterogeneity must correspond *ad hoc*, well-targeted policy interventions based on a local and place-based perspective that considers the territorial profile and sectoral specialization of local economic systems (Ascani et al., 2020). Only in this way will it be possible to attenuate the disruptive consequences of the COVID-19 upheavals and avoid the unfolding crisis that will further exacerbate pre-existing inequalities among Italian local economies.

REFERENCES

Abadie A, Diamond A, & Hainmueller, J (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490): 493–505.

Abrell J, Kosch M, & Rausch S (2019). How effective was the UK carbon tax? A machine learning approach to policy evaluation. A Machine Learning Approach to Policy Evaluation (April 15, 2019). CER-ETH–Center of Economic Research at ETH Zurich Working Paper, 19, 317.

Adams-Prassl A, Boneva T, Golin M, & Rauh C (2020). Inequality in the impact of the coronavirus shock: Evidence from real time surveys. *Journal of Public Economics*, 189: 1–33.

Andini M, Ciani E, de Blasio G, D'Ignazio A, & Salvestrini V (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization*, 156: 86–102.

Ascani A, Faggian A, & Montresor S (2020). The geography of COVID-19 and the structure of local economies: The case of Italy. *Journal of Regional Science,* first published online: 20 November 2020.

Athey S, & Imbens G (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Athey S, Bayati M, Doudchenko N, Imbens G, & Khosravi K (2018). Matrix completion methods for causal panel data models. National Bureau of Economic Research, No. w25132.

Athey S, Bayati M, Imbens G, & Qu Z (2019). Ensemble methods for causal effects in panel data settings. *AEA Papers and Proceedings*, 109: 65–70.

Bailey D., Clark J, Colombelli A, Corradini C, De Propris L, Derudder B, ... & Kemeny T (2020). Regions in a time of pandemic. *Regional Studies*, 54(9): 1163–1174.

Baker SR, Farrokhnia RA, Meyer S, Pagel M, & Yannelis C (2020). How does household spending respond to an epidemic? consumption during the 2020 covid-19 pandemic. *The Review of Asset Pricing Studies*, 10(4): 834–862.

Barbieri T, Basso G, & Scicchitano S (2020). Italian Workers at Risk During the Covid-19 Epidemic. GLO Discussion Paper Series No. 513.

Bartik AW, Bertrand M, Cullen Z, Glaeser EL, Luca M, & Stanton C (2020). The impact of COVID-19 on small business outcomes and expectations. *Proceedings of the National Academy of Sciences*, 117(30): 17656–17666.

Belloni A, Chernozhukov V, Fernández-Val I, & Hansen C (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1): 233–298.

Benatia D (2020). Reaching new lows? The pandemic's consequences for electricity markets. USAEE Working Paper No. 20–454.

Benatia D, de VILLEMEUR EB (2019). Strategic reneging in sequential imperfect markets. CREST Working Papers 2019-19.

Benedetti FC, Sedláček P, & Sterk V (2020). EU start-up calculator: impact of COVID-19 on aggregate employment. EUR 30372 EN, Publications Office of the European Union, Luxembourg, 2020.

Bick A, & Blandin A (2020). Real-time labor market estimates during the 2020 coronavirus outbreak. SSRN Electronic Journal No. 3692425.

Bijnens G, Karimov S, Konings J (2019). Wage indexation and jobs. A machine learning approach. VIVES Discussion Paper No. 82.

Blundell R, Costa Dias M, Joyce R, & Xu X (2020). COVID-19 and Inequalities. *Fiscal Studies*, 41(2): 291–319.

Buchheim L, Krolage C, & Link S (2020). Sudden Stop: When Did Firms Anticipate the Potential Consequences of COVID-19? CESifo Working Paper No. 8429.

Burlig F, Knittel CR, Rapson D, Reguant M, & Wolfram C (2020). Machine learning from schools about energy efficiency. *Journal of the Association of Environmental and Resource Economists*, 7(6): 1181–1217.

Cainelli G, Iacobucci D, & Micozzi A (2013). Determinants of territorial differences in entrepreneurial rates. An empirical analysis of Italian local systems. 53rd Congress of the European Regional Science Association (ERSA), Palermo, Italy.

Cajner T, Crane LD, Decker RA, Grigsby J, Hamins-Puertolas A, Hurst E, ... & Yildirmaz A (2020). The US labor market during the beginning of the pandemic recession. National Bureau of Economic Research No. w27159.

Carvalho VM, Hansen S, Ortiz A, Garcia JR, Rodrigo T, Rodriguez Mora S, & Ruiz de Aguirre P (2020). Tracking the COVID-19 crisis with high-resolution transaction data. CEPR Discussion Papers No. 14642.

Casarico A, & Lattanzio S (2020). The heterogeneous effects of COVID-19 on labor market flows: Evidence from administrative data. Covid Economics, 52: 152–174.

Caselli M, Fracasso A, & Scicchitano S (2020). From the lockdown to the new normal: An analysis of the limitations to individual mobility in Italy following the Covid-19 crisis. GSSI Discussion Paper Series in Regional Science & Economic Geography No.7/2020.

Cerqua A, Di Stefano R, Letta M, & Miccoli S (2020). Local mortality estimates during the COVID-19 pandemic in Italy. GSSI Discussion Paper Series in Regional Science & Economic Geography No.6/2020**.**

Chetty R, Friedman JN, Hendren N, & Stepner M (2020). The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data. National Bureau of Economic Research No. w27431.

Chudik A, Mohaddes K, Pesaran MH, Raissi M, & Rebucci A (2020). A counterfactual economic analysis of Covid-19 using a threshold augmented multi-country model. National Bureau of Economic Research No. w27855.

Fini R, & Sobrero M (2020). Why Italy needs an entrepreneurial renaissance after COVID-19, in Bellettini G. and Goldstein A. *The Italian economy after Covid-19. Short term costs and long-term adjustments*, Bononia University Press, Bologna.

Forsythe E, Kahn LB, Lange F, & Wiczer D (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. *Journal of Public Economics*, *189*: 1–7.

Giupponi G, & Landais C (2020). Subsidizing Labor Hoarding in Recessions: The Employment & Welfare Effects of Short Time Work. Mimeo, May 2020 version.

Gourinchas PO, Kalemli-Özcan Ṣ, Penciakova V, & Sander N (2020). *Covid-19 and SME Failures*. National Bureau of Economic Research No. w27877.

Hastie T, Tibshirani R, & Friedman J (2009). *The elements of statistical learning: data mining,* *inference, and prediction*. Springer Science & Business Media.

Kleinberg J, Ludwig J, Mullainathan S, & Obermeyer Z (2015). Prediction policy problems. *American Economic Review*, 105(5): 491–495.

Lantz B (2019). *Machine learning with R: expert techniques for predictive modeling*. Packt Publishing Ltd.

Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, & Yu B (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44): 22071–22080.

Rossi N, & Mingardi A (2020). Italy and COVID-19: Winning the war, losing the peace? *Economic Affairs*, 40(2): 148–154.

Sedláček P (2020). Lost generations of firms and aggregate labor market dynamics. *Journal of Monetary Economics*, 111: 16–31.

Souza M (2019). Predictive counterfactuals for treatment effect heterogeneity in event studies with staggered adoption. SSRN Electronic Journal No. 3484635.

Varian HR (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences*, 113(27): 7310–7315.

Von Gaudecker HM, Holler R, Janys L, Siflinger B, & Zimpelmann C (2020). Labour supply in the early stages of the CoViD-19 Pandemic: Empirical Evidence on hours, home office, and expectations. IZA Discussion Paper Series No. 13158.

Wager S, & Athey S (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

Appendix

Table A1 – Definition of the variables included in the analysis

Variable name	**Definition**	**Time period**	**Source**
	Counterfactual analysis		
Employment	Overall employment of private non-financial sector firms	2014 Q3 – 2020 Q3	Business Register
Employment in manufacturing	Overall manufacturing employment	2014 Q3 – 2020 Q3	Business Register
Employment in services	Overall services employment	2014 Q3 – 2020 Q3	Business Register
Business births	Companies that have registered in the period under review	2014 Q1 – 2020 Q3	Business Register
Business deaths	Companies that went out of business in the period under review	2014 Q1 – 2020 Q3	Business Register
Economic classification dummies	Without specialization, non-manufacturing (touristic), non-manufacturing (non-touristic), made in Italy, other manufacturing	2011	Istat
Geographical dummies	North-East, North-West, Centre, South		Istat
Population density	Resident population per unit area	2014-2019	Istat
Unemployment rate	Resident population aged 15+ not in employment but currently available for work	2014-2019	Istat
Activity rate	The number of people employed and those unemployed as a % of the total population	2014-2019	Istat
	Association analysis		
Employment change Q3 2020	Treatment effect of the COVID-19 crisis on overall employment levels	2020 Q3	Estimated via the MLCM
Unemployment rate	Resident population aged 15+ not in employment but currently available for work	2019	Istat
Economic classification dummies	Without specialization, non-manufacturing (touristic), non-manufacturing (non-touristic), made in Italy, other manufacturing	2011	Istat
Excess mortality estimates	Local excess mortality estimated via applying ML techniques to all-cause deaths data	From Feb 21, 2020 to Aug 31, 2020	Cerqua et al. (2020)
Share of jobs having a high risk of social aggregation	Number of employees exposed to a medium-high or high risk of social aggregation divided by the number of employees	2019	Own calculations using Business Register data

Table A1 – Continued

Share of jobs having a high integrated risk	Number of employees exposed to a medium-high or high integrated risk divided by the number of employees	2019	Own calculations using Business Register data
Share of short-term contracts	Number of employees with short-term contracts in October divided by the number of employees in October	2015	Istat
Share of jobs in suspended economic activities	Share of jobs in activities suspended in March 2020 by the Italian Government due to the spread of the pandemic	2017	Istat
Income per capita	The amount of money earned per person	2018	Ministry of Economy and Finance
Share of innovative start-ups	The ratio between innovative start-ups and the universe of firms registered in the Business Register	Average (2016-2019)	Business Register
Share of firms having employees in CIGS	The number of firms with employees in CIGS divided by the universe of firms registered in the Business Register	Average (2015-2018)	Ministry of Labor and Social Policies
Number of road accidents per 10,000 inhabitants	The number of road accidents with injuries to persons divided by resident population * 10,000.	2019	Istat
Dependency ratio	The ratio of those typically not in the labor force (the dependent part, ages 0 to 14 and 65+) and those typically in the labor force (the productive part, ages 15 to 64)	Jan 1, 2020	Istat
Share of population living in peripheral areas	Share of population living in areas defined by Istat as peripheral or ultra-peripheral	Jan 1, 2020	Istat
Index of relational intensity (IIRFL)	The percentage of flows within a LLM that connect different municipalities on the total of flows within the LLM. This indicator ranges from values close to 0 to 100 (case in which all the workers of the municipalities of the LLM go to work in another municipality). The higher the indicator, the greater the inter-municipal turbulence in terms of flows.	2011	Istat

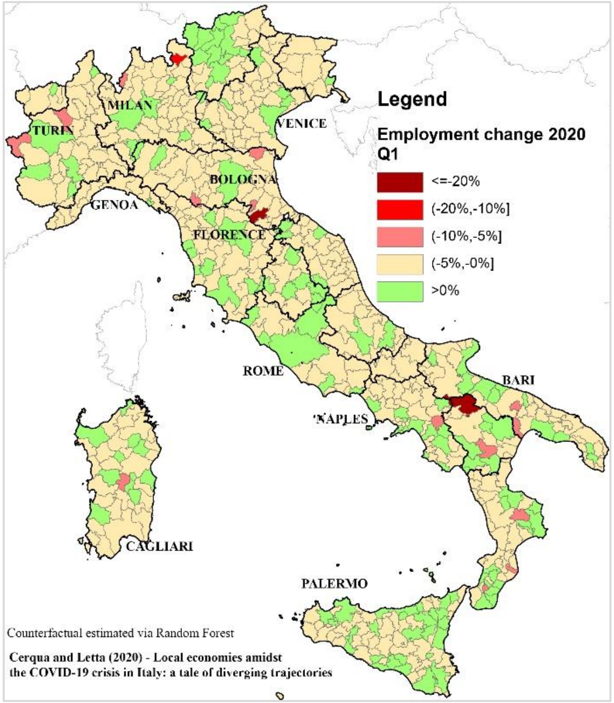
Notes: To determine the flow of registrations in a given period – e.g. 2nd trimester 2019 - the firms' universe extracted from the archive on June 30 is compared with that extracted in the previous quarter (March 31). Firms that are present in the 2nd (1st) quarter but not in the 1st (2nd) are classified as new registrations (companies that went out of business). Outcome variables in bold.

Table A2 – Descriptive statistics

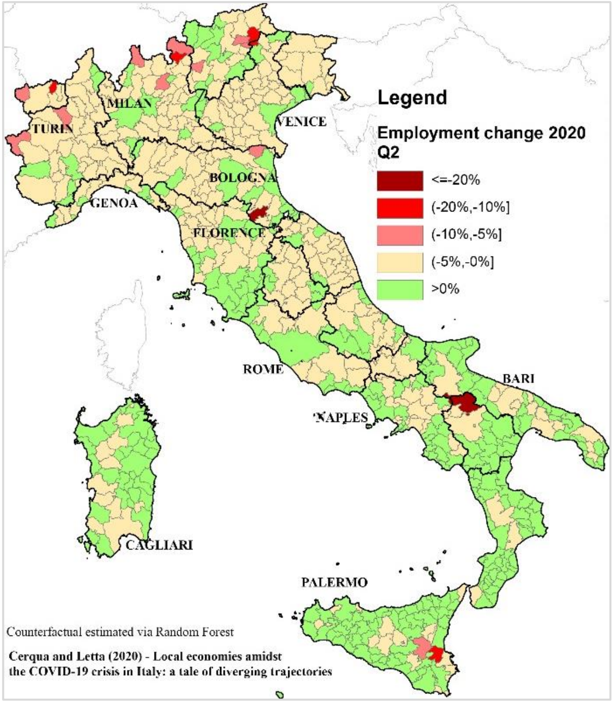
Variable name	Mean	SD	Min	Max
Counterfactual analysis				
Employment (log)	9.31	1.25	5.95	14.41
Employment in manufacturing (log)	7.53	1.61	3.37	12.65
Employment in services (log)	8.89	1.29	5.51	14.22
Business births	55.97	236.18	0	5173
Business deaths	44.63	202.79	0	9685
Share of LLMs without specialization	0.19	0.39	0	1
Share of touristic LLMs	0.14	0.34	0	1
Share of non-manufacturing (non-touristic) LLMs	0.23	0.42	0	1
Share of *made in Italy* LLMs	0.31	0.46	0	1
Share of manufacturing LLMs	0.14	0.35	0	1
<=10,000 inhabitants	0.08	0.28	0	1
(10,000; 50,000]	0.46	0.50	0	1
(50,000; 100,000]	0.25	0.43	0	1
(100,000; 500,000]	0.18	0.39	0	1
> 500,000 inhabitants	0.03	0.16	0	1
Activity rate	48.26	6.66	30.15	63.91
Unemployment rate	11.85	6.17	1.19	39.08
Population density	0.21	0.30	0.01	3.17
Association analysis				
Employment change Q3 2020 (%)	-5.17	5.50	-44.73	6.78
Share of LLMs without specialization	0.19	0.39	0	1
Share of touristic LLMs	0.14	0.34	0	1
Share of non-manufacturing (non-touristic) LLMs	0.23	0.42	0	1
Share of *made in Italy* LLMs	0.31	0.46	0	1
Share of manufacturing LLMs	0.14	0.35	0	1
Unemployment rate (%)	10.99	5.91	1.19	36.19
Excess mortality estimates (%)	7.98	22.38	-32.28	173.01
Share of jobs in suspended economic activities	0.47	0.08	0.25	0.79
Income per capita (€)	12705	3588	5882	22118
Share of firms having employees in CIGS	0.0008	0.0007	0	0.0046
Share of population living in peripheral areas	0.29	0.40	0	1
Share of short-term contracts	0.19	0.08	0.10	0.56
Number of road accidents per 10,000 inhabitants	2.18	1.20	0	6.94
Index of relational intensity (IIRFL)	25.70	14.48	0.2	66.1
Dependency ratio	0.58	0.05	0.43	0.78
Share of innovative start-ups	0.003	0.003	0	0.017
Share of jobs having a high risk of social aggregation	0.23	0.11	0.06	0.76
Share of jobs having a high integrated risk	0.06	0.03	0.01	0.37
Number of LLM-quarters (whole sample)	10,370			
Number of LLMs	610			

Figure A1 – 2020 Employment change by quarter

Quarter 1



Quarter 2



Quarter 3

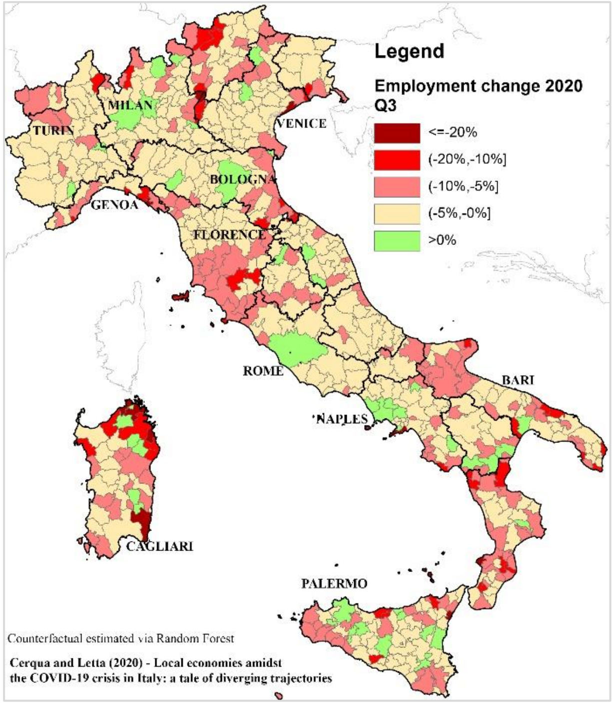
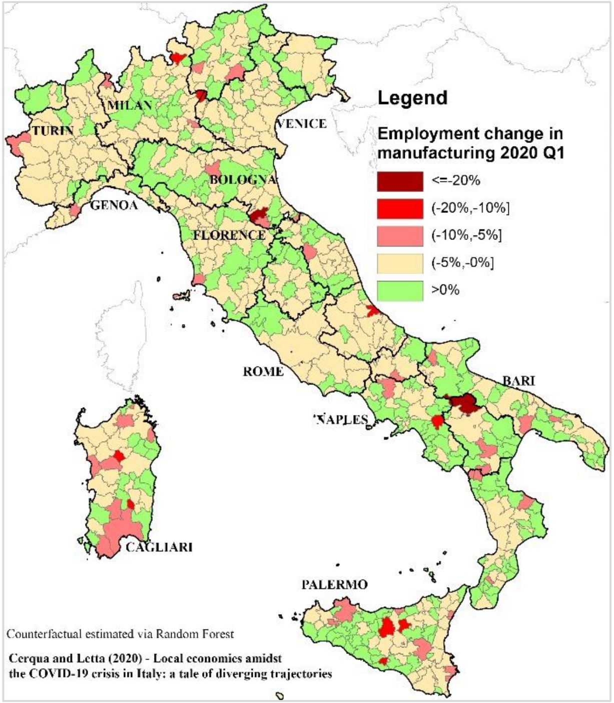
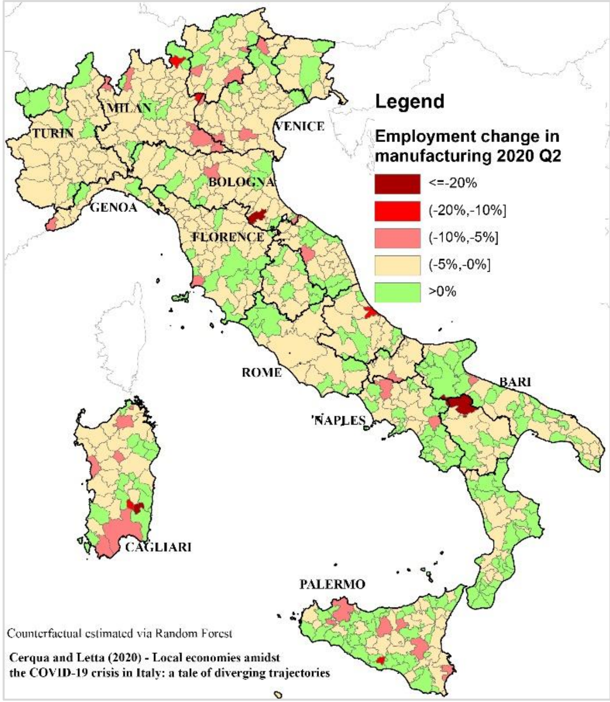


Figure A2 – 2020 Employment change in manufacturing by quarter

Quarter 1



Quarter 2



Quarter 3

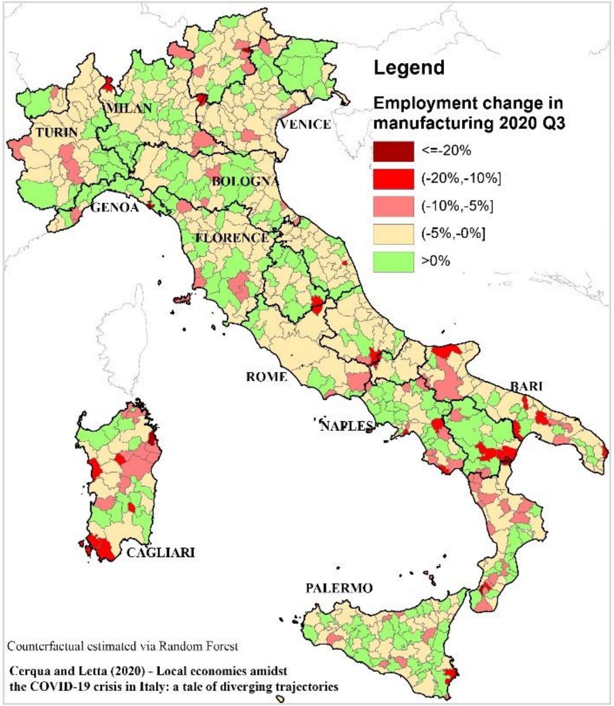


Figure A3 – 2020 Employment change in services by quarter

Quarter 1 Quarter 2 Quarter 3

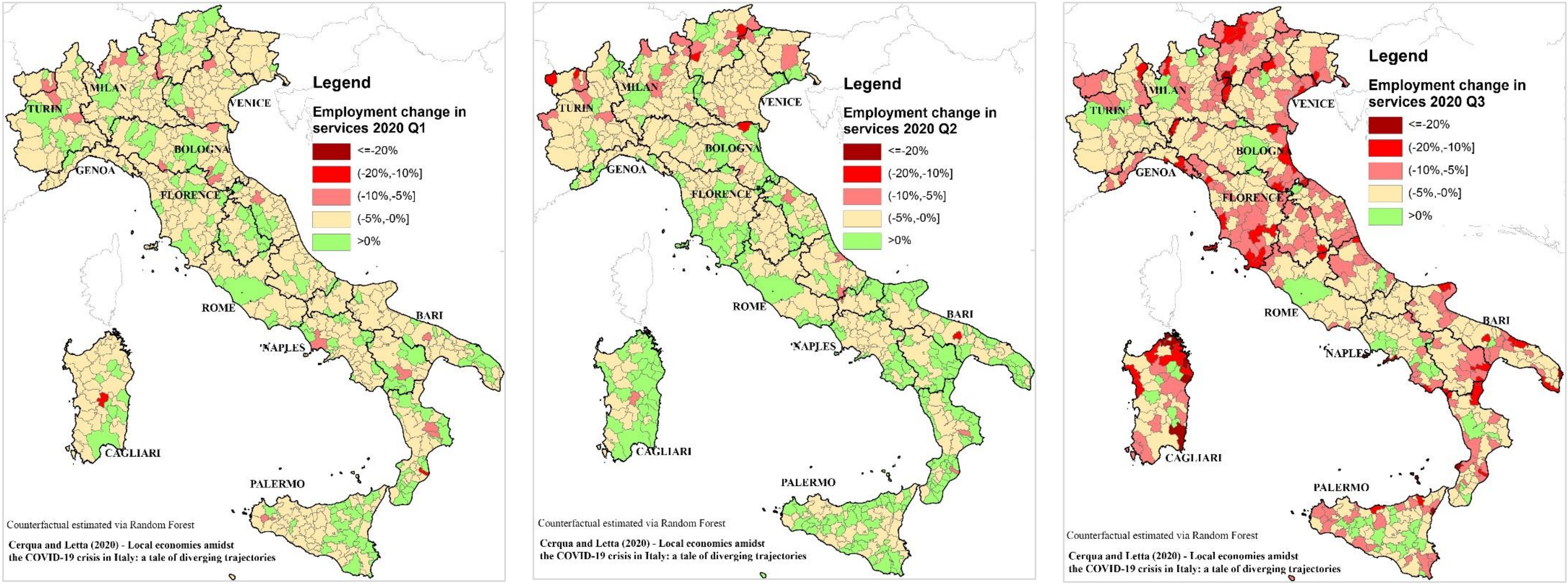
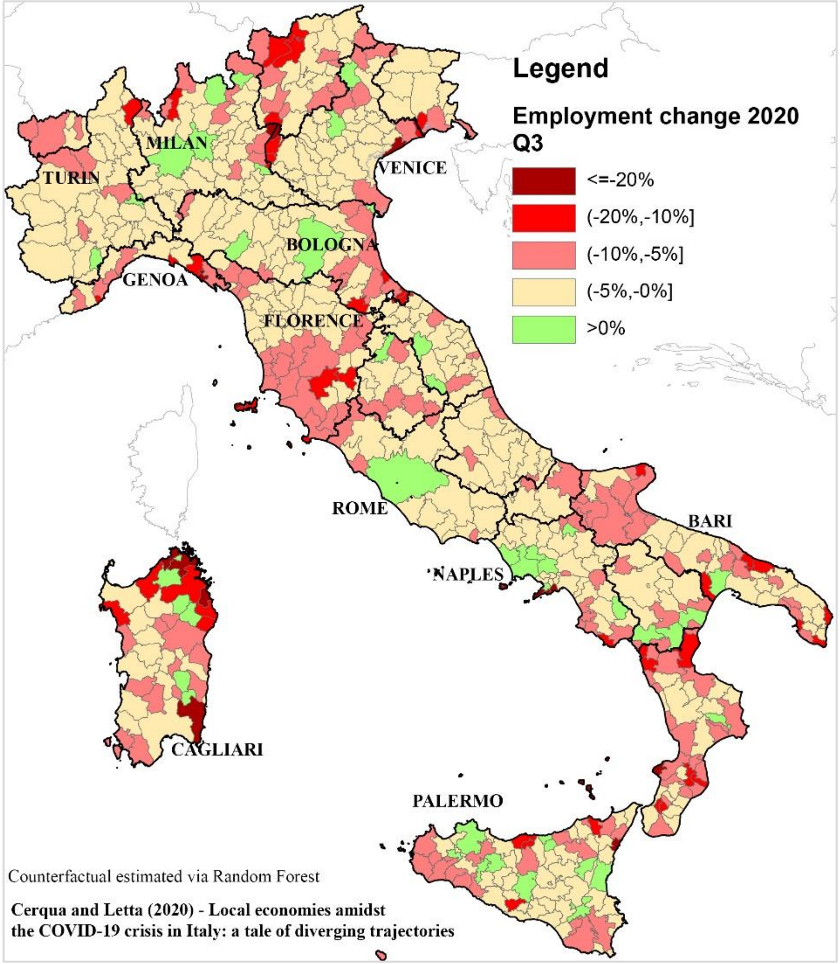
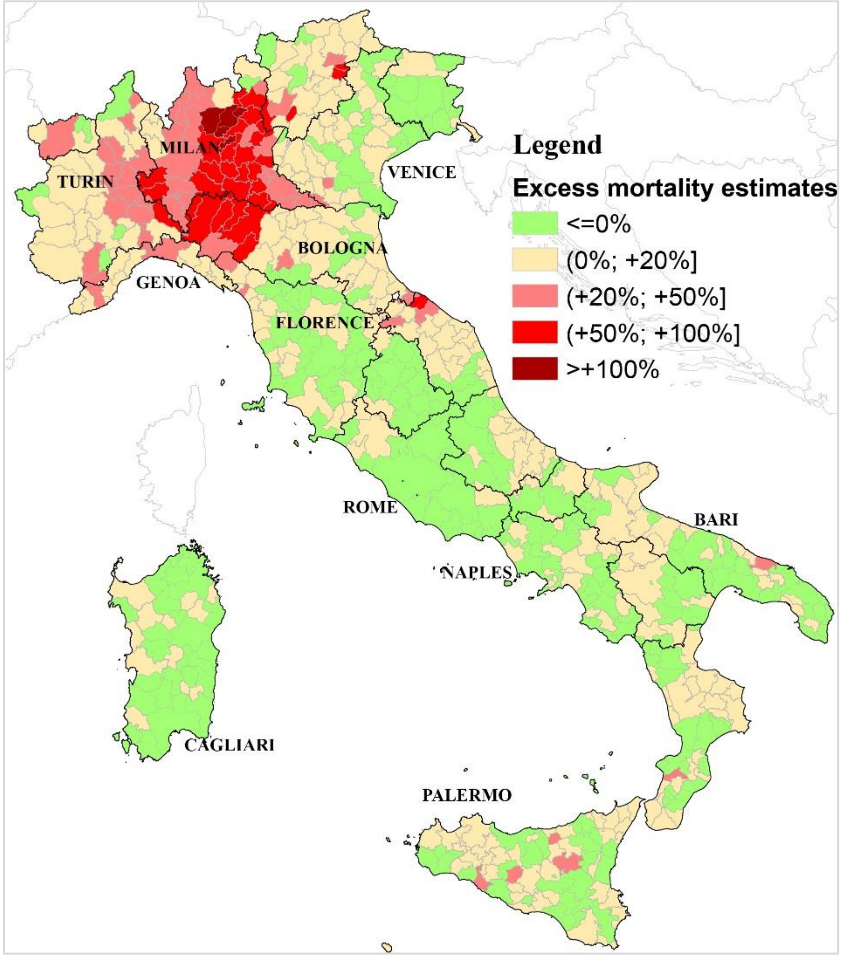


Figure A4 – Economic vs epidemiological impacts of the COVID-19 pandemic across Italy

Employment change 2020 Q3



Excess mortality (21 Feb 2020 – 31 Aug 2020)



Notes: The excess mortality estimates from Feb 21 2020 to Aug 31 2020 are from Cerqua et al. (2020).