



Attività essenziali, lockdown e contenimento della pandemia da COVID-19

Direzione Centrale Studi e Ricerche (DCSR) – INPS

Sommario

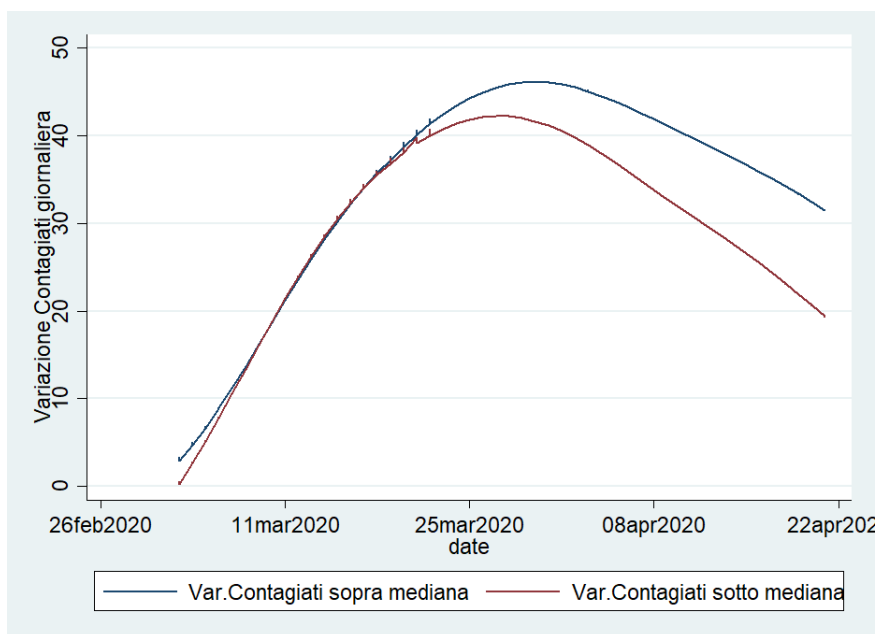
In questo documento si prendono in considerazione i dati provinciali della protezione civile sulla dinamica dei contagiati Covid19 e i dati amministrativi di fonte Inps relativi alla quota dei rapporti di lavoro nei settori essenziali a seguito dei due provvedimenti restrittivi del marzo 2020 emessi dal governo Italiano. Si vuole verificare l'ipotesi che, a seguito dei provvedimenti governativi, le province con una maggiore quota di rapporti di lavoro nei settori essenziali, cioè che non sono stati bloccati, hanno registrato una crescita superiore del numero dei contagiati. La figura riportata in questo sommario fornisce una statistica descrittiva inerente la rappresentazione grafica dei nostri risultati. Nella figura si mostrano due curve, quella dell'evoluzione dei contagiati nel gruppo di province al di sopra della mediana della distribuzione dei settori essenziali (le province con il 50% più elevato di incidenza dei settori essenziali) e quella di province al di sotto della mediana (il 50% con valori più bassi di settori essenziali)¹. Si evince che nel primo periodo le due curve crescono nello stesso identico modo, prima dei provvedimenti governativi, ma che dal periodo che comincia intorno al 22 marzo, con il provvedimento più restrittivo di lockdown, le due curve cominciano ad allontanarsi, con la curva delle province con settori essenziali sopra la mediana che cresce di più (fino ad inizio aprile) per poi decrescere più lentamente delle altre. Negli ultimi giorni del campione la differenza fra le due curve mostra che nei settori sopra la mediana dei settori essenziali vi sono in media 10 contagiati in più, un numero non trascurabile dato che la media della variabile dei contagiati dopo il 22 marzo è di 37 (l'impatto è circa il 25% della media).

Nello studio viene svolta un'analisi rigorosa dal punto di vista econometrico per verificare questa prima evidenza descrittiva, applicando il metodo delle differenze nelle differenze. I risultati descrittivi sono ampiamente confermati, anche applicando diverse specificazioni del modello. In particolare, si evince che all'aumentare di 1 punto percentuale della quota di settori essenziali in

¹ Il grafico riporta delle *locally weighted regression* come strumento per sintetizzare l'informazione statistica.

una provincia il numero di contagiati aumenta di 1.5 unità al giorno. Un altro modo per fornire una dimensione dell'impatto è che la differenza fra una provincia al 75esimo percentile della distribuzione della quota di essenziali e una al 25esimo percentile è di circa 10 contagiati al giorno.

Inoltre, si pone a verifica l'ipotesi che oltre all'incidenza dei settori essenziali possa giocare un ruolo rilevante la densità occupazionale a livello provinciale. Aggiungendo tale dimensione di analisi si deriva che l'impatto dell'esposizione a settori essenziali risulta essere leggermente più elevato rispetto al caso di considerare solo i settori essenziali (13 vs 10 contagiati in più al giorno), e maggiormente significativo da un punto di vista statistico.



Introduzione

Il contenimento dell'epidemia in Italia è stato conseguito soprattutto attraverso le misure di lockdown. Le misure di chiusura hanno in particolare riguardato due date: il decreto emanato il 9 marzo (con alcune estensioni dell'11 marzo), che limitavano il commercio all'ingrosso e al dettaglio, prevedendo la chiusura di bar, ristoranti e palestre e quello del 23 marzo che ha bloccato alcune attività economiche indicando un limitato numero di attività consentite, le cosiddette attività essenziali.

Le attività essenziali non sono distribuite in maniera omogenea sul territorio nazionale, e pertanto in alcune province, che saranno la nostra unità elementare di analisi, il numero di persone che sono venute in contatto, dopo il lockdown, tramite l'apertura delle attività essenziali sarà verosimilmente più grande che in altre. Ci si può di fatto aspettare che le misure di contenimento siano meno efficaci se l'economia è più attiva. Testare tale ipotesi non è banale. Tuttavia le tempistiche delle misure del



governo in questo specifico contesto generano una situazione favorevole alla ricerca sociale in quanto può essere analizzata come un esperimento naturale che ha come finalità di verificare, attraverso l'utilizzo di dati amministrativi, se le province con un maggior numero di attività essenziali abbiano avuto un trend di contagi differente rispetto alle altre dal momento dell'inizio del lockdown.

Verificare tale assunto è importante soprattutto ora che il numero dei contagi sta scendendo e che il governo si prepara a varare una riapertura delle attività, sotto il rischio di generare una seconda ondata di contagi.

Altri studi hanno provato ad analizzare gli effetti del lockdown, con risultati contrastanti. In un primo lavoro

(https://www.unive.it/pag/14024/?tx_news_pi1%5Bnews%5D=8829&cHash=d96192d9853e97c7b6b6e606e10b8ca4) si suggerisce che il primo decreto abbia rallentato il progredire dei contagi, mentre un altro lavoro (<https://www.lavoce.info/archives/65771/lockdown-delleconomia-un-primo-bilancio>) mostra che il decreto del 22 marzo ha contribuito alla decrescita ma in modo non significativo². Tuttavia, gli autori suggeriscono che diversi problemi legati alla qualità dei dati potrebbero aver inficiato le loro stime che mostrano comunque un trend di decrescita associato al secondo decreto.

In questa nota utilizziamo una strategia classica di differenze nelle differenze su un panel di 106 province italiane distribuito su 58 giorni dall'inizio dell'epidemia, 24 febbraio, fino al 21 Aprile 2020. I nostri risultati mostrano come l'eterogenea distribuzione delle attività essenziali sulla penisola abbia rallentato la decrescita dei contagi in quelle province con una quota di attività essenziali superiore alla mediana. Tale rallentamento è statisticamente significativo. Utilizzando diverse specificazioni del nostro modello di differenze nelle differenze verifichiamo che la nostra analisi è robusta, mostrando che le assunzioni alla base del modello di stima utilizzato sono valide. In particolare mostriamo che il trend dei contagi prima delle misure di contenimento è simile tra le varie province caratterizzati da diversi livelli di esposizione alle attività essenziali.

Ulteriori specificazioni mostrano come la densità delle attività essenziali sul territorio della provincia impatti in maniera significativa sul rallentamento del trend dei contagi, suggerendo che non solo tenere aperte le attività aumenta il rischio di diffusione del virus ma che tale aumento è più pronunciato quando le attività sono molto concentrate sul territorio provinciale.

² I nostri risultati differiscono da quelli di questo lavoro, ma ciò non deve sorprendere visto che vi sono differenze fra le due analisi: l'intervallo temporale (il nostro campione si estende fino al 21 aprile); la variabile inerente la quota di settori essenziali è determinata in modo diverso (in questo lavoro è calcolata sull'universo dei rapporti di lavoro Uniemens a 6 digit per i dipendenti privati); la specificazione econometrica è diversa, dato che in questo lavoro si utilizza la tecnica del metodo delle differenze nelle differenze.



I dati utilizzati

Abbiamo utilizzato dati provinciali di fonte Protezione Civile sulla progressione giornaliera dei contagi. Per contagiati la protezione civile intende il numero di nuovi positivi (a seguito dei nuovi tamponi e test) più il numero di decessi e dei malati guariti. Il numero dei contagiati pertanto è una cumulata nel tempo in ogni provincia. Purtroppo a livello provinciale la Protezione Civile non fornisce dati su tamponi, dimessi, guariti e deceduti disponibili solo a livello regionale, perciò non è possibile calcolare il saldo degli attualmente positivi ad un giorno specifico. Per tali variabili utilizziamo la loro release regionale in alcuni controlli di robustezza. Per ovviare all'informazione poco utile fornita a livello provinciale, nelle nostre stime abbiamo utilizzato come variabile dipendente la differenza tra due giorni consecutivi nel numero dei contagiati. In questo modo otteniamo una misura su quanti nuovi contagiati rispetto al giorno precedente vengono individuati a livello provinciale. In generale il trend dei contagi a livello nazionale è sempre in crescita e perde accelerazione negli ultimi giorni fino quasi ad appiattirsi.

Per i dati sulle attività essenziali abbiamo utilizzato dati di fonte INPS. La quota di attività essenziali è stata costruita a partire dai dati sui rapporti di lavoro dei dipendenti annuali di fonte Uniemens per il settore privato extra-agricolo. Una nota della DCSR riporta in maniera specifica come tale variabile sia costruita e quale sia la sua distribuzione geografica³. Per alcune specificazioni si utilizzerà anche una variabile legata alla densità dell'occupazione, calcolata come il rapporto fra il numero dei rapporti di lavoro a livello provinciale derivato dalla banca dati Uniemens e l'estensione territoriale delle province italiane di fonte Istat.

La tabella 1 riporta alcune statistiche descrittive per il nostro campione. La variazione giornaliera di contagiati è di 33.62, e la mediana 12. Ovviamente tale variabile ha una dinamica crescente nel tempo, con una media precedente all'8 marzo di 5 e successiva all'8 marzo di 37. Per quanto riguarda la variabile quota di rapporti di lavoro essenziali la media e la mediana sono del 48%, il 75esimo percentile del 52% e il 25esimo del 45%⁴.

³ Si veda il documento "Settori essenziali vs settori bloccati per la crisi pandemica: un'analisi dei rapporti di lavoro Uniemens", nel sito Inps Studi e Analisi:

https://www.inps.it/docallegatiNP/Mig/Allegati/NOTA_Settore_Bloccati_DCSR_INPS_final3.pdf

⁴ Si noti che rispetto alla nota precedente pubblicata dalla DCSR-INPS sui settori essenziali l'incidenza dei settori essenziali è leggermente inferiore, 48% contro circa il 50%. Ciò è dovuto al fatto che in questo contributo l'unità di osservazione è la provincia, e pertanto si calcola una media fra province. Nella nota precedente l'unità di osservazione era il rapporto di lavoro, e quindi si calcolava sul relativo universo.



Tabella 1. Descrittive statistiche della variazione dei contagiati e della quota di rapporti di lavoro essenziali. Dati provinciali.

	variazione contagiati	quota essenziali
Media	33,62	0,48
Mediana	12	0,48
p10	0	0,42
p25	3	0,45
p75	37	0,52
p90	83	0,55
Deviazione standard	64,83	0,05
Minimo	-186	0,29
Massimo	868	0,61

Legenda. Per la variabile variazione contagiati le descrittive sono calcolate dal primo giorno di contagio per ogni provincia. Esistono dei valori negativi della variabile variazione contagiati, probabilmente dovuti a revisioni ai dati poste in essere dalla Protezione Civile.

La strategia di identificazione

Per la nostra analisi specifichiamo un modello di differenze nelle differenze standard nella letteratura econometrica. In particolare, stimiamo:

$$\Delta Y_{p,t} = \alpha + \mathbf{G}_t + \delta * (QE_p * \mathbf{G}_t) + X_{r,t} + \eta_p + trend + \varepsilon_{p,t} \quad [1]$$

dove $\Delta Y_{p,t}$ è la nostra variabile dipendente calcolata come già illustrato nella sezione dati, la quale varia per ognuna delle 106 province p e per ognuno dei 58 giorni t analizzati. \mathbf{G}_t è una matrice composta da vettori contenenti variabili dummy, una per ognuno dei giorni della nostra analisi⁵. Il termine $QE_p * \mathbf{G}_t$ rappresenta una serie di 58 variabili dummy create dall'interazione tra quota di essenziali e i giorni. Il modello si chiude con effetti fissi provinciali η_p , il termine costante α e un termine di errore $\varepsilon_{p,t}$, stimato tramite una matrice robusta all'eteroschedasticità. In alcune specificazioni abbiamo inserito $X_{p,t}$, una serie di variabili di controllo rappresentate qui in forma matriciale che variano a livello regionale come il numero di tamponi, dei dimessi guariti e dei decessi. Inoltre, per tener conto del fatto che per ogni provincia può cambiare il giorno nel quale si è registrato il primo caso di contagio, si costruisce una variabile trend che prende il valore zero per tutti i giorni per i quali non vi sono casi di contagio nella provincia, e poi tale indicatore cresce di una unità dal primo giorno di contagio all'ultimo giorno disponibile nei dati (21 aprile). Nelle specificazioni dei modelli [1] e [2] tale trend è introdotto con un polinomio di quarto grado.

Vogliamo valutare l'impatto differenziale in termine di trend dei contagi dovuto alle misure di lockdown dopo il 9 e il 23 Marzo. Il modello [1] consente di stimare una serie di coefficienti δ , uno

⁵ Per ragioni statistiche di identificazione del modello, e come è prassi fare in questi casi, viene selezionata una categoria omessa, la dummy che identifica il giorno 8 di Marzo, e pertanto tutti i coefficienti possono essere interpretati relativamente a tale categoria. Si noti inoltre che essendo presenti nel modello effetti fissi provinciali non è possibile stimare l'effetto base della quota delle attività essenziali (che appunto non varia nel tempo per provincia).

per ognuno dei 58 giorni analizzati e questi identificano proprio il contributo marginale dovuto ad un aumento della quota di attività essenziali sull'andamento giornaliero della variabile dipendente. Abbiamo poi stimato con una seconda specificazione molto simile alla prima attraverso il seguente modello:

$$\Delta Y_{p,t} = \alpha + G_t + \delta * (QED_p * G_t) + X_{p,t} + \eta + \varepsilon_{p,t} \quad [2]$$

La differenza tra il modello [1] e [2] è nella variabile QED_p che è una dummy uguale ad 1 se la quota delle attività essenziali nella provincia supera la mediana. In questo secondo modello i valori stimati dei coefficienti δ rappresentano la differenza in media per ogni giorno che esiste tra province con quota di lavoratori essenziali sopra la mediana rispetto a quelle sotto la mediana.

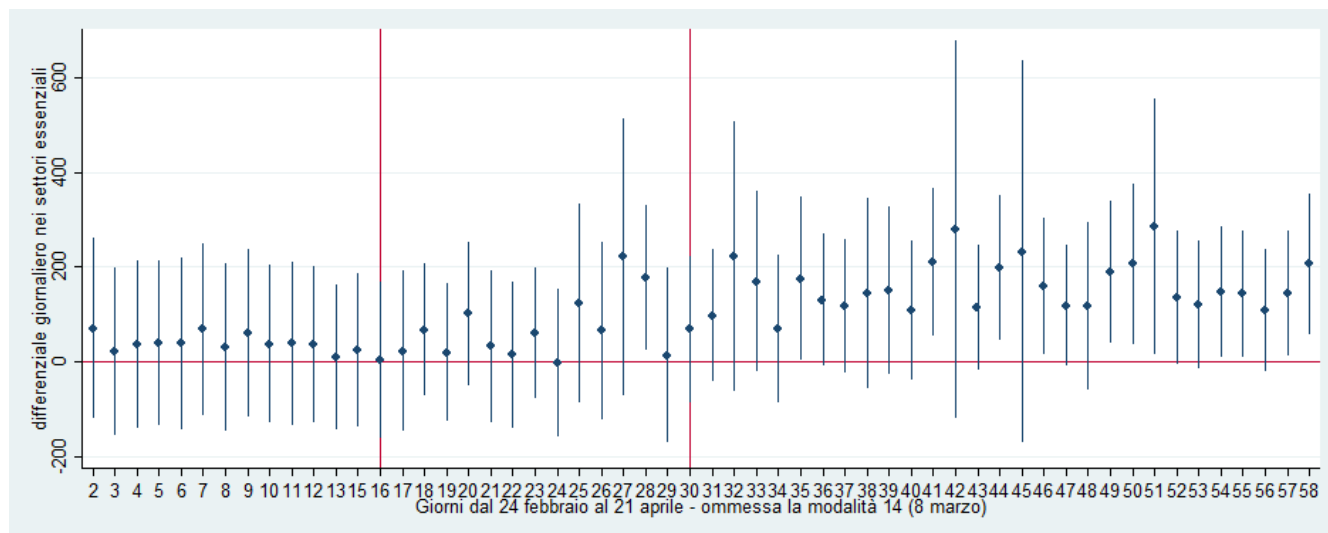
I risultati dell'impatto dei settori essenziali sul numero dei contagiati

La figura 1 mostra i coefficienti delle interazioni fra la quota di rapporti di lavoro essenziali e i giorni dal 24 febbraio al 21 aprile⁶. Le linee rosse riportate nel grafico fanno riferimento al 9 marzo (la prima a sinistra) e al 22 marzo, le due date di riferimento per i due provvedimenti di lockdown. Tali coefficienti vanno pertanto interpretati come un differenziale rispetto alle dummy di giorno incluse nella regressione (e non riportate nel grafico). Si evince come nel periodo precedente all'8 marzo, cioè prima delle prime decisioni di chiusura, i coefficienti siano praticamente nulli e largamente non significativi, e ciò indica la presenza dei cosiddetti 'pretrend', cioè i due gruppi, essenziali verso bloccati, non giocano alcun ruolo nell'andamento del contagio. Tra l'8 marzo e il 22 marzo i coefficienti cominciano a essere positivi, anche se fino al 22 marzo rimangono nella maggior parte dei casi non statisticamente significativi. Dopo il 22 marzo invece i coefficienti diventano positivi e nella maggior parte dei casi significativi, suggerendo che l'impatto differenziale di essere in una provincia con una quota più elevata di settori essenziali determina un aumento dei contagiati. In particolare, aumentare di un punto percentuale la quota di settori essenziali ha un impatto di circa 1,5 contagiati in più al giorno (visto che la media dei coefficienti dopo il 22 marzo è di circa 150). Un altro modo per quantificare tale impatto, e che tiene conto della distribuzione della variabile esplicativa (la quota dei settori essenziali), consiste nel considerare cosa succede se andassimo a comparare una provincia al 75esimo percentile della distribuzione dei servizi essenziali con una provincia al 25esimo percentile. La differenza fra i percentili 75 e 25 della quota di lavori essenziali è di circa 7 punti percentuali, come mostrato nella tabella 1. Dato che come già evidenziato la media dei coefficienti dopo il 22 marzo è circa 150, si può calcolare un impatto medio di essere in una

⁶ Le barre in corrispondenza di ogni punto per ogni giorno rappresentano le deviazioni standard e permettono di comprendere se il coefficiente per quel giorno è statisticamente diverso da zero. Inoltre, essendo la variabile dipendente in differenze non ne è possibile calcolare il valore per il primo giorno, che quindi viene omesso dalla regressione e dai grafici.

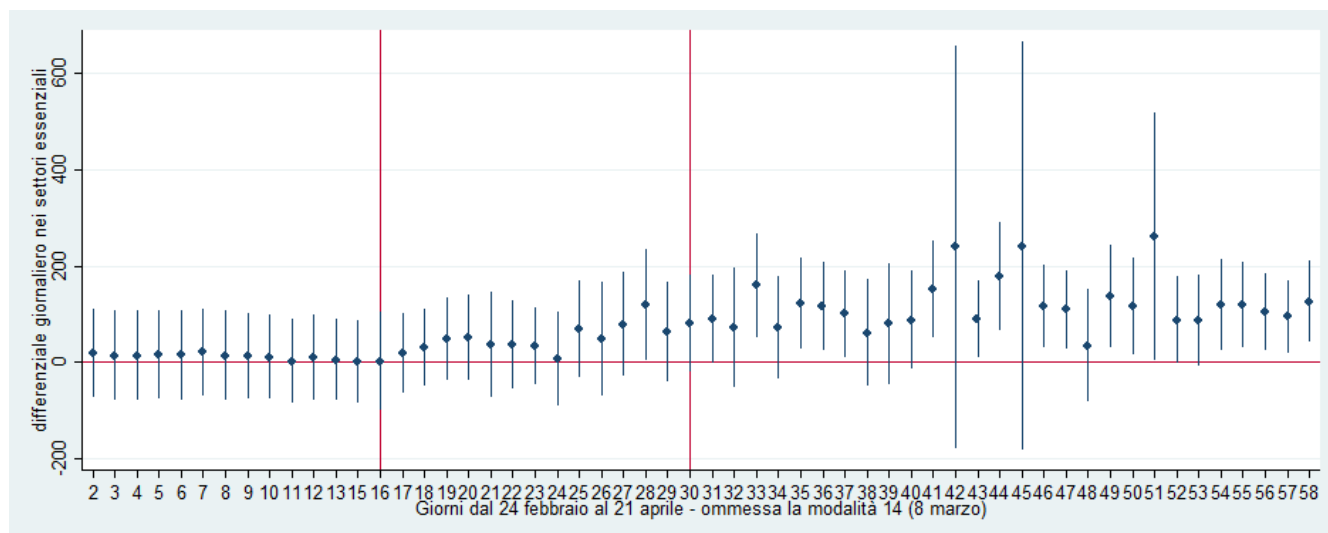
provincia al 75 percentile (ad esempio Cremona) verso una al 25 percentile (ad esempio Ferrara) uguale a circa 10 contagiati in più al giorno (circa il 27% in più rispetto alla media della variabile dipendente, la variazione dei contagi, dopo il 22 marzo). L'impatto non è pertanto trascurabile, e va nella direzione attesa di un aumento dei contagiati nelle province maggiormente esposte ad attività essenziali.

Figura 1. Impatto differenziale della quota di settori essenziali come variabile continua (per provincia per giorno).



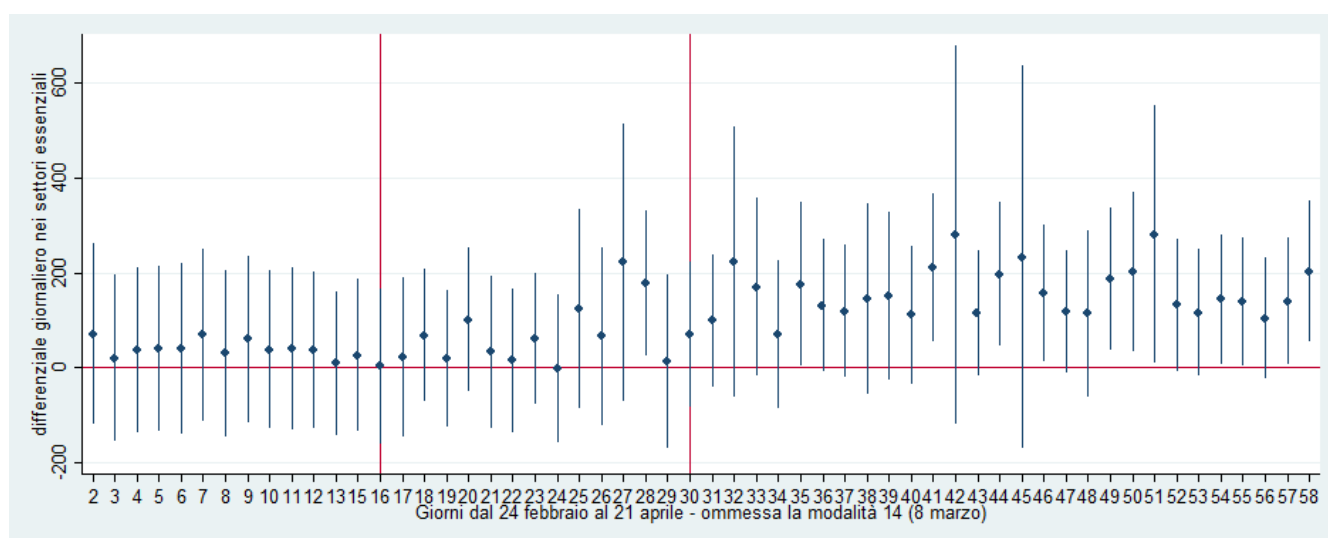
Abbiamo implementato vari test di robustezza ai nostri risultati. Un primo check consiste nell'eliminare le province lombarde che a detta di molti hanno rappresentato dei dati outlier all'interno della dinamica dei contagi (ed è stata anche l'unica regione che ha registrato una saturazione delle terapie intensive). La figura 2 mostra come l'andamento dei coefficienti è molto simile al precedente, e che pertanto l'esclusione della Lombardia non incide sulle conclusioni.

Figura 2: Impatto differenziale della quota di settori essenziali come variabile continua (per provincia per giorno), eliminando le province lombarde.



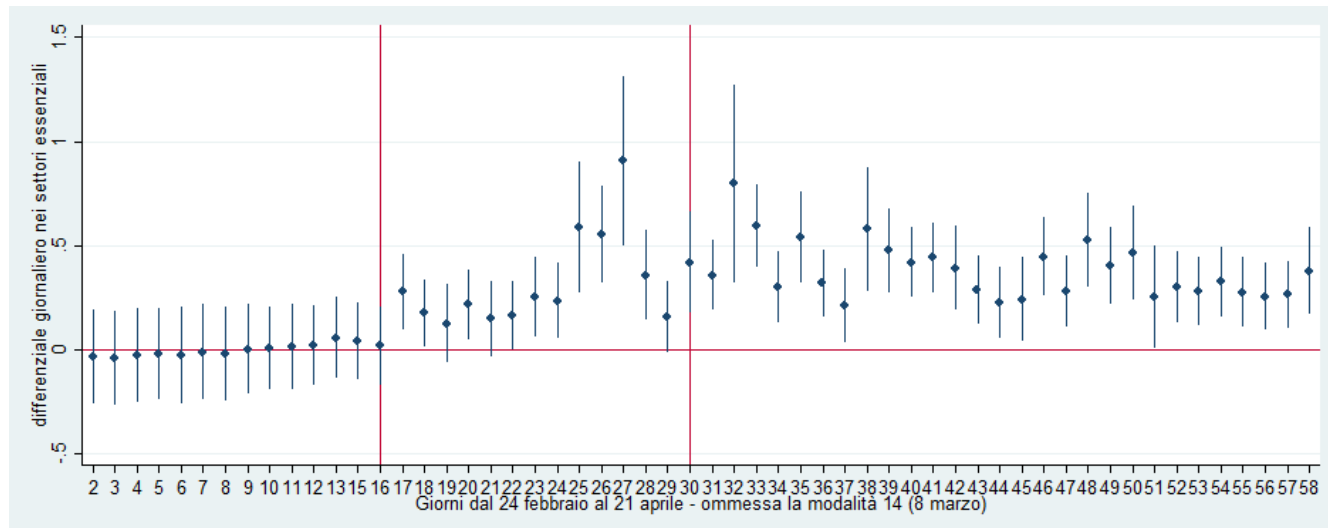
Un altro test di robustezza che abbiamo portato a termine è quello di inserire alcune variabili di controllo che variano giornalmente ma definite a livello regionale, e che potrebbero essere collegate alla diffusione del contagio, come il numero di tamponi effettuati (che ha un impatto sul numero di nuovi positivi), dei decessi (influenzati dalla saturazione delle terapie intensive) e dei dimessi guariti. La figura 3 mostra come anche inserendo tali controlli l'andamento rimane simile alle altre figure riportate.

Figura 3: Impatto differenziale della quota di settori essenziali come variabile continua (per provincia per giorno), con controlli regionali (numero di tamponi, decessi e guariti)



Un ulteriore approfondimento prende in considerazione la densità occupazionale delle attività nella provincia. E' plausibile sostenere che in province dove la densità per km² è più elevata, e quindi dove è più probabile una maggiore vicinanza dei lavoratori e prossimità anche per quanto riguarda i mezzi di trasporto, si ha una maggiore diffusione del virus. Al fine di verificare questa ipotesi di ricerca interagiamo la quota di lavoratori essenziali con la densità occupazionale per provincia (calcolata come il rapporto tra i lavoratori dipendenti Uniemens e la superficie in km² della provincia). Una volta derivata questa variabile continua applichiamo la stessa specificazione econometrica, nella quale interagiamo le dummy giornaliere con tale variabile continua. La figura 4 riporta un andamento dei coefficienti simile a quello delle figure precedenti, anche se la significatività statistica è decisamente superiore, suggerendo che considerare anche la densità occupazionale rende i risultati inerenti l'esposizione alla attività essenziali più netti. Per quanto riguarda la comparabilità dei coefficienti, confrontiamo anche in questo caso una provincia al 75esimo percentile della distribuzione dell'interazione della quota essenziali e densità e una provincia al 25esimo percentile. Tale differenza è uguale a 33, che moltiplicata per la media dei coefficienti dopo il 22 marzo, che è pari a 0.4, porta ad un aumento di contagiati di 13 (contro i 10 contagiati quando si considera la variabile 'quota di rapporto di lavoro essenziali' senza associarla alla densità occupazionale).

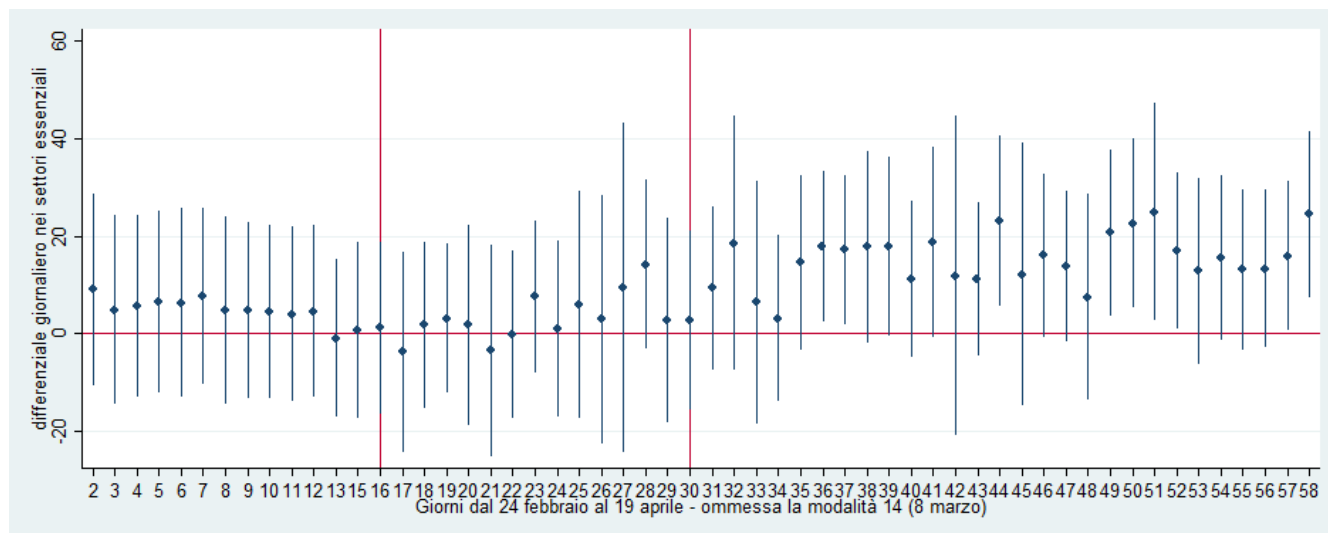
Figura 4: Impatto differenziale della variabile continua data dall'interazione quota di lavoratori essenziali e densità per km² (per provincia e per giorno).



Nella figura 5 implementiamo un ulteriore check nel quale cambiamo la specificazione econometrica. Invece che utilizzare la variabile continua della quota di essenziali, utilizziamo una dummy uguale a 1 per tutte le province che sono sopra la mediana di tale variabili, e zero per le province sotto la mediana. Si nota come l'andamento dei coefficienti è molto simile a quello individuato attraverso la variabile continua. Anche la dimensione dei coefficienti è comparabile. In

questo caso l'impatto dopo il 22 marzo oscilla tra i 10 e i 20 contagiati al giorno, ed è appunto un effetto medio fra tutte le province sopra la mediana rispetto alle province al di sotto. Con la specificazione continua avevamo determinato che la differenza fra una provincia al 75esimo e una al 25esimo percentile della quota di essenziali era di 10 contagiati, quindi lo stesso ordine di grandezza.⁷

Figura 5: Impatto differenziale della quota di settori essenziali come variabile binaria (sopra e sotto la mediana).



⁷ Non mostriamo gli stessi check di robustezza che abbiamo implementato per la specificazione con la variabile continua togliere le province lombarde, aggiungere variabili regionali come numero di tamponi, decessi, guariti), che confermano i risultati e che sono disponibili su richiesta.