ORIGINAL ARTICLE

The effect of a strict facial-mask policy on the spread of COVID-19 in Switzerland during the early phase of the pandemic

Emanuel Nussli*[1], Simon Hediger*[2†], Meta-Lina Spohn*[1] and Marloes H. Maathuis[1]

Abstract
Between July 2020 and October 2020 the cantons of Switzerland could choose to adopt the government-determined facial-mask policy, corresponding to mandatory facial-mask wearing on public transport, or a strict facial-mask policy, corresponding to mandatory facial-mask wearing on public transport and in all public or shared spaces where social distancing was not possible. We estimate the effect of introducing the strict facial-mask policy on the spread of COVID-19 in Switzerland during this first phase of the pandemic in 2020, using the cantonal heterogeneity in facial-mask policies. We adjust for social distancing behaviour, weather, other non-pharmaceutical policies and further variables. We estimate a significant reduction in the expected spread of COVID-19 in the early pandemic if the strict facial-mask policy is adopted.

Keywords: SARS-CoV-2; Policy Making; Random Effects; Fixed Effects; Facial Mask.

1 Introduction

The Coronavirus disease (COVID-19) pandemic has presented large challenges to societies around the world. In the early pandemic in 2020, where the Alpha variant of the SARS-CoV-2 virus was dominant, knowledge about the spread of the virus and about COVID-19 was scarce. In close collaboration with science, politicians and decision makers have been trying to contain the spread of COVID-19 while avoiding unnecessary restrictions. Non-pharmaceutical interventions such as school closures, restrictions on public and private gatherings and enforcement of home-office were employed.

In this paper, we focus on the effect of introducing a strict facial-mask policy on the containment of COVID-19 in Switzerland during the first phase of the pandemic in 2020. Studying the effect of facial masks is especially interesting as it is arguably one of the most debated policies. This might be partially due to the position that the Federal Office of Public Health of the Swiss Confederation (BAG) took in March 2020, communicating that healthy people do not need to wear facial masks[1]. A second reason for focusing on facial masks is that it is a relatively cheap and non-invasive policy when compared to other non-pharmaceutical interventions.

After the country-wide lockdown in Switzerland from mid-March to the end of April 2020, the 26 cantons were given partial autonomy in introducing COVID-19 containment measures in late June 2020[2]. The federal government-determined country-wide lower bounds on containment measures, and the cantons were allowed to deviate by mandating stricter policies. On July 6, 2020, wearing facial masks on

[*]Authors with equal contribution.
†Correspondence: simon.hediger@uzh.ch
[2]Department of Economics, University of Zurich, Zurich, CH
Full list of author information is available at the end of the article

[1]This was in alignment with the perspective of the World Health Organization (WHO). In July, 2020, this recommendation was overthrown due to increasing scientific evidence for the effectiveness of facial masks.

[2]See https://www.bag.admin.ch/bag/de/home/das-bag/aktuell/medienmitteilungen.msg-id-79522.html, last visited October 11, 2022, for further details.

public transport was made obligatory and thus formed the country-wide baseline for facial-mask policies[3]. We compare this to a strict facial-mask policy, corresponding to mandatory mask wearing on public transport and in all public or shared spaces where social distancing was not possible. In late December 2020 vaccinations against COVID-19 were initiated in Switzerland, marking a massive change point in the pandemic.

To estimate the effect of the strict facial-mask policy on the spread of COVID-19 during the early phase of the pandemic in Switzerland, we use the cantonal heterogeneity in facial-mask policies from July 2020 to December 2020. We quantify the spread of COVID-19 by two different, but related response variables: the estimated effective reproductive number (Huisman et al., 2022) and the approximated weekly growth rate of reported new cases (Chernozhukov et al., 2021). To identify the effect of the stricter facial-mask policy, we impose causal assumptions similar to Chernozhukov et al. (2021). We use a directed acyclic graph (DAG) to visualize the causal relationships among the facial-mask policy variable, the response variable and different sets of control variables. In addition to the observed relationships, we allow for canton-specific as well as time-specific effects in order to capture unobserved confounding via fixed and random effect approaches. By applying the backdoor-criterion and thus regressing the response variable on its parents in the DAG, we identify both the direct and total effect of the strict facial-mask policy variable on each of the two response variables. The direct effect, where direct means with respect to the variables we consider, captures changes in the response variable due to changes in the strict facial-mask policy variable, keeping the rest fix. The total effect additionally captures changes in the response variable that are mediated through changes in the social distancing behaviour.

We use publicly available data of different sources on the strict facial-mask policy variable, the response variables and the control variables. The data have a balanced panel structure, as the observations stem from the 26 cantons of Switzerland, measured during the 24 weeks of analysis. For both response variables we assume a linear regression model with a two-way error component, including a canton- and week-specific part. This model allows us to account for dependencies between the observations. Depending on the specific assumptions on the error components, the linear model is estimated with either a fixed-effects or random-effects approach. The coefficient of the strict facial-mask policy variable is our target of inference.

For both response variables and both direct and total effect, with both fixed- and random-effects approaches, we obtain significant negative point estimators of the coefficient of the strict facial-mask policy variable. In other words, we estimate a reduction in the spread of COVID-19 in the early pandemic if the facial-mask policy variable changes from the government-determined country-wide baseline to the strict facial-mask policy, while the rest of the variables in the model remain fixed. There is almost no difference between the estimates of the direct and total effect. In the model with the estimated effective reproductive number as response variable, the point estimators of both direct and total effect lie between -0.15 and -0.22. In the model with the approximated weekly growth rate of reported new cases as the response variable, the point estimators of both direct and total effect lie between -0.18 and -0.29.

To our knowledge, this is the first study that statistically analyzes the effect of the strict facial-mask policy on the spread of COVID-19 in Switzerland during the early phase of the pandemic. Pleninger et al. (2022) analyze the combination of all COVID-19 related policies in Switzerland, as measured by the Stringency Index of the Konjunkturforschungsstelle (KOF). They do not examine the isolated effect of the strict facial-mask policy. For Switzerland and Germany, Huber and Langen (2020) study the impact of the timing of lockdown measures on the cumulative death and hospitalization rates due to COVID-19, where they find that an early introduction reduces said

[3] See https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-79711.html, last visited October 11, 2022, for further details.

rates substantially compared to a later one. In other countries the effect of facial-mask policies has been studied. For the USA, Chernozhukov et al. (2021) study the effect of mandatory mask-wearing at the workplace. They estimate a significant reduction of the approximated weekly growth rate of reported new cases by around 0.1. For Germany, Mitze et al. (2020) find a 15%-75% reduction of new cases 20 days after the introduction of mandating facial masks in public transport and stores. Zhang et al. (2020) find that mandatory facial masks considerably slow down infection growth for the analyzed entities of New York, Wuhan and Italy. There are studies confirming the functionality of facial masks in hindering transmission of viral droplets in laboratory settings (see e.g., Kähler and Hain (2020)). However, in observational settings it is the effect of facial masks policies that is analysed, which includes mechanisms such as changes in risk-taking behaviour and miss-use of facial masks.

Direct comparison of our results to estimates in other countries and time spans is hard, due to different facial-mask policies and general differences between countries and their population. In essence, however our findings support the existing literature regarding the sign and significance of the effect of a strict facial-mask policy on the spread of COVID-19.

The article is organized as follows. Section 2 explains the data and the causal assumptions on the variables. Section 3 describes the methodology. Section 4 presents the results. Finally, we conclude in Section 5.

2 Data and Causal Assumptions

Our data are measured in each of the $i = 1, \ldots, N = 26$ cantons in Switzerland in each of the $t = 1, \ldots, T = 24$ weeks in the period of analysis, ranging from July 6, 2020 until December 21, 2020. We use weekly data because data at the daily resolution would artificially increase the sample size with highly dependent observations, which would lead to faulty statistical inference. Our variable of interest, the so-called treatment variable, is the strict facial-mask policy variable. To quantify the spread of COVID-19, we use two different response variables: the first is the estimated effective reproductive number and the second is the approximated weekly growth rate of reported new cases. We assume a directed acyclic graph (DAG), which is a graphical model displaying the causal relationships among the variables. Based on this DAG we identify the effect of the treatment variable on the response variables. A similar approach is used by Chernozhukov et al. (2021).

We consider eight groups of variables, observed in canton $i = 1, \ldots, N$ in week $t = 1, \ldots, T$[4]:

- $Y_{i,t}$: response variable quantifying the spread of COVID-19 (estimated effective reproductive number or approximated weekly growth rate of reported new cases),
- $M_{i,t}$: strict facial-mask policy variable (treatment variable),
- $B_{i,t}$: (social distancing) behavior variable, quantified by financial transactions,
- $\boldsymbol{D}_i$: demographic variables that are canton-specific,
- $H_{i,t}$: holiday indicator variable,
- $\boldsymbol{W}_{i,t}$: meteorological variables reflecting the weather situation,
- $\boldsymbol{P}_{i,t}$: non-pharmaceutical policy variables (excluding $M_{i,t}$),
- $Y_{i,t'}$: response variable $Y_{i,t}$ lagged to the past with lag t',

Subsequently, we present the specific variables (with their `short name` in brackets) in each of these categories as well as the assumptions on their causal relationships in the form of a DAG. We present a list of the 16 variables with their sources and descriptive statistics in Table 1.

2.1 Response Variables ($Y_{i,t}$)

The first response variable is the estimated effective reproductive number (`r`) of Huisman et al. (2022). The estimated effective reproductive number at day d is an estimate of the expected number of secondary infections at day d caused by a previously infected person. Its estimation involves multiple steps: 1) estimation of

[4] As convention, we use bold letters for multivariate variables and normal letters for univariate variables.

the number of newly infected people based on the number of newly confirmed cases, adjusting for reporting cycles and irregular reporting practices, 2) a deconvolution step using suitable delay distributions between transmission and reporting of the case to infer the actual infection incidence, 3) application of the EpiEpstim method developed by Cori et al. (2013) to estimate the effective reproductive number from the time series of newly infected people. Only cases stemming from infections within Switzerland are used for estimation. In each canton i, to obtain an observation for week t, we average the daily values within week t. We denote this response variable by $Y_{i,t} = R_{i,t}$. We obtain the data from the Federal Office of Public Health of Switzerland[5].

The second response variable is the approximated weekly growth rate of reported new cases. In each canton i in week t, it is defined by

$$G_{i,t} := \ln\left(\frac{C_{i,t}}{C_{i,t-1}}\right),$$

where $C_{i,t}$ represents the number of new confirmed cases in canton i in week t. Due to the delay between the reporting of a new case and the actual infection time with the virus, $G_{i,t}$ does not represent the pandemic situation in week t but of a time period before t. Therefore, to obtain a measurement of the approximated weekly growth rate of reported new cases in week t, we need to to use a future value of $G_{i,t}$. We employ the same time shift of 2 weeks to the future as Chernozhukov et al. (2021), resulting in the response variable $Y_{i,t} = G_{i,t+2}$ (`growth.new.cases`). We obtain the data on new reported confirmed cases from the Federal Office of Public Health of Switzerland[6].

We plot both responses in Figure 2 for all 26 cantons in the period of analysis. The plots and the Pearson correlation coefficients ρ highlight that the two responses are similar but there isn't a one-to-one correspondence between the estimated effective reproductive number and the approximated weekly growth rate of reported new cases.

2.2 Strict Facial Mask Policy Variable ($M_{i,t}$)

The strict facial-mask policy variable (`facial.mask`) is our treatment variable. In each canton i, it has a value of 1 if the strict policy is applied and a value of 0 if the government-determined baseline policy is applied. During the period of analysis, a total of 10 cantons deviate from the baseline policy by preemptively introducing the strict policy. At a daily resolution, there are 1833 observations where the strict policy is in place and 2561 where the baseline policy is implemented. For more details, see Table 1 and Figure 5. To obtain an observation for week t, we average the daily values within week t. We obtain the data from KOF, via their CRAN R-package `kofdata` (Bannert and Thoeni, 2022).

2.3 Social Distancing Behaviour Variable ($B_{i,t}$)

As a proxy for social distancing behaviour we use household spending, similar to Pleninger et al. (2022). In each canton i, we consider the logarithm of the ratio between the number of transactions in week t and the number of transactions in week $t-1$ in CHF in stores in Switzerland with credit cards, debit cards and bank transfers from mobile phones (`growth.transactions`). E-commerce is not considered. We obtain the data from Monitoring Consumption Switzerland[7].

2.4 Demographic Variables ($\boldsymbol{D}_i$)

Variables that affect all other types of variables except for the meteorological variables include demographic variables of a canton i, given by population size (`population`) and the percentage of people with age ≥ 80 years (`perc.o80`). Wheaton and Thompson (2020) show that infection growth is also strongly linked to residential density, that is the number of people per km^2 of settlement area (`density`), which we also include. These three variables can be

[5] See https://opendata.swiss/en/dataset/covid-19-schweiz, last visited October 11, 2022, for further details.

[6] See https://opendata.swiss/en/dataset/covid-19-schweiz, last visited October 11, 2022, for further details.

[7] A project by the universities of St. Gallen and Lausanne, see https://monitoringconsumption.com/, last visited October 11, 2022, for further details.

considered constant for all weeks t. We obtain the data from the Federal Statistical Office of Switzerland[8].

2.5 Holiday Indicator ($H_{i,t}$)

In each canton i the daily holiday indicator (`holiday`) has a value of 1 if the majority of public schools in the canton are on holiday and 0 otherwise. To obtain an observation for week t, we average the daily values within week t. We obtain the holiday data from the cantonal education departments[9].

2.6 Meteorological Variables ($\boldsymbol{W}_{i,t}$)

Zoran et al. (2020) suggest that weather conditions are closely linked to the spread of COVID-19. In particular, they find that dry air supports the transmission of COVID-19. Their findings are supported by Zhu et al. (2020) and Fattorini and Regoli (2020). To incorporate these effects, we assemble weather data from 14 representative weather stations, where we excluded weather stations located on mountains. We matched each canton i to the closest weather station to get a characterization of its daily weather, quantified by the number of minutes of sunshine (`sunshine`), the mean air temperature in °C (`temperature`), and the relative humidity in % (`humidity`). To obtain an observation for week t, we average the daily values within week t. We obtain the data from the Federal Office of Meteorology and Climatology of Switzerland[10].

2.7 Non-Pharmaceutical Policy Variables ($\boldsymbol{P}_{i,t}$)

The KOF Stringency Plus Index (Pleninger et al., 2022), the Government Response Index and the Economic Support Index (Hale et al., 2021)[11] compose different sets of policy variables into one index with the aim to reflect the stringency of a government in regards to COVID-19 policies. We do not use these indices, but use the policy variables directly, where we only consider those that vary at least across cantons or across weeks over the period of analysis. In each canton i, these policy variables are daily indicators for workplace closings (`work.closing`), school closings (`school.closing`), restrictions on gatherings (`rest.gatherings`), cancellation of public events (`canc.events`), and testing policy (`testing.policy`)[12]. These indicators have 2 to 5 levels, where a higher level indicates a stricter policy. To obtain an observation for week t, we average the daily values within week t. We obtain the data from KOF, provided through their CRAN R-package `kofdata` (Bannert and Thoeni, 2022).

2.8 Lagged Response Variables as Predictors ($Y_{i,t'}$)

We consider a lagged response variable as predictor (`Y.lagged`): We include the lagged response variable of which the value is known in week t, summarizing the information about the pandemic situation that is available to the public in week t. Knowledge about the current pandemic situation strongly drives the policy decisions and the behaviour of the population.

If we consider the weekly average estimated effective reproductive number as response variable, that is $Y_{i,t} = R_{i,t}$, the information variable is given by $Y_{i,t'} = R_{i,t-3}$, which corresponds to the estimated effective reproductive number of three weeks ago. This lag is due to time delays between the infection, the start of the symptoms and the report of a case, such that in week t only the value of three weeks ago is known. If we consider the weekly growth rate of reported new cases as response variable,

[8] See https://www.bfs.admin.ch/bfs/en/home/statistics/regional-statistics/regional-portraits-key-figures/cantons.assetdetail.20784336.html, last visited October 11, 2022, for further details.

[9] See https://www.edk.ch/en/education-system/websites-of-the-cantons, last visited October 11, 2022, for further details.

[10] See https://opendata.swiss/en/dataset/klimamessnetz-tageswerte, last visited October 11, 2022, for further details.

[11] See https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md, last visited October 11, 2022, for all indicators used to calculate the index.

[12] Indicator-coding: https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md,last visited October 11, 2022.

that is $Y_{i,t} = G_{i,t+2}$, the information variable is given by $Y_{i,t'} = G_{i,t}$, a value that we assume to be readily available in week t.

2.9 Causal Assumptions

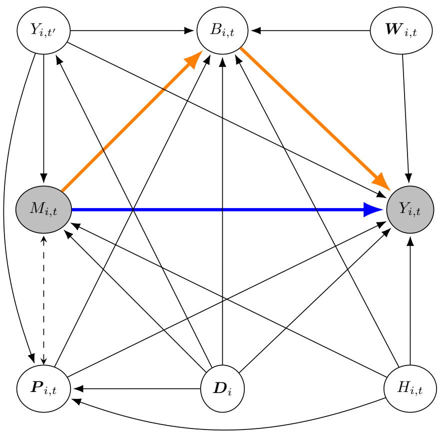

Figure 1: Assumed DAG on the eight sets of variables. The blue edge represents the direct effect of the strict facial-mask policy $M_{i,t}$ on the response variable $Y_{i,t}$. The blue edge in conjunction with the path along the orange edges represents the total effect of the strict facial-mask policy $M_{i,t}$ on the response variable $Y_{i,t}$.

To identify the effect of the strict facial-mask policy on the spread of COVID-19, we impose causal assumptions similar to Chernozhukov et al. (2021). We use a DAG to visualize the assumed causal relationships among the sets of variables, see Figure 1.

The causal relationships between the variables are assumed to be the same for both response variables. The gray nodes represent the strict facial-mask policy, our treatment variable, and the response variable. The white nodes represent the covariates. A directed edge $A \to B$ between nodes A and B represents a causal relationship, where an intervention on A results in changes in B. The bi-directed, dashed edge between $M_{i,t}$ and $\boldsymbol{P}_{i,t}$ means that they are allowed to have unmeasured common causes.

Note that the DAG in Figure 1 is not a DAG in the strict sense, since certain nodes represent groups of variables. We allow the variables within such groups to be arbitrarily causally related so that there are no cycles. An edge to or from such a group of variables indicates that we allow such an edge for each variable in the group.

The strict facial-mask policy variable $M_{i,t}$ is assumed to influence the response variable directly or indirectly. The blue edge, $M_{i,t} \to Y_{i,t}$, represents the *direct effect*. The path in orange, $M_{i,t} \to B_{i,t} \to Y_{i,t}$, represents the *indirect effect*, the effect of $M_{i,t}$ on the spread of COVID-19 through its effect on $B_{i,t}$[13]. This corresponds to differences in social distancing behaviour of the public due to the policy $M_{i,t}$: For example, some people might stay at home more while some people might maintain social contacts due to a feeling of safety wearing a facial mask. The sum of the direct and indirect effects results in the *total effect* of the strict facial-mask policy variable on the response variable.

Furthermore, note that the inclusion of canton-specific and time-specific effects capturing unobserved confounding via the fixed effects methodology corresponds two additional nodes, denoted by α_i and γ_t respectively in Section 3, that affect all other variables. The random effects approach translates to two additional nodes α_i and γ_t with an outgoing edge to $Y_{i,t}$. For more details, see Sections 3.1 and 3.2.

3 Methodology

Our data have a balanced panel structure where the observations stem from the same units (cantons), measured longitudinally at the same time points (weeks). In the following, for cantons $i = 1, \ldots, N$ and weeks $t = 1, \ldots, T$, let $\boldsymbol{Z}_{i,t}$ be a row vector of a valid adjustment set relative to the effect of interest, either the direct or total effect of $M_{i,t}$ on $Y_{i,t}$; we specify this set later. The two-way error component

[13] This is called a mediation structure, for a modern treatment of this field, see Robins et al. (2022).

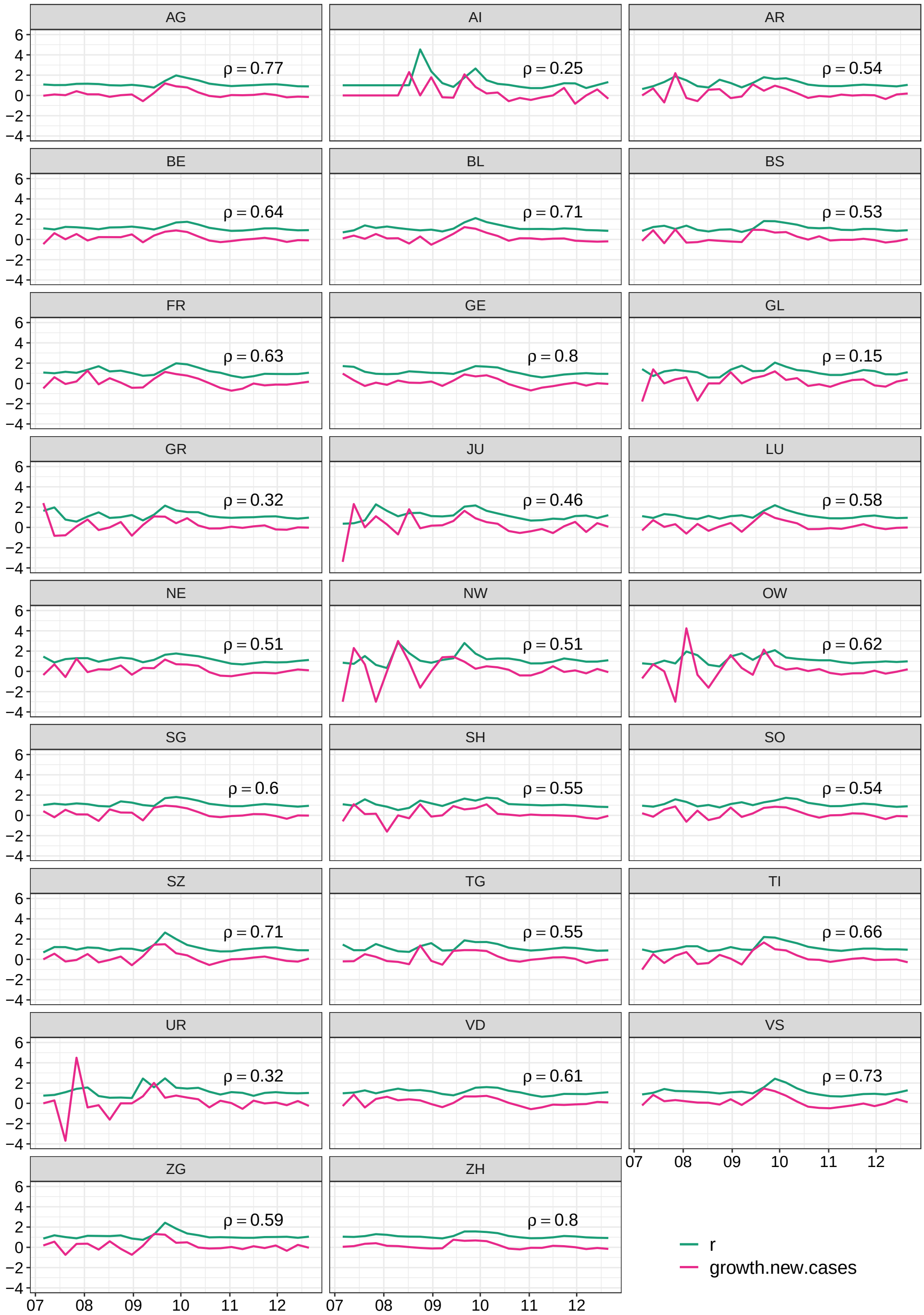

Figure 2: Time series of the two response variables, the estimated effective reproductive number $R_{i,t}$, and the approximated weekly growth rate of reported new cases $G_{i,t+2}$, between July 6, 2020 and December 21, 2020, for each of the 26 cantons. The labels on the x-axis represent the months between July and December 2020 (07-12). For each canton we report the Pearson correlation coefficient ρ between the two time series.

Short Name	Description	Descriptive Statistics (median, mean, sd)	Data Source
$Y_{i,t}$	Response Variables		
`r`	estimated effective reproductive number	$(1.05, 1.15, 0.38)$	COVID
`growth.new.cases`	$\ln$(weekly growth rate of reported new cases)	$(0.05, 0.15, 0.67)$	COVID
$M_{i,t}$	Strict Facial Mask Policy Variable		
`facial.mask`	strict facial-mask policy	$(0.00, 0.44, 0.49)$	Policies
$B_{i,t}$	Social Distancing Behaviour Variable		
`growth.transactions`	$\ln$(weekly growth rate of transactions)	$(0.00, -0.01, 0.08)$	Consumption
$\boldsymbol{D}_i$	Demographic Variables		
`population`	population	$(234'650, 326'311, 348'969)$	Population
`perc.o80`	age ≥ 80 years in $\%$	$(5.29, 5.44, 0.74)$	Population
`density`	people per km^2 of settlement area	$(2503, 2780, 1253)$	Population
$H_{i,t}$	Holiday Indicator		
`holiday`	official school holiday indicator	$(0.00, 0.08, 0.25)$	Holidays
$\boldsymbol{W}_{i,t}$	Meteorological Variables		
`sunshine`	sunshine in minutes per day	$(248, 288, 192)$	Weather
`temperature`	mean air temperature in $^\circ\mathrm{C}$	$(11.36, 11.89, 7.15)$	Weather
`humidity`	relative humidity in $\%$	$(78.94, 76.62, 9.10)$	Weather
$\boldsymbol{P}_{i,t}$	Non-Pharmaceutical Policy Variables		
`work.closing`	closing of workplaces policy	$(1.00, 1.49, 0.71)$	Policies
`school.closing`	closing of schools policy	$(1.00, 1.32, 0.47)$	Policies
`rest.gatherings`	restrictions on gatherings policy	$(2.00, 2.04, 1.12)$	Policies
`canc.events`	cancellation of public events policy	$(1.00, 1.37, 0.47)$	Policies
`testing.policy`	testing policy	$(2.00, 2.42, 0.48)$	Policies

Table 1: Short name, description, descriptive statistics and data source for all the variables and responses used in the analysis. The baseline for each of the social distancing behaviour variables is the respective median in the period of January 3, 2020 until February 6, 2020. For more details regarding the interpretation of the values of the non-pharmaceutical policy variables, see the description under `https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md`, last visited October 11, 2022.

linear regression model is given by

$$Y_{i,t} = \theta M_{i,t} + \boldsymbol{\beta}^\top \boldsymbol{Z}_{i,t} + \alpha_i + \gamma_t + \epsilon_{i,t}, \qquad (1)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where the coefficient θ is our target of inference, $\boldsymbol{\beta}$ is a coefficient vector of the same length as $\boldsymbol{Z}_{i,t}$, α_i is an unobserved canton-specific effect, γ_t is an unobserved week-specific effect, and $\epsilon_{i,t}$ is the remaining error. Depending on the assumptions on the error components α_i, γ_t and $\epsilon_{i,t}$, Model (1) can be handled with either a fixed-effects or random-effects approach[14]. Generally, the fixed-effects approach is more robust than the random-effects approach, while the latter is more efficient in case all assumptions are met. We briefly outline both approaches in the upcoming sections, for more details see for example Hansen (2022). The suitability of the

[14]We also assume linearity of the generating equation of $B_{i,t}$, which is required to identify the total effect with the linear fixed-effects and random-effects approaches. See Chernozhukov et al. (2021) for an explicit derivation of the total effect.

linearity assumption in Equation (1) is assessed via Tukey-Anscombe plots (residual v.s. fitted values), shown in Figure 6 in Appendix D.

3.1 Fixed-Effects Approach

The fixed-effects approach assumes that the stochastic structure of α_i and γ_t is unknown and possibly arbitrarily correlated with $M_{i,t}$ and $\boldsymbol{Z}_{i,t}$. In this case, we call α_i an unobserved cantonal fixed effect and γ_t an unobserved weekly fixed effect. The incorporation of fixed effects accounts for unobserved common causes of the treatment and response variable that are either canton-specific, but invariant across weeks, or week-specific, but invariant across cantons. The variance-covariance structure of the error terms $\epsilon_{i,t}$ can take many forms; see the upcoming Section 3.1.1. However, $\epsilon_{i,t}$ are always supposed to satisfy the exogeneity assumption,

$$\mathbb{E}[\epsilon_{i,t} \mid M_{i,t}, \boldsymbol{Z}_{i,t}, \alpha_i, \gamma_t] = 0, \qquad (2)$$

for all $i = 1, \dots, N$ and $t = 1, \dots, T$. This assumption implies no further unobserved confounding apart from α_i and γ_t. To eliminate α_i and γ_t, we apply the two-way within transformation,

$$\ddot{u}_{i,t} := u_{i,t} - \frac{1}{N}\sum_{i=1}^{N} u_{i,t} - \frac{1}{T}\sum_{t=1}^{T} u_{i,t} + \frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{i,t}, \qquad (3)$$

to $u_{i,t} \in \{Y_{i,t}, M_{i,t}, \boldsymbol{Z}_{i,t}, \alpha_i, \gamma_t, \epsilon_{i,t}\}$ of Model (1) and obtain the following equation

$$\ddot{Y}_{i,t} = \theta \ddot{M}_{i,t} + \boldsymbol{\beta}^\top \ddot{\boldsymbol{Z}}_{i,t} + \ddot{\epsilon}_{i,t}, \qquad (4)$$

where the interpretation of θ remains as in Model (1). Finally, we estimate the coefficient θ by estimating the whole coefficient vector $\boldsymbol{\eta} = (\theta, \boldsymbol{\beta})$, using Ordinary Least Squares (OLS). In the following, we use the acronym `FE` for this approach.

Apart from the basic OLS estimate, we also compute a debiased estimate of θ, $\hat{\theta}_{BC}$, by cross-over Jackknife bias correction (Chen et al., 2019, 2020; Chernozhukov et al., 2021). We employ this method as the estimation of dynamic linear panel models (i.e. including lagged instances of the response variable as covariates) using the fixed effects estimator yields a bias of the estimated coefficients of order $\mathcal{O}(1/T)$[15]. The debiased estimate is given by

$$\hat{\theta}_{BC} = 2\hat{\theta} - (\hat{\theta}_{S_1} + \hat{\theta}_{S_2})/2,$$

where $\hat{\theta}$ is the OLS regression coefficient based on the entire sample and $\hat{\theta}_{S_j}$ is the estimated coefficient computed on the sub-sample S_j, $j = 1, 2$. The sub-samples S_1 and S_2 are defined, as in Chernozhukov et al. (2021), by

$$\begin{aligned} S_1 :=& \{(i,t) : i \leq \lceil N/2 \rceil, t \leq \lceil T/2 \rceil\} \\ &\cup \{(i,t) : i \geq \lceil N/2 + 1 \rceil, t \geq \lceil T/2 + 1 \rceil\} \end{aligned}$$

and

$$\begin{aligned} S_2 :=& \{(i,t) : i \leq \lceil N/2 \rceil, t \geq \lceil T/2 \rceil\} \\ &\cup \{(i,t) : i \geq \lceil N/2 + 1 \rceil, t \leq \lceil T/2 + 1 \rceil\}, \end{aligned}$$

respecting the natural ordering of the weeks. Since there is no natural ordering of the cantons, we repeat the above procedure 20 times, where each time the cantons are randomly permuted. The final estimate is then the average of the 20 debiased estimates[16]. In the following, we use the acronym `DFE` for this debiased fixed-effects approach.

We now detail the specific sets of control variables $\boldsymbol{Z}_{i,t}$, which depend on whether we aim at estimating the direct or the total effect of the strict facial-mask policy on the spread of COVID-19. Due to the within transformation (3), apart from the unobservable α_i and γ_t, also all observable week-constant variables, such as policy indicators that do not vary over the period of analysis, and canton-constant variables, such as population or density, drop out of $\boldsymbol{Z}_{i,t}$. In the case of the direct effect, we must thus regress the response variable on all its remaining parents, i.e., $\boldsymbol{Z}_{i,t} = (B_{i,t}, H_{i,t}, \boldsymbol{W}_{i,t}, \boldsymbol{P}_{i,t}, Y_{i,t'})$. In the case of the

[15] Note that this is technically not a bias but rather a probability limit.

[16] The estimates are robust against changes in the number of permutations.

total effect, we need to remove the variable $B_{i,t}$ from the adjustment set, and obtain $\boldsymbol{Z}_{i,t} = (H_{i,t}, \boldsymbol{W}_{i,t}, \boldsymbol{P}_{i,t}, Y_{i,t'})$. Concretely, the following variables are contained in each category:

- $B_{i,t}$: `growth.transactions`,
- $H_{i,t}$: `holiday`,
- $\boldsymbol{W}_{i,t}$: `sunshine`, `temperature` and `humidity`,
- $\boldsymbol{P}_{i,t}$: `rest.gatherings`, `canc.events` and `work.closing`,

and the variable $Y_{i,t'}$ is specific for each of the two response variables, see Section 2.8.

3.1.1 Construction of Confidence Intervals

To construct 95%-confidence intervals for the coefficient θ, we use the normal approximation,

$$CI_{95\%} := \left[\hat{\theta} \pm 1.96\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}\right], \tag{5}$$

where $\hat{\theta}$ is the first entry in the estimated coefficient vector $\hat{\boldsymbol{\eta}} = (\hat{\theta}, \hat{\boldsymbol{\beta}})$, obtained either through the `FE` or `DFE` approach, and $\widehat{\mathrm{Var}}(\hat{\theta})$ is the corresponding estimated variance. The estimation of the variance requires careful consideration due to the panel structure of our data. In Appendix B, we present a detailed explanation of the estimators used, which are some of the most widely employed ones.

3.2 Random-Effects Approach

The random-effects approach assumes that the components of the error, α_i, γ_t, and $\epsilon_{i,t}$, satisfy the following exogeneity assumptions

$$\begin{aligned} &\mathbb{E}\left[\alpha_i \mid M_{i,t}, \boldsymbol{Z}_{i,t}\right] = 0, \\ &\mathbb{E}\left[\gamma_t \mid M_{i,t}, \boldsymbol{Z}_{i,t}\right] = 0, \\ &\mathbb{E}\left[\epsilon_{i,t} \mid M_{i,t}, \boldsymbol{Z}_{i,t}\right] = 0, \end{aligned} \tag{6}$$

for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$. These assumptions imply that α_i, γ_t, and $\epsilon_{i,t}$ are uncorrelated with $M_{i,t}$ and $\boldsymbol{Z}_{i,t}$, which implies the strong assumption of no unobserved confounding. The correlation within weeks and within cantons in the composite error $v_{i,t} = \alpha_i + \gamma_t + \epsilon_{i,t}$ is accounted for by the Feasible Generalized Least Squares (FGLS) approach, where we assume the following structure of the variance-covariance matrix,

$$\mathrm{Cov}(v_{i,t}, v_{j,s}) = \left\{ \begin{array}{ll} \sigma_i^{\alpha} + \sigma_i^{\gamma} + \sigma_i^{\epsilon}, & i = j,\ t = s, \\ \sigma_i^{\alpha} + \sigma_i^{\gamma}, & i = j,\ t \neq s, \\ 0, & \text{otherwise} \end{array} \right\},$$

where $\sigma_i^{\alpha} > 0$, $\sigma_i^{\gamma} > 0$ and $\sigma_i^{\epsilon} > 0$. With this approach we again obtain an estimator of the whole coefficient vector $\boldsymbol{\eta} = (\theta, \boldsymbol{\beta})$ and extract the estimator of θ. To construct 95%-confidence intervals for θ, we again use Equation (5), where we apply Formula (8) with a plug-in estimator of $\boldsymbol{\Omega}$. In the following, we use the acronym `RE` for the point estimator as well as the confidence interval of this approach.

In contrast to the fixed-effects approach, the variables $\boldsymbol{D}_i$ that are part of the parents of the response variable, do not drop out of $\boldsymbol{Z}_{i,t}$. Furthermore, the set of policy variables now includes `school.closing` and `testing.policy`, which were dropped in the `FE` approach as they contain constant values within cantons over the period of analysis and are thus eliminated by the within transformation. In the case of the direct effect we obtain $\boldsymbol{Z}_{i,t} = (B_{i,t}, \boldsymbol{D}_i, H_{i,t}, \boldsymbol{W}_{i,t}, \boldsymbol{P}_{i,t}, Y_{i,t'})$. In the case of the total effect, we need to remove the variable $B_{i,t}$ from the adjustment set, and obtain $\boldsymbol{Z}_{i,t} = (\boldsymbol{D}_i, H_{i,t}, \boldsymbol{W}_{i,t}, \boldsymbol{P}_{i,t}, Y_{i,t'})$. Concretely, the following variables are contained in each category:

- $B_{i,t}$: `growth.transactions`,
- $\boldsymbol{D}_i$: `population`, `density` and `perc.080`,
- $H_{i,t}$: `holiday`,
- $\boldsymbol{W}_{i,t}$: `sunshine`, `temperature` and `humidity`,
- $\boldsymbol{P}_{i,t}$: `rest.gatherings`, `canc.events`, `work. closing`, `school.closing` and `testing.policy`,

and the variable $Y_{i,t'}$ is specific for each of the two response variables, see Section 2.8.

3.3 Sensitivity Analysis

We perform an extensive sensitivity analysis to investigate the robustness of our results towards changes in our methodology and data preparation. Please see Appendix A.

3.4 Implementation

We implement the methodology in the software R, using the packages `plm`, `sandwich`, and `lmtest`.

4 Results

For each of the three modeling approaches (fixed-effects `FE`, debiased fixed-effects `DFE`, and random-effects `RE`) we distinguish between the direct and total effect of introducing a strict facial-mask policy (as compared to the government-determined country-wide baseline policy) on the two response variables. This results in 12 estimated effects along with their confidence intervals.

Figure 3 shows the results for the direct effect of the strict facial-mask policy. We see that all 95%-confidence intervals lie to the left of zero. Thus, the direct effect of the strict facial-mask policy was found to be significantly negative in all modelling approaches, implying a significant reduction in the spread of COVID-19. Figure 4 shows the results for the total effect of the strict facial-mask policy, and we see that the overall picture is very similar.

The point estimators of both the direct and total effect in the model with response variable `r` lie between -0.15 and -0.22. In the model with response variable `growth.new.cases`, the point estimators lie between -0.18 and -0.29. These numbers are estimates for the average treatment effect (ATE) of the strict facial-mask policy. Let $Y_{i,t}^{(1)}$ be the value of the response in canton i at time t had canton i implemented the strict facial-mask policy at time $t-\ell$, as described in detail in Subsection 2.1. Conversely, $Y_{i,t}^{(0)}$ denotes the value of the response for canton i and time t had the strict facial-mask policy not been implemented at time $t-\ell$. The ATE is defined as

$$\mathrm{ATE} := \mathbb{E}\left(Y_{i,t}^{(1)} - Y_{i,t}^{(0)}\right), \tag{7}$$

where the expectation is taken over all the observations. See Rubin (1974) for a treatment of the framework of potential outcomes. This allows us to interpret the estimates in a concise manner: We estimate that the introduction of the strict facial mask policy — on average – lowers the effective reproductive number by 0.15 to 0.22. Considering that the average effective reproductive number was 1.15 in our sample (see Table 1), this corresponds to a substantial effect. The analogous interpretation holds for the weekly growth rate of new cases where we estimate an average reduction by introducing the strict facial-mask policy of between 0.18 and 0.22, while the average weekly growth rate of new cases is 0.15 in our sample.

The fact that the estimated direct and total effects are very similar suggests that either the facial-mask policy worked mainly through the direct path by reducing the transmissibility of COVID-19, or the behavioural variable `growth.transactions` is not capturing the important changes in social distancing behaviour. It is plausible that the latter is at least part of the explanation, as it is for example unclear how this variable can reflect changes in behaviour in private spaces. This is a substantial limitation in the analysis of decomposing the effect of the facial mask policy into the direct effect and the effect via social distancing – the credibility of the total effect of the facial mask policy on the spread of COVID-19 is not reduced however. However, Chernozhukov et al. (2021) employ a closely related empirical approach for the U.S. in the early pandemic and they do not find the indirect effect to be significant either. They use Google Mobility Reports to capture social distancing behaviour, which has been investigated and deemed a good representation by Maloney and Taskin (2020). The Google Mobility Reports are not available at the cantonal level for Switzerland however, but Chernozhukov et al. (2021) may be taken as evidence that the facial mask policy mainly works through the direct effect.

For all 12 modeling approaches, the Tukey-Anscombe plots (residuals v.s. fitted values), displayed in Figure 6 in Appendix D, show no evidence against the assumption of linearity. The results of the extensive sensitivity analysis are shown in Table 2 in Appendix A. In line with our main analyses, all 35 reported point estimators are negative and 25 out of the 35 point estimators are significantly different from zero at the $\alpha = 0.05$ level.

5 Conclusion

We analyze the effect of the strict facial-mask policy on the spread of COVID-19 during the early phase of the pandemic in Switzerland, using the cantonal heterogeneity in facial-mask policies from July 2020 to December 2020. The obligation to wear a facial mask in public transportation formed the government-determined country-wide baseline for facial-mask policies. The strict facial-mask policy corresponds to mandatory mask wearing on public transport and in all public or shared spaces where social distancing is not possible.

We found a significant reduction in the expected spread of COVID-19 in the early pandemic if the facial-mask policy is changed from the government-determined country-wide baseline to the strict facial-mask policy. Importantly, we do not investigate whether the estimated effects are relevant in any given social context.

It is important to stress that in an observational study like the one at hand it is almost impossible to correct for all unobserved confounders. It is highly likely that our effect of interest is confounded by further social, cultural and economic traits that may differ between cantons and weeks. Implementation of and compliance with non-pharmaceutical policies like the strict facial-mask policy are subject to cultural norms, political backgrounds and defiance against political authorities and policy makers, to just name a few examples of such factors. Hence, the results should be treated with caution and interpreted in light of the mostly untestable assumptions inherent in our modelling approaches, described in Sections 2 and 3. In particular, we want to emphasize that the assumption no unmeasured confounding imposed using the `RE` methodology is very delicate and might not hold, considering we do not control for the development of COVID-19 in neighboring countries. The fact that the estimates using the `FE` and `DFE` approaches are very similar to those using `RE` are a good sign however that we capture the pre-policy dynamics well as these allow for unmeasured confounding as explained in Section 3.1.

Furthermore, our results are conditional on the specific time period between July and December 2020 during the early pandemic, when the Alpha variant of the SARS-CoV-2 virus was predominant and no vaccinations were available yet. However, even though direct comparison to results in other countries and time spans is hard, our findings are largely in line with those of other groups (see e.g., Pleninger et al. (2022), Chernozhukov et al. (2021), and Mitze et al. (2020)).

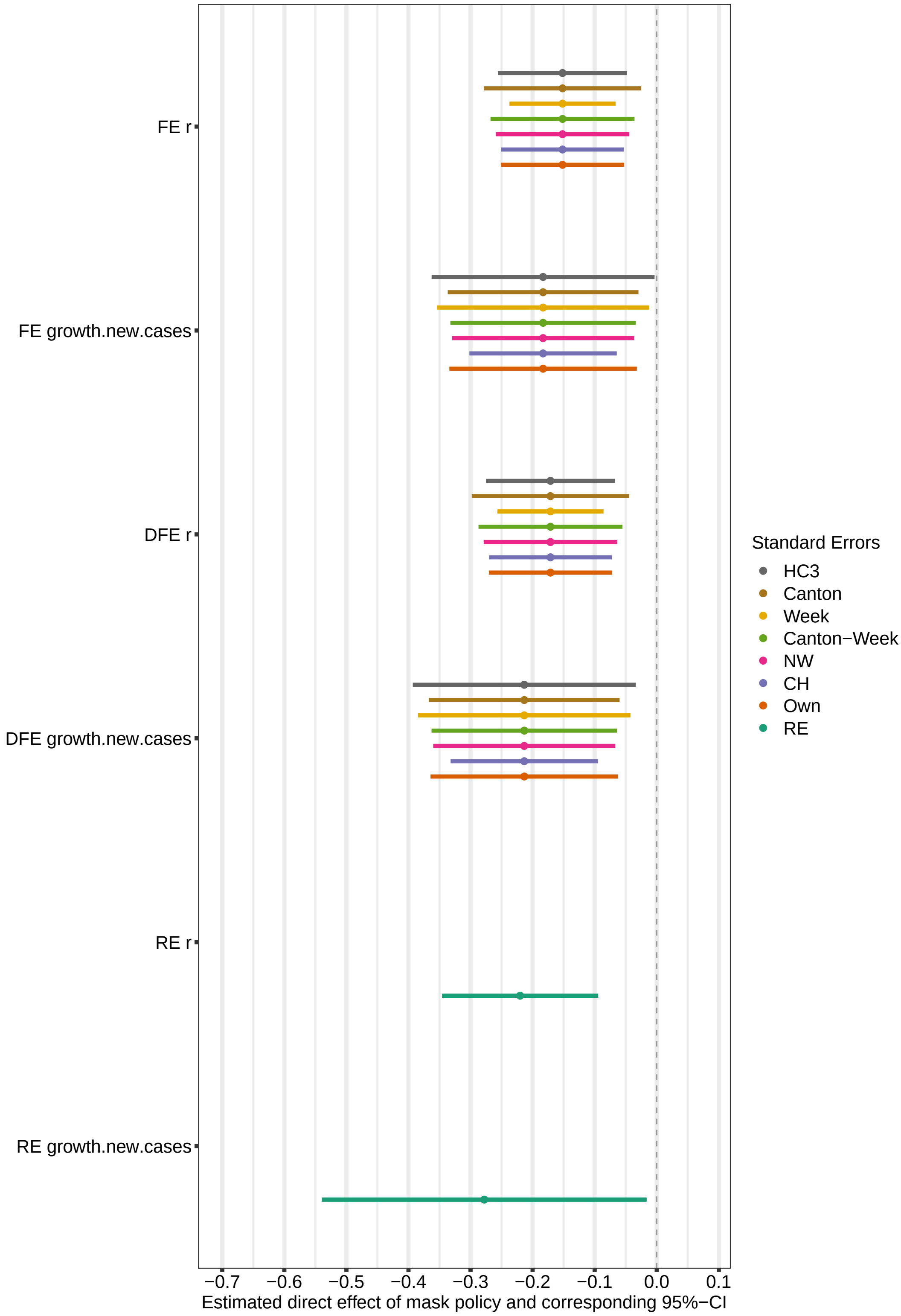

Figure 3: Point estimators and 95%-confidence intervals ($CI_{95\%}$) for the direct effect of the strict facial-mask policy variable (`facial.mask`) on the estimated effective reproductive number (`r`) or the approximated weekly growth rate of reported new cases (`growth.new.cases`) for different modeling approaches (described in Section 3): the fixed-effects approach (`FE`), a debiased variant of the fixed-effects approach (`DFE`), and a random-effects approach (`RE`). For `FE` and `DFE`, we construct seven different confidence intervals reflecting different assumptions on the dependencies in the data, as described in Section 3.

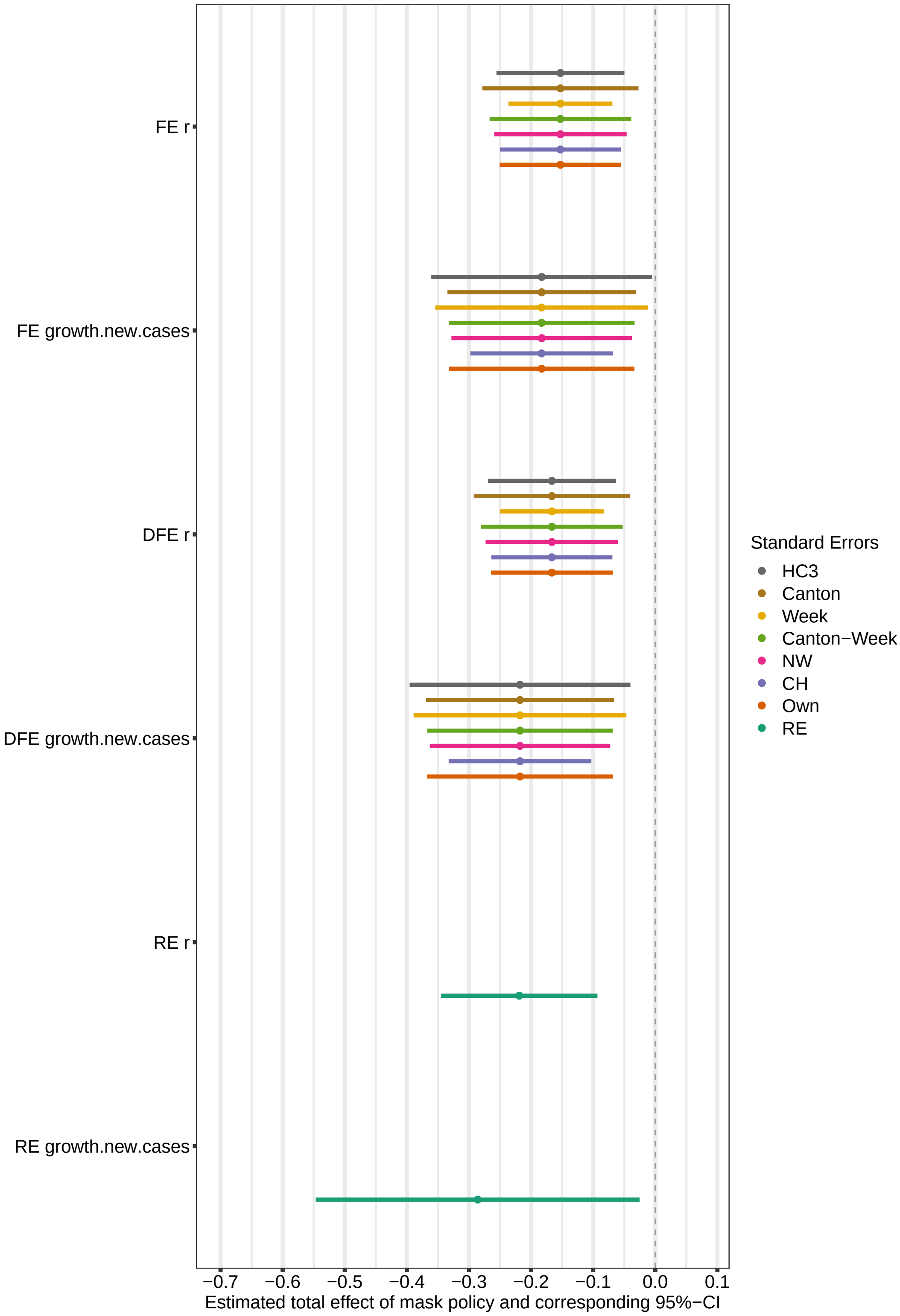

Figure 4: Point estimators and 95%-confidence intervals ($CI_{95\%}$) for the total effect of the strict facial-mask policy variable (`facial.mask`) on the estimated effective reproductive number (`r`) or the approximated weekly growth rate of reported new cases (`growth.new.cases`) for different modeling approaches (described in Section 3): the fixed-effects approach (`FE`), a debiased variant of the fixed-effects approach (`DFE`), and a random-effects approach (`RE`). For `FE` and `DFE`, we construct seven different confidence intervals reflecting different assumptions on the dependencies in the data, as described in Section 3.

Availability of data and materials
https://github.com/enussl/Facial-Mask-Policy-COVID-19

Competing interests
The authors declare that they have no competing interests.

Funding
Not applicable

Authors' contributions
E.N., S.H. and M-L.S. contributed equally to the manuscript. M.H.M. helped in the creation of this manuscript with her statistical expertise. All authors read and approved the final manuscript.

Acknowledgements
Not applicable

Abbreviations

- BAG: Bundesamt für Gesundheit
- DAG: directed acyclic graph
- KOF: Konjunkturforschungsstelle
- OLS: Ordinary Least Squares
- FGLS: Feasible Generalized Least Squares
- CI: confidence interval
- ATE: Average Treatment Effect
- FE: fixed effects
- RE: random-effects
- DFE: debiased fixed-effects
- DML: double machine learning

Author details
[1]Seminar for Statistics, ETH Zurich, Zurich, CH.
[2]Department of Economics, University of Zurich, Zurich, CH.

References

Bannert, M., Thoeni, S.: Kofdata: Get Data from the 'KOF Datenservice' API. (2022). R package version 0.2. https://CRAN.R-project.org/package=kofdata
Chen, S., Chernozhukov, V., Fernández-Val, I.: Mastering panel metrics: Causal impact of democracy on growth. AEA Papers and Proceedings **109** (2019)
Chen, S., Chernozhukov, V., Fernandez-Val, I., Kasahara, H., Schrimpf, P.: Cross-over jackknife bias correction for non-stationary nonlinear panel data. Forthcoming (2020)
Chernozhukov, V., Kasahara, H., Schrimpf, P.: Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. Journal of Econometrics **220**, 23–62 (2021)
Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J.: Double/debiased machine learning for treatment and structural parameters. Econometrics Journal **21** (2018)
Chiang, H.D., Hansen, B.E., Sasaki, Y.: Standard errors for two-way clustering with serially correlated time effects. arXiv preprint arXiv:2201.11304 (2022)
Colella, F., Lalive, R., Sakalli, S.O., Thoenig, M.: Inference with arbitrary clustering. IZA Discussion Paper No. 12584, Available at SSRN: https://ssrn.com/abstract=3449578 (2019)
Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S.: A new framework and software to estimate time-varying reproduction numbers during epidemics. American Journal of Epidemiology **178**, 1505–1512 (2013)
Emmenegger, C., Bühlmann, P.: Plug-in machine learning for partially linear mixed-effects models with repeated measurements. Scandinavian Journal of Statistics (2023). doi:10.1111/sjos.12639
Fattorini, D., Regoli, F.: Role of the chronic air pollution levels in the covid-19 outbreak risk in italy. Environmental Pollution **264**, 114732 (2020)
Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H.: A global panel database of pandemic policies (Oxford covid-19 government response tracker). Nature Human Behaviour **5**, 529–538 (2021)
Hansen, B.: Econometrics. Princeton University Press, 41 William Street, Princeton, New Jersey (2022)
Huber, M., Langen, H.: Timing matters: the impact of response measures on covid-19-related hospitalization and death rates in germany and switzerland. Swiss Journal of Economics and Statistics **156** (2020). doi:10.1186/s41937-020-00054-w
Huisman, J.S., Scire, J., Angst, D.C., Li, J., Neher, R.A., Maathuis, M.H., Bonhoeffer, S., Stadler, T.: Estimation and worldwide monitoring of the effective reproductive number of sars-cov-2. medRxiv, 2020–11 (2022)
Kähler, C.J., Hain, R.: Fundamental protective mechanisms of face masks against droplet infections. Journal of Aerosol Science **148** (2020)
Maloney, W., Taskin, T.: Determinants of Social Distancing and Economic Activity During COVID-19: A Global View, (2020). doi:10.1596/1813-9450-9242
Mitze, T., Kosfeld, R., Rode, J., Walde, K.: Face masks considerably reduce covid-19 cases in Germany. Proceedings of the National Academy of Sciences of the United States of America **117** (2020)
Newey, W.K., West, K.D.: A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica **55**(3), 703–708 (1987)
Pleninger, R., Streicher, S., Sturm, J.-E.: Do covid-19 containment measures work? evidence from Switzerland. Swiss Journal of Economics and Statistics **158**, 5 (2022)
Robins, J.M., Richardson, T.S., Shpitser, I.: An Interventionist Approach to Mediation Analysis. doi:10.1145/3501714.3501754
Rubin, D.B.: Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology **66**, 688–701 (1974). doi:10.1037/h0037350
Wheaton, W.C., Thompson, A.K.: The geography of covid-19 growth in the US: Counties and metropolitan areas. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3570540 (2020)
Zhang, R., Li, Y., Zhang, A.L., Wang, Y., Molina, M.J.: Identifying airborne transmission as the dominant route for the spread of covid-19. Proceedings of the National Academy of Sciences of the United States of America **117** (2020)
Zhu, Y., Xie, J., Huang, F., Cao, L.: Association between short-term exposure to air pollution and covid-19 infection: Evidence from China. Science of the Total Environment **727**, 138704 (2020)
Zoran, M.A., Savastru, R.S., Savastru, D.M., Tautan, M.N.: Assessing the relationship between surface levels of pm2.5 and pm10 particulate matter impact on covid-19 in Milan, Italy. Science of the Total Environment **738**, 139825 (2020)

A Sensitivity Analysis

We perform an extensive sensitivity analysis to investigate the robustness of our results. In the following, we explain the robustness checks conducted and subsequently present the results in Table 2. We restrict the robustness checks to the total effect of the strict facial-mask policy on both response variables and do not consider the direct effect.

A.1 Alterations to Confidence Interval Construction

For the `FE` and `DFE` approaches for point estimation, described in Sections 3.1 and 3.2, we compute the standard errors using one-way clustering on the month instead of the week in Formula (9) (`Month FE` and `Month DFE`) as well as two-way clustering on the canton and the month in Formula (10) (`Canton-Month FE` and `Canton-Month DFE`).

A.2 Alterations to the Data and Point Estimation

The upcoming sections describe changes to the data and point estimation. For all these adaptations we construct the confidence intervals with one method only: For the `FE` and `DFE` approaches we compute the standard errors using the two-way clustering on the canton and week (`Canton-Week`), given in Formula (10), to construct the confidence intervals. For the `RE` approach, we compute the standard errors as described in Section 3.2.

Additional Information Variables

For the `RE` approach and the weekly growth rate of reported new cases as response variable we include additional predictors as motivated by Chernozhukov et al. (2021). To describe them, consider the following new definitions,

$$G'_{i,t} := \ln\left(C_{i,t}\right),$$

$$G_t.\text{nat} := \ln\left(\frac{C_t.\text{nat}}{C_{t-1}.\text{nat}}\right),$$

$$G'_t.\text{nat} := \ln\left(C_t.\text{nat}\right),$$

where $C_{i,t}$ represents the number of new confirmed cases in canton i in week t and C_t.nat represents the national number of new confirmed cases in week t. We use $Y_{i,t} = G_{i,t+2}$, and the corresponding set $\boldsymbol{Y}_{i,t'} = (G_{i,t+1}, G_{i,t}, G'_{i,t}, G_t.\text{nat}, G'_t.\text{nat})$. To the adjustment set $\boldsymbol{Z}_{i,t}$ we also add $\ln\left(T_{i,t}\right)$, where $T_{i,t}$ represents the number Covid-tests performed in canton i in week t, representing an additional information variable. Note that this is only done for the `RE` model as the variables at the national level are omitted in the `FE` and `DFE` approaches through the within transformation.

Half-Cantons

The observations from the half-cantons Basel-Landschaft and Basel-Stadt, Appenzell-Innerrhoden and Appenzell-Ausserrhoden and Obwalden and Nidwalden are combined into one canton, respectively. The response variable `r` is computed by taking the average of the two respective values. We perform the analysis for the `FE`, `DFE` and `RE` approaches.

Timing of Information Variables

We examine the influence of the lag of the information variable that is part of the lagged response variable $Y_{i,t'}$. For the response `r`, we change the lag of the information variable from $t' = 3$ to $t' = 2$, resulting in the lagged response variable $Y_{i,t'} = R_{i,t-2}$. For the response `growth.new.cases`, we change the lag of the information variable from $t' = 2$ to $t' = 3$, resulting in the lagged response variable $Y_{i,t'} = G_{i,t-1}$. We perform the analysis for the `FE`, `DFE` and `RE` approaches.

Outliers

We fit the `FE` approach as described in Section 3.1, compute the the Cook's distance for each observation and exclude observations with a corresponding Cook's distance $> \frac{4}{NT}$. We then refit the model using the `FE` approach based on the reduced sample. In the model where `r` is the response, 28 observations are excluded. When `growth.new.cases` is the response, 27 observations are excluded. Since the calculation of the Cook's distance in the `DFE` and `RE` approaches is not straightforward, we restrict this robustness check to the `FE` model.

	r		growth.new.cases	
	Point Estimate	Confidence Interval	Point Estimate	Confidence Interval
A.1 Alterations to Confidence Interval Construction				
Month FE	-0.15	$[-0.22, -0.09]$	-0.18	$[-0.33, -0.03]$
Canton-Month FE	-0.15	$[-0.26, -0.05]$	-0.18	$[-0.34, -0.03]$
Month DFE	-0.17	$[-0.19, -0.07]$	-0.22	$[-0.37, -0.07]$
Canton-Month DFE	-0.17	$[-0.27, -0.06]$	-0.22	$[-0.37, -0.06]$
A.2 Alterations to the Data and Point Estimation				
Additional Information Variables				
RE	-	-	-0.24	$[-0.42, -0.07]$
Half-Cantons				
FE	-0.16	$[-0.26, -0.06]$	-0.13	$[-0.33, 0.07]$
DFE	-0.25	$[-0.35, -0.15]$	-0.20	$[-0.41, 0.00]$
RE	-0.21	$[-0.33, -0.09]$	-0.25	$[-0.47, -0.04]$
Timing of Information Variables				
FE	-0.18	$[-0.26, -0.09]$	-0.08	$[-0.26, 0.10]$
DFE	-0.28	$[-0.37, -0.20]$	-0.17	$[-0.35, 0.01]$
RE	-0.27	$[-0.46, -0.09]$	-0.18	$[-0.41, 0.05]$
Outliers				
FE	-0.08	$[-0.14, -0.02]$	-0.15	$[-0.28, -0.01]$
Sample Period				
FE	-0.26	$[-0.55, 0.04]$	-0.19	$[-0.43, 0.06]$
Double Machine Learning				
DML	-0.69	$[-1.40, 0.02]$	-0.69	$[-1.51, 0.13]$
Lag-1 Response Variable as Predictor				
FE	-0.11	$[-0.20, -0.03]$	-0.25	$[-0.43, -0.07]$
DFE	-0.13	$[-0.21, -0.04]$	-0.28	$[-0.46, -0.10]$
RE	-0.20	$[-0.30, -0.10]$	-0.44	$[-0.65, -0.22]$
DML	-0.59	$[-1.16, -0.02]$	-0.70	$[-1.54, 0.14]$

Table 2: Results of the sensitivity analyses for both responses `r` and `growth.new.cases`. For the adaptions in the point estimation we construct the 95%-confidence intervals ($CI_{95\%}$) with one method only: For the `FE` and `DFE` approaches we compute the standard errors using the two-way clustering on the canton and week (`Canton-Week`). For the `RE` approach, we compute the standard errors as described in Section 3.2. For the double machine learning approach, we compute the standard error using the two-way clustering on the canton and week (`Canton-Week`). The point estimates in the main analysis regarding the total effect for `r` as response are $(-0.15, -0.17, -0.22)$ for `FE`, `DFE` and `RE` respectively. For `growth.new.cases` as response, the point estimates are $(-0.18, -0.22, -0.28)$ for `FE`, `DFE` and `RE` respectively.

Sample Period
We restrict the period of analysis to the time between August 21, 2020 and October 19, 2020. On August 21, 2020, the canton of Neuchâtel was the first canton to introduce the strict facial-mask policy. On October 19, 2020, the federal government enforced the strict facial-mask policy nationwide. This period is short with only $T = 9$ weeks which constitutes a

problem for the `DFE` and `RE` approach. We thus perform the analysis only for the `FE` approach.

Double Machine Learning Approach
We relax the assumption of a linear regression model to a partially linear regression model, where the effect of the adjustment set $\boldsymbol{Z}_{i,t}$ on $Y_{i,t}$ is non-parametric. We use the adjustment set of the `RE` approach. Estimation is done via the double machine learning framework (Chernozhukov et al., 2018). This approach assumes the following model,

$$\begin{aligned} M_{i,t} &= m(\boldsymbol{Z}_{i,t}) + \nu_{i,t}, \\ Y_{i,t} &= \theta M_{i,t} + g(\boldsymbol{Z}_{i,t}) + \epsilon_{i,t}, \end{aligned}$$

where $\mathbb{E}[\nu_{i,t} \mid \boldsymbol{Z}_{i,t}] = \mathbb{E}[\epsilon_{i,t} \mid M_{i,t}, \boldsymbol{Z}_{i,t}] = 0$. We learn the functions $m(\cdot)$ and $g(\cdot)$ with random forests. See Chernozhukov et al. (2018) for more details. We implement the procedure with the R-package `DoubleML`.

Lag-1 Response Variable as Predictor
In addition to the information variable $Y_{i,t'}$, we consider for both response variables an additional lag of the response variable as predictor. We include the lag-1 response variable $Y_{i,t-1}$ in the models, since for both response variables we observe a possibly non-zero autocorrelation at lag one for some cantons.

A.3 Interpretation of Results from the Sensitivity Analysis

In all cases considered, we obtain a negative point estimate of the treatment effect, ranging from -0.70 to -0.08. The estimate is deemed significantly different from 0 in 25 out of the 35 (71%) supplementary analyses conducted. We will now briefly discuss the non-significant cases.

We start with the analysis for the half-cantons. For the response `growth.new.cases`, we get non-significant results. Note that the approximated weekly growth rate has much higher variance as compared to `r` (0.67 and 0.38 respectively) as the construction of `r` includes more smoothing steps as detailed in Section 2.1. By combining the half-cantons into one respective canton, we reduce the number of observations from $26 \times 24 = 624$ to $23 \times 24 = 552$ weekly observations. We think that the variability of `growth.new.cases` and the reduction of observations are the drivers of this result. Furthermore, the results are very close to being significant (or significant) with upper confidence interval values of $0.07, 0.00$ and -0.04 for the `FE`, `DFE` and `RE`.

For the timing of information variables, the same arguments as for the half-cantons apply. Additionally, note that the `DFE` approach produces more negative estimates compared to the `FE` approach in all the sensitivity analyses. Due to the properties discussed in Section 3.1, the `DFE` approach is more trustworthy.

For the sample period robustness check, we restrict the analysis to the period with the highest heterogeneity in facial-mask policies. The reduction in observations from 624 to 234 is radical and prohibits estimation using the `DFE` and `RE` methodologies. The `FE` approach produces similar point estimates over the full period vs. the short period (-0.15 vs. -0.26) for `r` and (-0.18 vs. -0.19) for `growth.new.cases`. The uncertainty in the estimated treatment effect is larger however in the short period, which is to be expected. Using the data at a daily frequency and estimating the treatment effect for the short periods yields similar point estimates that are deemed highly significant ($-0.21, p = 0.0002$ for `r`) and ($-0.14, p = 0.0076$ for `growth.new.cases`).

Lastly, for the `DML` approach, we get smaller point estimates for the treatment effects with -0.69 for both `r` and `growth.new.cases`. However, the uncertainty is very large, so the results are not significant. Using this methodology, we cannot include random or fixed effects, meaning this difference might be explained by unmeasured confounding variables. There is however new exciting research that synthesizes random effects and fixed effects approaches and double machine learning, see Emmenegger and Bühlmann (2023). As there is no clear pattern apparent in the Tukey-Anscombe plots in Figure 6, we suspect that the difference in point estimates between the linear methods and the non-linear `DML` methodology is mostly driven by the latter's lack to incorporate random effects or fixed effects – and not by an underlying non-linear relationship between the facial-mask policy and the response variables.

B Variance-Covariance Matrix of Estimated Coefficients

Remember that $\hat{\boldsymbol{\eta}} = (\hat{\theta}, \hat{\boldsymbol{\beta}})$ is estimated either via `FE` or `DFE`. We obtain $\widehat{\mathrm{Var}}(\hat{\theta})$ as the first diagonal entry in the estimated variance-covariance matrix of $\hat{\boldsymbol{\eta}}$, $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\eta}})$. To incorporate the panel structure of the data into the estimation of the variance-covariance matrix, we consider the following estimators:

Let in the following

$$\boldsymbol{X}_{i,t} := (M_{i,t}, \boldsymbol{Z}_{i,t})^\top \in \mathbb{R}^{P\times 1}$$

be the observed predictor of canton i and week t, where $P := 1 + |\boldsymbol{Z}_{i,t}|$. Further let

$$\boldsymbol{X} := \begin{pmatrix} \boldsymbol{X}_{1,1}^\top \\ \boldsymbol{X}_{1,2}^\top \\ \vdots \\ \boldsymbol{X}_{1,T}^\top \\ \boldsymbol{X}_{2,1}^\top \\ \vdots \\ \boldsymbol{X}_{N,T}^\top \end{pmatrix} \in \mathbb{R}^{NT\times P}$$

be the stacked predictor matrix. The conditional variance-covariance matrix of $\hat{\boldsymbol{\eta}}$ can be written as

$$\mathrm{Var}(\hat{\boldsymbol{\eta}} \mid \boldsymbol{X}) = \boldsymbol{Q}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}^{-1}, \tag{8}$$

where

$$\boldsymbol{Q} := \frac{1}{NT}\boldsymbol{X}^\top\boldsymbol{X},$$

and

$$\boldsymbol{\Omega} := \frac{1}{(NT)^2}\boldsymbol{X}^\top\mathrm{Var}(\boldsymbol{\epsilon})\boldsymbol{X},$$

with

$$\boldsymbol{\epsilon} := \big(\epsilon_{1,1}, \epsilon_{1,2}, \dots, \epsilon_{1,T}, \epsilon_{2,1}, \dots, \epsilon_{2,T}, \dots, \epsilon_{N,T}\big)^\top.$$

We denote by $\hat{\boldsymbol{\epsilon}}$ the empirical residuals obtained through the `FE` or `DFE` approach. The variance of $\hat{\boldsymbol{\eta}}$ is then estimated by plugging in an estimate of $\boldsymbol{\Omega}$ into Equation (8), resulting in $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\eta}}) = \boldsymbol{Q}^{-1}\hat{\boldsymbol{\Omega}}\boldsymbol{Q}^{-1}$[17].

Subsequently, we present seven estimators for $\boldsymbol{\Omega}$ (with their `short name` in brackets) corresponding to different assumptions on the structure of $\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s})$ for $i, j = 1, \dots, N$ and $t, s = 1, \dots, T$. These assumptions reflect the clustered, heteroskedastic and autocorrelated nature of the error terms.

1) Heteroscedastic-Robust (`HC3`) ($\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s}) \neq 0$ iff $i = j$ and $s = t$):

$$\hat{\boldsymbol{\Omega}}_1 := \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{X}_{i,t}\boldsymbol{X}_{i,t}^\top\tilde{\epsilon}_{i,t}^2,$$

where the residual $\tilde{\epsilon}_{i,t}$ is given by the classical HC3 representation

$$\tilde{\epsilon}_{i,t} := \frac{\hat{\epsilon}_{i,t}}{(1 - \boldsymbol{X}_{i,t}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_{i,t})}.$$

2) One-Way Clustering on Canton (`Canton`) ($\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s}) \neq 0$ iff $i = j$):

$$\hat{\boldsymbol{\Omega}}_2 := \frac{1}{(NT)^2}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^\top,$$

where $\hat{\boldsymbol{R}}_i := \sum_{t=1}^{T}\boldsymbol{X}_{i,t}\hat{\epsilon}_{i,t}$.

3) One-Way Clustering on Week (`Week`) ($\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s}) \neq 0$ iff $t = s$):

$$\hat{\boldsymbol{\Omega}}_3 := \frac{1}{(NT)^2}\sum_{t=1}^{T}\hat{\boldsymbol{S}}_t\hat{\boldsymbol{S}}_t^\top, \tag{9}$$

where $\hat{\boldsymbol{S}}_t := \sum_{i=1}^{N}\boldsymbol{X}_{i,t}\hat{\epsilon}_{i,t}$.

4) Two-Way Clustering on Canton and Week (`Canton-Week`) ($\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s}) \neq 0$ iff $t = s$ or $i = j$):

$$\hat{\boldsymbol{\Omega}}_4 := \frac{1}{(NT)^2}\Bigg(\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^\top + \sum_{t=1}^{T}\hat{\boldsymbol{S}}_t\hat{\boldsymbol{S}}_t^\top - \sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{X}_{i,t}\boldsymbol{X}_{i,t}^\top\hat{\epsilon}_{i,t}^2\Bigg). \tag{10}$$

[17] Note that estimates of $\mathrm{Var}(\hat{\theta})$ based on $\hat{\theta}$ remain valid for the debiased estimator $\hat{\theta}_{BC}$ (Chen et al., 2019).

5) Newey-West (NW) (Newey and West, 1987) ($\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s}) \neq 0$ iff $i = j$, where $\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s})$ is decreasing with $|t-s|$ increasing):

$$\hat{\boldsymbol{\Omega}}_5 := \frac{1}{(NT)^2}\Bigg(\sum_{i=1}^{N} \hat{\boldsymbol{R}}_i \hat{\boldsymbol{R}}_i^\top + \sum_{t=1}^{T} \hat{\boldsymbol{S}}_t \hat{\boldsymbol{S}}_t^\top - \sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{X}_{i,t}\boldsymbol{X}_{i,t}^\top \hat{\epsilon}_{i,t}^2 + \sum_{m=1}^{M} w(m,M)(\hat{\boldsymbol{G}}_m + \hat{\boldsymbol{G}}_m^\top - \hat{\boldsymbol{H}}_m - \hat{\boldsymbol{H}}_m^\top)\Bigg), \tag{11}$$

where $\hat{\boldsymbol{G}}_m := \sum_{t=1}^{T-m} \hat{\boldsymbol{S}}_t \hat{\boldsymbol{S}}_{t+m}^\top$, $\hat{\boldsymbol{H}}_m := \sum_{i=1}^{N}\sum_{t=1}^{T-m} \boldsymbol{X}_{i,t}\hat{\epsilon}_{i,t}\boldsymbol{X}_{i,t+m}^\top \hat{\epsilon}_{i,t+m}$, $w(m,M) = 1 - m/(M+1)$ are triangular weights and $M = \lfloor T^{1/4} \rfloor$.

6) Chiang-Hansen (CH) (Chiang et al., 2022) ($\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s}) \neq 0$, where $\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s})$ for arbitrary $i \neq j$ is decreasing with $|t-s|$ increasing):
The estimator $\hat{\boldsymbol{\Omega}}_6$ is given by Equation (11), where $w(m,M)$ are the triangular weights as in NW and M is data driven.

7) Informal Own Specification (Own) (motivated by Colella et al., 2019) ($\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s}) \neq 0$ if $i = j$ or i and j are neighboring cantons, where $\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s})$ is decreasing with $|t-s|$ increasing):

$$\hat{\boldsymbol{\Omega}}_7 := \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} \omega_{itjs} V_{itjs},$$

where

$$V_{itjs} := \boldsymbol{X}_{i,t}\hat{\epsilon}_{i,t}\hat{\epsilon}_{j,s}\boldsymbol{X}_{j,s}^\top,$$

and the weights ω_{itjs} specify the dependence between two error terms $\epsilon_{i,t}$ and $\epsilon_{j,s}$ and are given by

$$\omega_{itjs} := \left\{ \begin{array}{lr} 1, & i=j,\ t=s, \\ \lambda_{ij} 0.5^{|t-s|}, & \text{otherwise} \end{array} \right\},$$

and

$$\lambda_{ij} := \left\{ \begin{array}{lr} 1, & i=j, \\ 0.5, & i,j \text{ neighbors}, \\ 0, & \text{otherwise} \end{array} \right\}.$$

C Adaptation of Strict Facial-Mask Policy

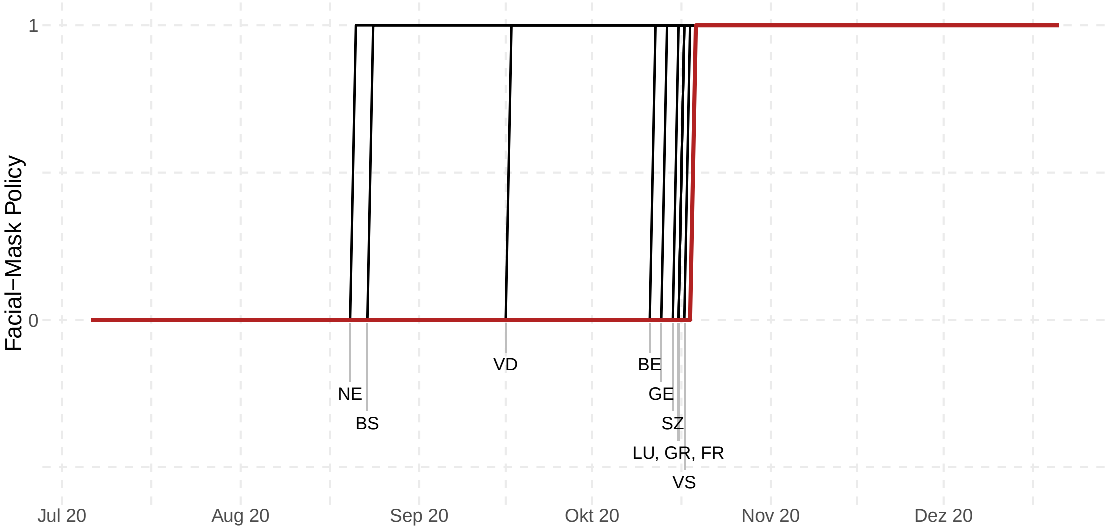

Figure 5: Plot of adaptation of strict facial-mask policy: A value of 0 corresponds to the government-determined baseline policy while a value of 1 indicates a strict facial-mask policy as described in Subsection 2.2. The red line denotes the government-determined baseline policy.

D Tukey-Anscombe Plots (Residuals v.s. Fitted Values)

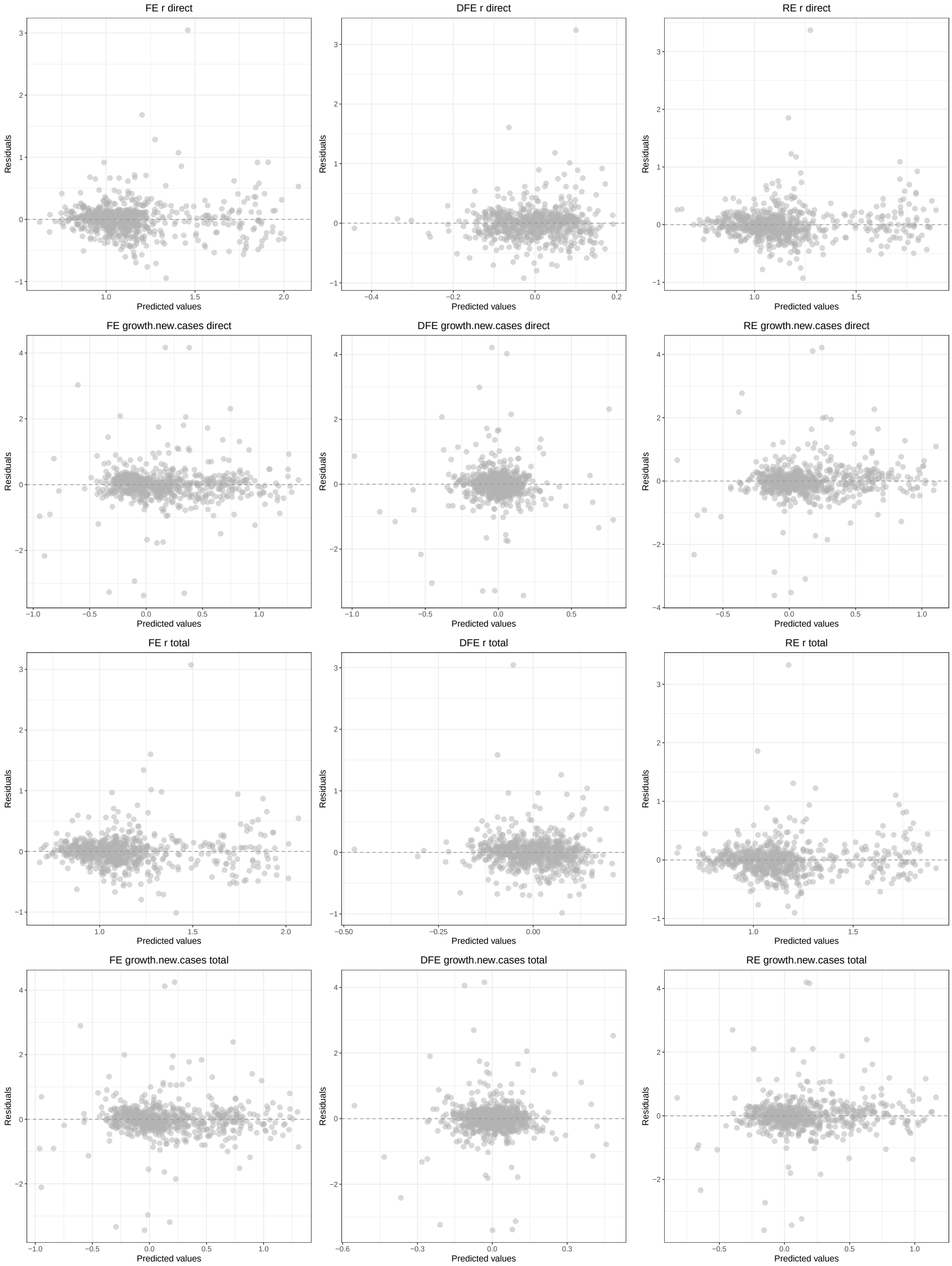

Figure 6: Tukey-Anscombe plots (residuals v.s. fitted values) for all 12 modeling approaches: both direct and total effect for both response variables with the three model approaches `FE`, `DFE` and `RE`, as described in Section 3.