Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Local inequalities of the COVID-19 crisis

Augusto Cerqua*, Marco Letta

Department of Social Sciences and Economics, Sapienza University of Rome, P.le Aldo Moro, 5, 00185, Rome, Italy

ARTICLE INFO

JEL classification:
C53
D22
E24
R12

Keywords:
Impact evaluation
Counterfactual approach
Machine learning
Local labor markets
COVID-19
Italy

ABSTRACT

This paper assesses the pandemic’s impact on Italian local economies with the newly developed machine learning control method for counterfactual building. Our results document that the economic effects of the COVID-19 shock vary dramatically across the Italian territory and are spatially uncorrelated with the epidemiological pattern of the first wave. The largest employment losses occurred in areas characterized by high exposure to social aggregation risks and pre-existing labor market fragilities. Lastly, we show that the hotspots of the COVID-19 crisis do not overlap with those of the Great Recession. These findings call for a place-based policy response to address the uneven economic geography of the pandemic.

1. Introduction

With over 132,200 deaths and more than 4,780,000 cases (as of November 4, 2021), Italy ranks among the worst-hit countries by COVID-19.1 The Italian government was the first in Europe to declare, on March 9, 2020, 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 authorities to issue new social distancing policies, including the reintroduction of restrictive measures targeting economic activities.

The Italian government tried to attenuate the impacts of such disruptive events 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, 2020, which introduced two exceptional labor market policies: a special COVID-19 short-time work retroactive compensation scheme and a freezing of layoffs, which have later been repeatedly extended and are still partially in place at the time of writing.

Despite the implementation of a wide range of policy interventions, Italy’s GDP contracted by 8.9% in 2020.3 Besides, the Bank of Italy reports, for 2020, a reduction of 11% in the number of hours worked and a decrease of 2.1% in the number of persons employed.

Remarkably, ex-post evaluations of the spatial distribution of the economic effects of the COVID-19 emergency are still missing. Such a vacuum was hardly surprising initially, as real-time microdata are scarce, but as the epidemiological impacts finally attenuate, there is a strong case for a comprehensive analysis of the economic geography of the pandemic crisis. On top of data scarcity, rigorous impact evaluation is challenging for econometric issues: the COVID-19 shock virtually left no part of the Italian territory unaffected, and the national lockdown involved the entire country. In econometric jargon, it is hard to find a control group because the treatment affected all units simultaneously or with short lags. Therefore, while in countries and areas where no total lockdowns were
implemented one can exploit staggered or heterogenous policy responses to generate a counterfactual scenario, standard evaluation techniques, such as difference-in-difference or the synthetic control method (SCM), are not applicable to the Italian context. This is probably the reason why, although the literature on the pandemic is flourishing (Adams-Prassl et al., 2020; Baker et al., 2020; Bartik et al., 2020; Beneventi et al., 2020; Bick and Blandin, 2020; Bloom et al., 2021; Blundell et al., 2020; Cajner et al., 2020; Carvalho et al., 2020; Chetty et al., 2020; Chudik et al., 2021; Forsythe et al., 2020; Gourinchas et al., 2020; Von Gaudecker et al., 2020), the Italian case still remains relatively unexplored.

Among the few exceptions, Ascani et al. (2021) provide evidence of a close relationship between COVID-19 disease patterns and local economies' characteristics. Through a linear probability model, Casarico and close relationship between COVID-19 disease patterns and local economies (LLMs), investigate the main territorial features of such unequal impacts of the pandemic (Bailey et al., 2020). Indeed, in Europe as elsewhere, the current crisis is undoubtedly a recession are temporary and less-educated workers. Such workers are being poorer and characterized by high unemployment rates and vulnerable labor markets. Besides, the micro-level evidence mentioned above shows that those who have been more affected by the pandemic shock is hitting different industries, including leisure, hospitality, and tourism, which is also confirmed by our analysis. Finally, our results are in line with those of Chetty et al. (2020), who document sharp differences in the geographic patterns of employment changes between the COVID crisis and Great Recession in the US, and point out that job losses during the pandemic tend to be more concentrated in affluent areas. In conclusion, we provide empirical support in favor of the view that the pandemic crisis is different from a typical recession.
2. Data

Our primary dependent variable is the log of overall employment in the private non-financial sector. 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 are generated using restricted-access data collected from the Business Register kept by the Union of the Italian Chambers of Commerce (Unione camere). 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 have been 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 2016 Q3 to 2020 Q4 and employ the random forest algorithm described in Section 3.7 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, 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 (North-East, North-West, Centre and South), log of population, population density, unemployment rate, activity rate, yearly and quarterly fixed effects, share of foreign population 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, all the available yearly lags, the last pre-treatment lag of the outcome variable, and the same lags of the other outcome variables.

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 variables plausibly 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. (2021), updated to September 30, 2020.7 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., 2021): the share of jobs having a high risk of social aggregation and the share of jobs having a high ‘integrated’ risk. These variables proxy for the demand-side changes due to peoples’ immediate response to the pandemic and are generated on the basis of the work conducted by an ad hoc task force,8 which linked a level of social aggregation to each economic sector (2-digit NACE Rev.2 classification) and integrated risks from low to high. Activities at high integrated risk are those associated with the risk of coming into contact with sources of contagion at work, especially those connected to work processes (e.g. human health services, sewerage, public administration and defense), while activities at high risk of social aggregation are those that involve contact with other subjects in addition to the company’s workers (e.g. catering, entertainment, hospitality).

As the geography of industries highly exposed to the ‘COVID-19 shock’ is heterogeneous (Krueger et al., 2020), we create a variable that incorporates the predicted supply-side sectoral shocks to each LLM. Specifically, we generate the share of jobs in suspended economic activities from March to May 2020.9 In addition, we build the share of temporary contracts as a metric for temporary jobs’ local relative importance.10 Other economic variables included in this phase of the analysis are per capita income, unemployment rate, 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.11 We also add two variables that consider the densities of health care personnel and hospitals: i) the number of hospital beds per 1,000 inhabitants, and ii) the share of workers employed in the NACE 2-digit sectors ‘human health activities’ and ‘residential care activities’.

Lastly, as mobility is one of the critical aspects linked to the epidemiological spread of COVID-19, we take this 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 (IIRFL) within the local labor market. The higher the IIRFL, the greater the inter-municipal turbulence in terms of flows.

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

3. Methods

Our empirical exercise consists of three tasks – a counterfactual analysis, an association analysis, and a comparative analysis. For all three steps, we harness ML’s statistical power: the random forest algorithm for the counterfactual and comparative analyses, and a regression tree for the association analysis. Below, we separately discuss the three empirical analyses.

3.1. Counterfactual analysis: the machine learning control method

To tackle the econometric challenges related to the pandemic shock’s pervasive nature and establish causality, we draw on the newly developed MLCM to generate a counterfactual scenario in which the COVID-19 pervasive nature and establish causality, 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

---

6 In our business demography analyses, we consider all types of firms, including those registered to the Business Register but having 0 employees.
7 Please note that for business demography variables, instead, the sample starts from 2015 Q1.
8 The ‘made in Italy’ manufacturing LLMs are characterized by industrial districts. Most of them are specialized in the manufacture of food products, furniture, textiles, apparel, leather and footwear.
9 These data are publicly available here: https://www.stimecomunalicovid19.com/.
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 The selection of these activities was carried out on the basis of the NACE Rev.2 classification.
12 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).
13 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 and Landais, 2020).
fundamental problem of causal inference, i.e. the impossibility of observing the potential outcome in the no-treatment scenario, a curse that afflicts 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 and Imbens, 2016; Athey et al., 2021; Athey et al., 2019; Belloni et al., 2017; Hofman et al., 2021; Varian, 2014, 2016; Wager and Athey, 2018). Varian (2014, 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, 2019; Bijiens et al., 2019; Burlig et al., 2020; Cerqua et al., 2021; 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. (2021) 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. (2021) employ three different ML routines (LASSO, random forest, and stochastic gradient boosting) to derive municipality-level excess mortality estimates during the first wave of 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 an 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 the validity of this approach 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 method: the random forest, which has been defined the most successful general-purpose algorithm in modern times (Howard and Bowles, 2012; Varian, 2014). The random forest is a fully non-linear technique based on the aggregation of many decision trees. In particular, random forest builds many trees (1,000, 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 established by Cerqua et al. (2021), 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 into a training sample, made up of 80% of the LLMs, and a disjoint test set, consisting of the remaining 20%14;
2) We train our random forest algorithm on the training set and perform a 10-fold cross-validation to select the best-performing tuning hyperparameter15;
3) We test the out-of-sample predictive performance on the corresponding pre-2019 testing sample16;
4) We test model accuracy on the entire 2019 sample and compare its predictive performance with those of a battery of other alternative approaches, namely a before-after analysis, which has become a common and intuitive metric to gauge the magnitude of the pandemic’s impact, a traditional OLS model, and a simpler ML algorithm, the regression tree;
5) We repeat the same routine on the entire pre-2020 dataset (which, having four more quarters of observations, enhances random forest’s predictive power compared to the one on which we develop the model) and finally predict, for 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 the causal impact of the COVID-19 shock. We deem it plausible given the massive disruption to the economy brought about by the sudden unexpected arrival of the pandemic. Lastly, please note that, by ‘COVID-19 shock’, we mean the economic shock, i.e. we refer not only to the epidemiological spread of the virus per se, but also to the related behavioral changes and, above all, to the national lockdown and the other non-pharmaceutical interventions that were adopted to contain the health emergency. This implies that, via our counterfactual approach, we capture the total impact of the pandemic on each LLM.

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 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 most in predicting the heterogeneous local economic impact of the pandemic.

Regression trees are an ideal tool to fulfill this purpose for two reasons: 1) 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

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.
15 We use cross-validation to solve the bias-variance trade-off and maximize the out-of-sample performance of the random forest algorithm. (Hastie et al., 2009). Specifically, we employ 10-fold cross-validation on the training sample to select, among different alternatives (p/2, p/3, and p/6), the optimal value of the tuning hyperparameter m, i.e. the number of features p randomly sampled as candidates at each split.
16 This pre-2019 test is necessary to get a measure of the true predictive performance of our models on data from previously unseen LLMs.
be particularly valuable when the model must be shared to support public decision-making (Andini et al., 2018; Lantz, 2019); ii) unlike traditional approaches, 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 Q4, on the full sample of Italy’s LLMs. In other words, we are not interested in out-of-sample prediction, but only in the associations between the features and the outcome. Second, and related, we do not apply cross-validation to select the hyperparameter of the regression tree method (named ‘complexity parameter’, cp) and adopt the commonly adopted default value of 0.01.

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.

Lastly, for the sake of comparison with a traditional regression approach, we also run an OLS-based association analysis using as outcome variable the employment treatment effects estimated using the standard before-after predictive method. This additional check, which draws on micro literature works using a linear model to retrieve associations with the estimated changes (Casarico and Lattanzio, 2020), allows us to show that more advanced and fully nonparametric statistical techniques do not only lead to more accurate impact estimates, but also to a more comprehensive assessment of territorial-level associations.

3.3. Comparative analysis: the COVID-19 crisis vs the Great Recession

Is the pandemic crisis different from previous recessions? To answer this question, we compare the estimated employment losses due to the COVID-19 shock with those of the 2008–2009 Great Recession. Specifically, we collect annual employment data from Istat for the period 2006–2019. These are yearly data that cover all public and private employment. Note that these data have a correlation of 98.9% with the Unioncamere employment data used for the main analysis during the period 2016–2019.

We could not use such data in the counterfactual analysis because they do not cover 2020 and do not allow the subsample split between manufacturing and services employment. Note that these data have a correlation of 98.9% with the Unioncamere employment data used for the main analysis during the period 2016–2019.

For Median Squared Error.

16 Tables A1 and A2 in the Appendix provide further details on the data used for this analysis. Finally, after estimating and mapping employment losses during the Great Recession, we compare their spatial distribution with that of the employment change registered during the COVID-19 crisis.

4. Counterfactual analysis

We begin by reporting in Table 1 the random forest technique’s predictive performance compared to a battery of alternative predictive methods. First, we compare it with the intuitive before-after method often adopted to gauge the magnitude of the COVID-19 shock. The before-after analysis estimates the impact of COVID-19 as the difference between the trend of a given outcome (in this case, employment) in 2020 (after the pandemic’s arrival) and the pre-pandemic average figures of the past year(s). The underlying assumption is that, without the pandemic, the number of employees would have remained constant (zero growth). Examples of this intuitive approach to gauge the magnitude of the COVID-19 impact on employment and firm outcomes in the Italian context can be found in Casarico and Lattanzio (2020), Giacomelli et al. (2021), and Viviano (2020). Next, we also assess the predictive power of a more traditional fixed-effects OLS model (fitted on the full 2016–2018 sample) and a simpler machine learning routine, the regression tree (here used following the same routine implemented for the random forest). In this way, we can get a full picture of the predictive ability of our methodology compared to other approaches. To assess the predictive performance of the various methods, we use two different measures of the typical prediction error, i.e. the Mean

| Table 1 | Predictive performances for 2019 (log) overall employment levels. |
|---------|---------------------------------------------------------------|
| Panel A – Performance on all LLMs | Predictive method | MSE | MEDSE |
| | Corresponding quarter – last year (2018) | 0.00101 | 0.00040 |
| | Corresponding quarter – 3-year average (2016–2018) | 0.00320 | 0.00214 |
| | OLS | 0.08510 | 0.04799 |
| | Regression tree | 0.05669 | 0.02907 |
| | Random forest | 0.00082 | 0.00026 |
| Panel B – Performance by population size | | | |
| ≤ 50,000 inhabitants | Corresponding quarter – last year (2018) | 0.00120 | 0.00042 |
| | Corresponding quarter – 3-year average (2016–2018) | 0.00325 | 0.00189 |
| | OLS | 0.08907 | 0.03839 |
| | Regression tree | 0.04337 | 0.02366 |
| | Random forest | 0.00107 | 0.00034 |
| Between 50,000 and 200,000 inhabitants | Corresponding quarter – last year (2018) | 0.00075 | 0.00037 |
| | Corresponding quarter – 3-year average (2016–2018) | 0.00312 | 0.00224 |
| | OLS | 0.06415 | 0.05289 |
| | Regression tree | 0.04382 | 0.02943 |
| | Random forest | 0.00052 | 0.00021 |
| ≥ 200,000 inhabitants | Corresponding quarter – last year (2018) | 0.00090 | 0.00042 |
| | Corresponding quarter – 3-year average (2016–2018) | 0.00324 | 0.00274 |
| | OLS | 0.20479 | 0.18258 |
| | Regression tree | 0.19768 | 0.08120 |
| | Random forest | 0.00056 | 0.00020 |

Notes: Estimates on the 2019 sample. MSE stands for Mean Squared Error; MEDSE for Median Squared Error.
Squared Error (MSE) and Median Squared Error (MEDSE). The figures reported in Table 1 reveal that random forest predictions substantially outperform all the other methodologies in the out-of-sample predictive test on the 2019 sample. Let us first focus on the mean predictive performance across the entire 2019 sample (Panel A). Using MSE as the reference metric, the predictive gain of the random forest performance is of about 19% compared to last year’s before-after figures, and of almost 75% compared to the three-year (2016–2018) average of the outcome variable. MEDSE performances are even more dramatically unbalanced in favor of the random forest. The random forest also strongly outperforms an OLS model run on the full 2016–2018 sample, with a reduction of the 2019 MSE larger than 90% compared to this traditional approach, which also performs far worse than both the before-after approaches and seems to be very sensitive to extreme values, in line with the results of Burlig et al. (2020). Finally, the regression tree predictive performance is in between, with MSE and MEDSE values lower than OLS but definitely larger than the before-after metrics and the random forest. This suggests that one needs to apply more powerful ensemble methods such as the random forest in order to maximize out-of-sample performance and unleash the full predictive potential of ML for counterfactual building. In addition, the relatively good performance of the before-after approach using the last pre-shock corresponding quarter’s average as counterfactual suggests that this straightforward methodology can be an acceptable empirical solution for short-term assessments during emergency situations.

Overall, the random forest outperforms all the other methods. This key insight is not just valid on average, but also for each of the three population subsamples in which we split our full sample (Panel B). The subsample comparison reveals that the random forest predictive gain becomes especially sizable when moving from less-populated LLMs (in which employment annual growth rates can vary greatly due to small employment levels) to bigger local economies, where the gap between this complex algorithm and all the other approaches becomes large. In sum, this test demonstrates that complex data-driven methodologies can lead to consistently more accurate predictions of potential outcomes in a given, ‘ordinary’ year.¹⁹

Having established that the random forest exploits past information to predict future outcomes much better than standard methods, we take a quick look at the aggregate treatment effects of the coronavirus crisis (weighted by LLM population) for the employment outcome. By the end of 2020, the pandemic has entailed a 3% decrease in overall employment in Italy, compared to what employment levels would have been had the pandemic never reached the country. This national-level estimate is larger than the 2.1% annual reduction in the number of persons employed in 2020 estimated by the Bank of Italy.²⁰ The reason for the discrepancy is that this institution measures average employment change with respect to 2019, which does not account for the potential growth in employment that could have been registered in 2020 in the absence of the shock.

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

Fig. 1 shows the map of employment change in 2020 Q4 at the LLM level. The degree of treatment effect heterogeneity is striking. At first glance, it is evident that the crisis hit more severely local economies located in the Centre-North (–3.7% vs –1.8% in the South). Nevertheless, many LLMs in Southern regions and in the islands have also been sharply affected. We also remark that, in the so-called Mezzogiorno, the weight of the informal economy is substantial, and we cannot capture those losses with official data. The North-South gap in employment losses might thus be smaller than it is apparent from this figure. Despite these caveats, the map looks different from what one would expect based on micro-level evidence (Casarico and Lattanzio, 2020; Carta and De Filippis, 2021), because temporary and less-educated workers with unstable and less protected jobs are more commonly employed in Southern regions, less developed than the richer North. Our finding that losses are more substantial in the affluent areas of the country is in line with the evidence provided by Chetty et al. (2020) for the US.

More generally, some local economies have been hit much harder than others, with impacts ranging from drops larger than 2.5% in most LLMs of the entire Centre-North, Abruzzi, Basilicata and the costs of Sardinia, to small decreases or even mildly positive effects in parts of Lazio, Campania, Calabria, Sicily, and central Sardinia. What is even more relevant than the macro-area patterns is, in our view, the within-region degree of heterogeneity, which shows how, in virtually all Italian regions, some LLMs fared much better than others despite being geographically close and often contiguous. Such local disparities would be averaged away if using regional data. Besides, it is worth noting that the documented employment impacts are net of the Italian government’s protective measures. This means that without these protective measures (the layoff freeze and CIGS extensions in particular), local impacts would have been more sizeable.

Where does such a striking heterogeneity come from? We first inspect the geographic distribution of the employment and epidemiological outcomes engendered by COVID-19. Fig. 2 presents a visual comparison between the economic vs epidemiological effects of COVID-19 in Italy. Looking at the maps, the geographic distribution of impacts does not mirror the COVID-19 epidemiological spread during the first wave, which is proxied by excess mortality estimates from February 21, 2020, to September 30, 2020. To estimate 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 (−0.092), which suggests a lack of significant spatial correlation between employment and epidemiological impacts.

4.2. Employment by sector

If LLMs’ COVID-19 death toll 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 Fig. 3, the tertiary sector was more severely affected than the manufacturing one.²¹ This is not unexpected, as workplace closures primarily affected 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 layoff freeze measure. Given these facets, it comes as no surprise that employment losses primarily affected LLMs

¹⁹ Albeit it is untestable due to the shock, the predictive gain of the random forest is likely to be even higher in the 2020 sample as the training set contains four more quarters of information.

²⁰ See here: https://www.bancaditalia.it/pubblicazioni/relazione-annuale/2020/rel_2020.pdf (in Italian).

²¹ This is confirmed by the population-weighted national-level estimates, which reveal an aggregate 2.1% decrease in manufacturing compared to a 3.6% decrease in services.
specialized in services, and tourism in particular. For instance, the case of Tuscany, home to many cities renowned for their art and culture, is particularly emblematic of the losses suffered in the tourism sector.

4.3. Business demography

We then look at how COVID-19 affected business demography outcomes. At the national level, by the end of 2020, the crisis determined a 19.5% decrease in business births and a 7.5% decrease in business deaths. Fig. 4 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) in 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. By contrast, the impact on firm exits is more polarized and geographically dispersed, with several LLMs experiencing substantial reductions in cessations of trading, whereas others saw a significant increase in firm exits.

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. 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 (Sedlářek, 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 and Mingardi, 2020). The results on firm closures, instead, should be interpreted with caution, as many firm exits could have been temporarily frozen by the supportive measures adopted by the government.

5. Association analysis

The counterfactual analysis revealed a substantial heterogeneity of the pandemic’s economic effects. Such heterogeneity is partly driven by the territorial sectoral specialization. Nevertheless, we want to go further than this and understand the factors that matter the most in generating such a fragmented landscape. Therefore, we use a regression tree to examine the main predictors of employment losses.

Fig. 5 illustrates the regression tree of the LLM-specific overall employment treatment effects. The tree reveals interesting patterns. First, only four variables generate the tree: the share of jobs having a high aggregation risk, the share of temporary contracts, the unemployment rate, and per capita income. Second, the most severely affected LLMs are those in which there is a substantial share of jobs at a high risk of social aggregation and 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 21% drop in employment. Third, excess deaths are not picked up by the tree, confirming that there is no discernible association between employment outcomes and the spread of COVID-19.

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., 2021). In turn, the relevance of the 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 emergency, in

---

22 Note that (cf. Table A2) the average LLM share of jobs having a high risk of social aggregation is 23% (with a 11% standard deviation) and the average share of temporary contracts is 19% (standard deviation = 8%).
Finally, local economies specialized in sectors with a high aggregate risk, due to their structural characteristics that made them less exposed to the protective measures adopted by the government without paying a high price for the measures. Rural areas of Calabria and Sicily, might have benefited from the protective measures adopted by the government without paying a high price due to their structural characteristics that made them less exposed to the crisis.23 Finally, local economies specialized in sectors with a high aggregation risk, lower share of temporary contracts, but higher per capita income also experienced more severe losses. Such insights add nuance to the micro-level evidence cited above and complement the key insights from Fig. 1 that the largest employment changes have been observed in the Centre-North.

As a sensitivity check, we replace our variables on the share of jobs having a high risk of social aggregation risks, the share of jobs having a high ‘integrated’ risk, and the share of jobs in suspended economic activities with alternative measures of the expected sectoral shocks: the observed demand- and supply-side changes. These two variables weigh the expected supply and demand sectoral shocks reported in del Rio-Chanona et al. (2020) by each LLM’s sectoral composition (see Tables A1 and A2 in the Appendix for definitions and descriptive statistics), and are highly correlated with the three replaced features.24 The corresponding regression tree is presented in Fig. A1. The tree confirms that the largest employment losses originate from severe demand shocks experienced in areas characterized by high risks of social aggregation and more vulnerable labor markets.

Lastly, Table A3 in the Appendix reports the results of the OLS-based association analysis run on the 1-year before-after treatment effects. As one can see, while a traditional model can detect the importance of the predominant predictors, as the coefficients of the key variables (social aggregation risks, the share of temporary contracts, per capita income, and the unemployment rate) are all strongly significant and negative (except for the unemployment rate, whose sign is positive), it is intrinsically unable to capture the relevant interactions among these features that are instead promptly detected by the tree. Consider, for instance, the role played by the share of temporary contracts. According to OLS, it is negatively associated with the treatment effects. But the regression tree suggests that, conditional on threshold values of other key features (risks of social aggregation and unemployment rate), this variable can be either negatively or positively associated with the employment losses. The switch depends on the way in which it combines with the other features. Based on the OLS association analysis, instead, one would conclude that all areas with a high share of temporary contracts experienced negative effects. Instead, the areas with the highest values of this variable, which are more isolated and with higher unemployment rates, saw slight increases in employment levels. The reason for this contradiction is that, unlike the regression tree, which is a fully nonparametric and flexible statistical learning method designed to automatically capture the interactions and non-linearities among features with relevant predictive power, OLS cannot capture the important interactions and non-linearities at play due to its parametric functional form restrictions. Lastly, OLS suggests that other variables, such as the share of jobs in suspended economic activities and the dependency ratio, are significantly associated with treatment effects, but these features do not appear at all in the tree. Such differences suggest that simple linear techniques provide only an incomplete assessment of the more significant territorial-level associations. In sum, the replacement of traditional methods with more advanced data-driven techniques leads not only to more accurate

---

23 Furthermore, these areas were depressed well before the pandemic and were experiencing decreasing employment trends in the pre-COVID years. Paradoxically, the pandemic might have reverted this trend.

24 The reason why we opt for a replacement, rather than an enrichment, of variables related to supply and demand shocks, is rooted in the interpretative perspective provided by Mullainathan and Spiess (2017): If variables are highly correlated with each other, then such variables are substitutes, rather than complements, in predicting the outcome of interest.

---

Fig. 4. Business births and deaths changes in 2020.
estimates of treatment effects, but also to qualitatively different conclusions regarding their most relevant territorial factors and, ultimately, to different policy prescriptions.

6. Comparative analysis

The foregoing has provided evidence about the magnitude, heterogeneity, and characteristics of the COVID-19 shock in Italy. But is this recession different from previous ones? Fig. 6 reports an LLM-level comparison between two maps, both estimated using a counterfactual generated through the random forest: the employment losses of the COVID-19 crisis, i.e. the same map of Fig. 1, and those of the Great Recession.

Three key insights stand out: i) in absolute terms, the magnitude of the national-level impact is stronger in the pandemic recession: −3% in 2020 vs −1.8% in 2008–2009, in line with the findings of Cajner et al. (2020) and Coibion et al. (2020) for the US; ii) the two crises hit different territories: the spatial correlation between the two sets of LLM-level estimates is −0.104, confirming the differences in the geographic patterns of employment losses during the two recessions documented by Chetty et al. (2020) for the US; iii) the macro-patterns are also clearly different: while Southern regions were the most severely affected during the Great Recession (−3.8%), the COVID-19 crisis has a more homogeneous impact across macro-areas and the largest employment losses are predominantly concentrated in the Centre-North.

Even though, due to the lack of granular data by sector, we cannot replicate the subsample analysis for manufacturing and services that we presented in Fig. 3, official data by the Bank of Italy report that the manufacturing sector was substantially more affected than the services sector during the Great Recession.25 In sum, while the Great Recession mainly hit the South and manufacturing activities, the pandemic recession triggered employment drops across the entire country and damaged the services sector more. The COVID-19 crisis is thus less similar to previous economic downturns than one would perhaps expect, not just in absolute terms but also concerning the most affected territories and sectors.

Fig. 5. Regression tree on 2020 employment change.

---

25 Cf. Table 9.1 of the 2009 Annual Report: https://www.bancaditalia.it/pubblicazioni/relazione-annuale/2009/rel09_totale.pdf (in Italian).
7. Conclusion

We have documented the striking local inequalities of the coronavirus crisis across the Italian territory. The largest employment losses are more concentrated in the Centre-North and associated with a combination of LLM-specific features such as sectoral specialization, exposure of economic activities to high social aggregation risks, and pre-existing labor market vulnerabilities. By contrast, there is no discernible spatial correlation between the economic and epidemiological patterns of the pandemic. Lastly, the territorial hotspots of the COVID-19 crisis do not overlap with those of the Great Recession.

These insights are provisional. First, the pandemic did not end in 2020. Second, as the exceptional protective labor market policies implemented by the government will be gradually lifted, the magnitude and spatial distribution of local impacts may change. More research is thus needed to monitor the evolution of local economies and provide rigorous, granular, and constantly updated empirical evidence. As it stands, we deem these local and spatial dimensions of the crisis to be policy-relevant, especially in light of the forthcoming resources from the NextGenerationEU initiative. There is growing evidence that the pandemic is increasing inequality at all levels, including territorial and regional divides (Stantcheva, 2021). Our work corroborates these findings. In this respect, the Italian case is emblematic, and the results we provide can be thoughtfully extended to similar dynamics that might be taking place in other countries too.

From a prescriptive viewpoint, our analysis calls for a place-based approach in the policy response to the crisis. As national policies and top-down plans will be insufficient to lead the recovery (Bailey et al., 2020), policymakers should not neglect the local evolution of this unprecedented shock. Therefore, such diverging trajectories emphasize the need for ad hoc policy interventions based on the territorial profile and sectoral specialization of local economic systems (Ascani et al., 2021). The place-based policy perspective we advocate is, in our view, the best possible approach to target the local hotspots of the COVID-19 crisis and prevent the pandemic from further exacerbating pre-existing territorial vulnerabilities.

Credit author statement

Augusto Cerqua: Conceptualization, Data curation, Validation, Visualization, Investigation, Writing—Reviewing and Editing, Funding acquisition, Marco Letta: Methodology, Software, Formal analysis, Writing—original draft preparation, Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to the Editor Stephen L. Ross and five anonymous referees for their valuable advice. We thank Guglielmo Barone and Charles Wyplosz for their comments and suggestions on earlier versions of this work. Finally, we thank seminar participants to the SIS, COMPIE, AISRe, and EALE 2021 annual conferences.
## Appendix A

### Table A1

Definition of the variables included in the empirical analysis.

| Variable name | Definition | Time period | Source |
|---------------|------------|-------------|--------|
| **Counterfactual analysis** | | | |
| Employment | Overall employment of private non-financial sector firms | 2014 Q3–2020 Q4 | Business Register |
| Employment in manufacturing | Overall manufacturing employment | 2014 Q3–2020 Q4 | Business Register |
| Employment in services | Overall services employment | 2014 Q3–2020 Q4 | Business Register |
| Business births | Companies that have registered in the period under review | 2014 Q1–2020 Q4 | Business Register |
| Business deaths | Companies that went out of business in the period under review | 2014 Q1–2020 Q4 | Business Register |
| Economic classification dummies | Without specialization, non-manufacturing (touristic), non-manufacturing (non-touristic), made in Italy, other manufacturing | | Istat |
| Geographical dummies | North-East, North-West, Centre, South | | Istat |
| Log of population | Log of the resident population | 2014–2019 | 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 |
| Share of foreign population | Foreigners/population | 2014–2019 | Istat |

| **Association analysis** | | | |
| Employment change in 2020 | Treatment effect of the COVID-19 crisis on overall employment levels | 2020 Q4 | Estimated via the MLCM |
| Unemployment rate | Resident population aged 15+ not in employment but currently available for work | 2019 | Istat |
| Excess mortality estimates | Municipality-level excess mortality estimated by applying ML techniques to all-cause death data, aggregated at the LLM level | From Feb 21, 2020 to Sep 30, 2020 | Cerqua et al. (2021), |
| 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 |
| 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 temporary contracts | Number of employees with temporary 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 |
| Per capita income | The amount of money earned per person | 2019 | 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) | Ministry of Economy and Finance |
| 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 (IRFL) | The percentage of flows within an 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 |
| Number of hospital beds per 1000 inhabitants | Number of hospital beds divided by resident population * 1000 | 2018 | Ministry of Health |
| Share of workers employed in health care occupations | Share of jobs in the NACE 2-digit sector ‘human health activities’ and ‘residential care activities’ | 2019 | Own calculations using Business Register data |
| Supply-side changes | Supply-side changes due to the closure of non-essential industries and workers not able to perform their activities at home | 2019 | Own calculations using forecasts by del Rio-Chanona et al. (2020) |
| Demand-side changes | Demand-side changes due to people's immediate response to the pandemic, such as reduced demand for goods or services that are likely to place people at risk of infection | 2019 | Own calculations using forecasts by del Rio-Chanona et al. (2020) |

| **Great Recession analysis** | | | |
| Employment | Overall employment, including also all public employees | 2006–2009 | Istat |
| Economic classification dummies | Without specialization, non-manufacturing (touristic), non-manufacturing (non-touristic), made in Italy, other manufacturing | | Istat |
| Geographical dummies | North-East, North-West, Centre, South | | Istat |
| Business births | Companies that have registered in the period under review | 2005–2007 | Business Register |
| Business deaths | Companies that went out of business in the period under review | 2005–2007 | Business Register |
| Population density | Resident population per unit area | 2005–2007 | Istat |
| Per capita income | The amount of money earned per person | 2005–2007 | Ministry of Economy and Finance |

Notes: Outcome variables in bold.
Table A2
Descriptive statistics. Association analysis.

| Variable name | Mean  | SD    | Min   | Max    |
|---------------|-------|-------|-------|--------|
| **Counterfactual analysis** |       |       |       |        |
| Employment (log) | 9.30  | 1.25  | 5.95  | 14.41  |
| Employment in manufacturing (log) | 7.52  | 1.61  | 3.50  | 12.65  |
| Employment in services (log) | 8.88  | 1.28  | 5.51  | 14.22  |
| Business births | 57.96 | 243.92 | 1 | 5173 |
| Business deaths | 46.72 | 204.10 | 1 | 9685 |
| Share of LLMs without specialization | 0.19  | 0.39  | 0     | 1      |
| Share of non-manufacturing (non-touristic) LLMs | 0.14  | 0.34  | 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 (%) | 48.10 | 6.67  | 29.79 | 63.91  |
| Unemployment rate (%) | 12.14 | 6.19  | 1.19  | 39.08  |
| Population (log) | 10.71 | 1.13  | 8.05  | 15.18  |
| Population density | 0.21  | 0.30  | 0.01  | 3.17   |
| Share of foreign population | 0.07  | 0.04  | 0.00  | 0.18   |
| **Association analysis** |       |       |       |        |
| Employment change in 2020 (%) | -3.47 | 4.20  | -26.90 | 9.71   |
| Unemployment rate (%) | 10.99 | 5.91  | 1.19  | 36.19  |
| Excess mortality estimates (%) | 7.99  | 19.72 | -34.30 | 148.07 |
| Share of jobs in suspended economic activities | 0.47  | 0.08  | 0.25  | 0.79   |
| Per capita income (€) | 16559 | 4109  | 8050  | 27664  |
| 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 temporary 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 (JIRFL) | 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 hospital beds per 1000 inhabitants | 2.43  | 3.16  | 0    | 24.27  |
| Share of workers employed in health care occupations | 0.0253 | 0.0265 | 0 | 0.3530 |
| Supply-side changes (used in the sensitivity check) | -0.27 | 0.06  | -0.51 | -0.10  |
| Demand-side changes (used in the sensitivity check) | -0.21 | 0.08  | -0.08 | -0.61  |
| **Great Recession analysis** |       |       |       |        |
| Employment (log) – Istat data | 9.71  | 1.17  | 7.21  | 14.31  |
| Share of LLMs without specialization | 0.19  | 0.39  | 0    | 1      |
| Share of non-manufacturing (non-touristic) LLMs | 0.14  | 0.34  | 0    | 1      |
| Share of made in Italy LLMs | 0.23  | 0.42  | 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] | 0.45  | 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      |
| Population density | 0.20  | 0.29  | 0.01  | 3.12   |
| Per capita income (€) | 15408 | 3136  | 8757  | 2572   |
| **Number of LLM-quarters** | 10,980 |       |       |        |
| **Number of LLM-years** | 610   |       |       |        |

A. Cerqua, M. Letta Regional Science and Urban Economics 92 (2022) 103752
Table A3
OLS-based association analysis

| Feature                                      | Coefficient | Standard error | p-value     |
|----------------------------------------------|-------------|----------------|-------------|
| Unemployment rate                            | 12.525      | 3.498          | 0.000370    |
| Excess mortality estimates                   | 0.00901     | 0.00647        | 0.165       |
| Share of jobs in suspended economic activities| –6.352      | 1.577          | 0.0000639   |
| Per capita income                            | –0.000250   | 0.0000623      | 0.0000684   |
| Share of firms having employees in CIGS      | –107.763    | 208.675        | 0.606       |
| Share of population living in peripheral areas| 0.464       | 0.367          | 0.206       |
| Share of temporary contracts                 | –13.783     | 2.160          | 0.000       |
| Number of road accidents per 10,000 inhabitants| 0.0859      | 0.131          | 0.511       |
| Index of relational intensity                | –0.00919    | 0.0106         | 0.385       |
| Dependency ratio                             | –5.990      | 2.542          | 0.0188      |
| Share of innovative start-ups                | 63.971      | 45.158         | 0.157       |
| Number of hospital beds per 1,000 inhabitants| –0.0108     | 0.0437         | 0.805       |
| Share of workers employed in health care occupations| 14.931      | 8.177          | 0.0684      |
| Share of jobs having a high risk of social aggregation| –19.656      | 1.236          | 0.000       |
| Share of jobs having a high integrated risk  | –4.517      | 7.213          | 0.531       |

Notes: The outcome variable is 2020 employment change estimated with the 1-year before-after methodology. Intercept not reported.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.regsciurbeco.2021.103752.

References

Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: estimating the effect of California’s tobacco control program. J. Am. Stat. Assoc. 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.

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. J. Econ. Behav. Org. 156, 86–102.

Ascani, A., Faggian, A., Montresor, S., 2021. The geography of COVID-19 and the structure of local economies: the case of Italy. J. Reg. Sci. 61 (2), 407-441.
Athey, S., Imbens, G., 2016. Recursive partitioning for heterogeneous causal effects. Proc. Natl. Acad. Sci. Unit. States Am. 113 (27), 7353–7360.
Athey, S., Bayati, M., Imbens, G., Qu, Z., 2019. Ensemble methods for causal effects in panel data settings. AE AJ Papers and Proceedings 109, 65–70.
Athey, S., Bayati, M., Doudchenko, N., Imbens, G., Kborravi, K., 2021. Matrix completion methods for causal panel data models. J. Am. Stat. Assoc. 1–15.
Bailey, D., Clark, J., Colombelli, A., Corradini, C., De Propris, L., Derudder, B., Kemeny, T., 2020. Regions in a time of pandemic. Reg. Stud. 54 (9), 1163–1174.
Baker, S.R., Farrokhnia, R.A., 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.
Bartik, A.W., Bertrand, M., Cullen, Z., Glaser, E.L., Luca, M., Stanton, C., 2020. The impact of COVID-19 on small business outcomes and expectations. Proc. Natl. Acad. Sci. Unit. States Am. 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., de Villemeur, E.B., 2019. Strategic reneging in sequential imperfect markets. CREST Working Papers No. 19.
Benedetti, F.C., Sedlček, 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. IVES Discussion Paper No. 82.
Bloom, N., Fletcher, R.S., Yeh, E., 2021. The Impact of COVID-19 on US Firms. National Bureau of Economic Research No. p. w28314.
Blundell, R., Costa Dias, M., Joyce, R., Xu, X., 2020. COVID-19 and inequalities. J. Econ. Perspect. 34 (2), 291–319.
Burlig, F., Knittel, C.R., Rapson, D., Reguant, M., Rodano, G., 2021. Fallimenti d’impresa in Epoca Covid. Note Covid-19. Bank of Italy, Giupponi, G., Landais, C., 2020. Subsidizing Labor Hoarding in Recessions: the Employment & Welfare Effects of Short Time Work. CEPR Discussion Papers No. 13310, May 2020 version.
Cerqua, A., Di Stefano, R., Letta, M., Miccoli, S., 2021. Predictive counterfactuals for treatment effect heterogeneity in event studies with staggered adoption. SSRN Electronic Journal 3484635.
Chetty, R., Friedman, J.N., 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 (NBER) Working Paper No w27431.
Chudik, A., Mohaddes, K., Pesaran, M.H., Raisi, M., Rebecchi, A., 2021. A counterfactual economic analysis of Covid-19 using a threshold augmented multi-country model. J. Int. Finance 119, 102977.
Coibion, O., Gorodnichenko, Y., Weber, M., 2020. Labor Markets during the COVID-19 Crisis: A Preliminary View. National Bureau of Economic Research (NBER) Working. Paper No.27017.
del Rio-Chanona, R.M., Mealy, P., Pichler, A., Lafond, F., Farmer, D., 2020. Supply and demand shocks in the COVID-19 pandemic: an industry and occupation perspective. Oxf. Rev. Econ. Pol. 36 (Issue Suppl. 1), S94-S137.
Forsythe, E., Kahn, L.B., Lange, F., Wiczer, D., 2020. Labor demand in the time of COVID-19: evidence from vacancy postings and UI claims. J. Publ. Econ. 189, 104236.
Giacomelli, S., Mocetti, S., Rodano, G., 2021. Fallimenti d’impresa in Epoca Covid. Note Covid-19. Bank of Italy, Giupponi, G., Landais, C., 2020. Subsidizing Labor Hoarding in Recessions: the Employment & Welfare Effects of Short Time Work. CEPR Discussion Papers No. 13310, May 2020 version.
Gourinchas, P.O., Kalemli-Ozcan, S., Penciakova, V., Sander, N., 2020. Covid-19 and SME Failures. National Bureau of Economic Research No. p. w27877.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
Hofman, J.M., Watts, D.J., Athey, S., Garip, F., Griffiths, T.L., Kleinhag, J., et al., 2021. Integrating explanation and prediction in computational social science. Nature 595 (7866), 181–188.
Howard, J., Bowles, M., 2012. The two most important algorithms in predictive modeling today. In: Strata Conference Presentation, February, vol. 28.
Kleinhag, J., Ludwig, J., Mullainathan, S., Obermeyer, Z., 2015. Prediction policy problems. Am. Econ. Rev. 105 (5), 491–495.
Krueger, D., Ublig, H., Xie, T., 2020. Macroeconomic Dynamics and Reallocation in an Epidemic. National Bureau of Economic Research No. p. w27047.
Lanz, B., 2019. Machine learning with R: Expert Techniques for Predictive Modeling. Packt Publishing Ltd.
Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. J. Econ. Perspect. 31 (2), 87–106.
Rossi, N., Mingardi, A., 2020. Italy and COVID-19: winning the war, losing the peace? Economia 40 (2), 148–174.
Sedlček, P., 2020. Lost generations of youth: Evidence from administrative data. Covid Economics 52, 152.
Varian, H.R., 2014. Big data: new tricks for econometrics. J. Econ. Perspect. 28 (2), 3–28.
Varian, H.R., 2016. Causal inference in economics and marketing. Proc. Natl. Acad. Sci. Unit. States Am. 113 (27), 7310–7315.
Viviano, E., 2020. Alcune stime preliminari degli effetti delle misure di sostegno sul mercato del lavoro. Note Covid-19. Bank of Italy.
Von Gaudecker, H.M., Holler, R., Janys, L., Siffinger, 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. p. 13158.
Wagner, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. J. Am. Stat. Assoc. 113 (525), 1228–1242.