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

Annual Congress of the Swiss Society of Economics and Statistics

---

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

June 30, 2023

[1]Seminar for Statistics, ETHZ

[2]Department of Economics, UZH

Table of contents

Introduction

Introduction I: Motivation

- Heatedly debated policy: How effective were facial-masks in reducing the spread of COVID-19 in Switzerland?[1]
- Heterogeneity in facial-mask policies from July 2020 to October 2020 at the cantonal level due to two different policies:
  - *Baseline policy*: mandatory facial-mask wearing on public transport.
  - *Strict policy*: mandatory facial-mask wearing on public transport and mandatory facial-mask wearing in all public or shared spaces where social distancing is not possible.

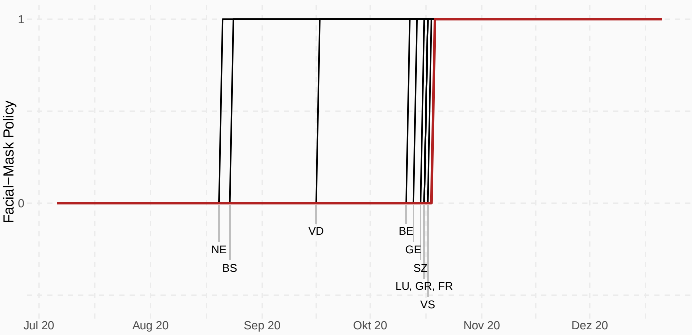
[1] https://www.swissinfo.ch/eng/politics/cover-up–how-shifting-policies-affect-swiss-attitudes-toward-masks/45978462.

Introduction I: Motivation

- Heatedly debated policy: How effective were facial-masks in reducing the spread of COVID-19 in Switzerland?[1]
- Heterogeneity in facial-mask policies from July 2020 to October 2020 at the cantonal level due to two different policies:
  - *Baseline policy*: mandatory facial-mask wearing on public transport.
  - *Strict policy*: mandatory facial-mask wearing on public transport and mandatory facial-mask wearing in all public or shared spaces where social distancing is not possible.

[1] https://www.swissinfo.ch/eng/politics/cover-up–how-shifting-policies-affect-swiss-attitudes-toward-masks/45978462.

Introduction II: Research Question

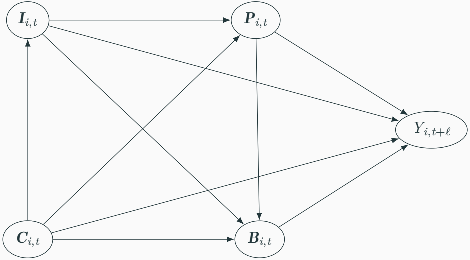


- Research question: How effective was the introduction of the strict facial-mask policy relative to the country-wide baseline facial-mask policy?

Causal Assumptions, Data & Estimation

Causal Assumptions I

- Following Chernozhukov et al. [2021], we assume the following structural causal model describing the dynamics of COVID-19 under policy interventions:



where

- $\boldsymbol{I}_{i,t}$ are information variables
- $\boldsymbol{P}_{i,t}$ are the COVID-19 related policies
- $\boldsymbol{C}_{i,t}$ are confounding factors
- $\boldsymbol{B}_{i,t}$ are behaviour variables
- $Y_{i,t+\ell}$ is the forward health outcome

Causal Assumptions II

- This causal ordering corresponds to the following timing of events:
  - Information and confounders are determined at time t.
  - Policies are determined given information and confounders at time t.
  - Behaviour is decided given policies, information and confounders at time t.
  - The health outcome is settled at time $t+\ell$ given behaviour, policies, information and confounders.

Causal Assumptions II

- This causal ordering corresponds to the following timing of events:
  - Information and confounders are determined at time t.
  - Policies are determined given information and confounders at time t.
  - Behaviour is decided given policies, information and confounders at time t.
  - The health outcome is settled at time $t+\ell$ given behaviour, policies, information and confounders.

Causal Assumptions II

- This causal ordering corresponds to the following timing of events:
  - Information and confounders are determined at time t.
  - Policies are determined given information and confounders at time t.
  - Behaviour is decided given policies, information and confounders at time t.
  - The health outcome is settled at time $t+\ell$ given behaviour, policies, information and confounders.

Causal Assumptions II

- This causal ordering corresponds to the following timing of events:
    - Information and confounders are determined at time t.
    - Policies are determined given information and confounders at time t.
    - Behaviour is decided given policies, information and confounders at time t.
    - The health outcome is settled at time $t+\ell$ given behaviour, policies, information and confounders.

Causal Assumptions & Data I

- We instantiate the structural causal model by assigning measured variables to $(\boldsymbol{I}_{i,t}, \boldsymbol{P}_{i,t}, \boldsymbol{C}_{i,t}, \boldsymbol{B}_{i,t}, Y_{i,t+\ell})$ by considering previous research[2].
- Using the measured variables, we refine the structural causal model presented previously while neither changing the overall idea nor the implications.

[2] Chernozhukov et al. [2021], Huisman et al. [2022], Pleninger et al. [2022], Zoran et al. [2020], Zhu et al. [2020], Fattorini and Regoli [2020], Hale et al. [2021].

Causal Assumptions & Data I

- We instantiate the structural causal model by assigning measured variables to $(\boldsymbol{I}_{i,t}, \boldsymbol{P}_{i,t}, \boldsymbol{C}_{i,t}, \boldsymbol{B}_{i,t}, Y_{i,t+\ell})$ by considering previous research[2].
- Using the measured variables, we refine the structural causal model presented previously while neither changing the overall idea nor the implications.

[2] Chernozhukov et al. [2021], Huisman et al. [2022], Pleninger et al. [2022], Zoran et al. [2020], Zhu et al. [2020], Fattorini and Regoli [2020], Hale et al. [2021].

Causal Assumptions & Data II

- Forward health outcomes:
  - Effective reproductive number $R_{i,t+\ell}$
  - Approximated weekly growth rate of new cases $G_{i,t+\ell}$
- Policies:
  - Strict facial mask policy variable $M_{i,t}$
  - Non-pharmaceutical policy variables $\boldsymbol{P}_{i,t}$
    - Work closings
    - School closings
    - Restriction of gatherings
    - Cancellation of events
    - Testing policy
- Behaviour:
  - Approximated weekly growth rate of household spending $B_{i,t}$

Descriptive Statistics of Data

Causal Assumptions & Data II

- Forward health outcomes:
  - Effective reproductive number $R_{i,t+\ell}$
  - Approximated weekly growth rate of new cases $G_{i,t+\ell}$
- Policies:
  - Strict facial mask policy variable $M_{i,t}$
  - Non-pharmaceutical policy variables $\boldsymbol{P}_{i,t}$
    - Work closings
    - School closings
    - Restriction of gatherings
    - Cancellation of events
    - Testing policy
- Behaviour:
  - Approximated weekly growth rate of household spending $B_{i,t}$

Descriptive Statistics of Data

Causal Assumptions & Data II

- Forward health outcomes:
  - Effective reproductive number $R_{i,t+\ell}$
  - Approximated weekly growth rate of new cases $G_{i,t+\ell}$
- Policies:
  - Strict facial mask policy variable $M_{i,t}$
  - Non-pharmaceutical policy variables $\boldsymbol{P}_{i,t}$
    - Work closings
    - School closings
    - Restriction of gatherings
    - Cancellation of events
    - Testing policy
- Behaviour:
  - Approximated weekly growth rate of household spending $B_{i,t}$

Descriptive Statistics of Data

Causal Assumptions & Data III

- Information:
  - Publicly available information about COVID-19 in week t
    - Lagged effective reproductive number $R_{i,t}$
    - Lagged approximated weekly growth rate of new cases $G_{i,t}$
- Confounding factors:
  - Demographic variables $\boldsymbol{D}_i$
    - Population size
    - Percentage over 80 years old
    - Population density in settlement area
  - Holiday indicator $H_{i,t}$
  - Meteorological variables $\boldsymbol{W}_{i,t}$
    - Minutes of sunshine
    - Air temperature
    - Relative humidity

Descriptive Statistics of Data

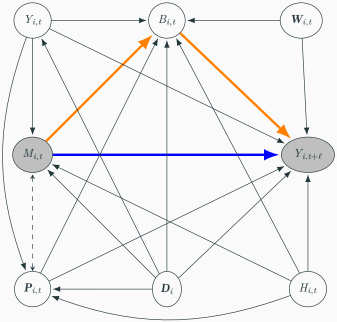
Causal Assumptions & Data III

- Information:
  - Publicly available information about COVID-19 in week t
    - Lagged effective reproductive number $R_{i,t}$
    - Lagged approximated weekly growth rate of new cases $G_{i,t}$
- Confounding factors:
  - Demographic variables $\boldsymbol{D}_i$
    - Population size
    - Percentage over 80 years old
    - Population density in settlement area
  - Holiday indicator $H_{i,t}$
  - Meteorological variables $\boldsymbol{W}_{i,t}$
    - Minutes of sunshine
    - Air temperature
    - Relative humidity

Descriptive Statistics of Data

Causal Assumptions & Data IV

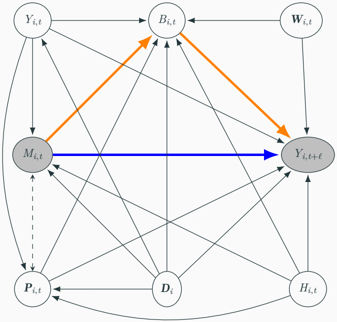
- This translates into the following, more refined causal graph:



- Note that this allows the decomposition of the effect of the strict facial-mask policy on the response variable into two channels:
  - Direct effect: $M_{i,t} \longrightarrow Y_{i,t+\ell}$
  - Indirect effect via behaviour: $M_{i,t} \longrightarrow B_{i,t} \longrightarrow Y_{i,t+\ell}$

Causal Assumptions & Data IV

- This translates into the following, more refined causal graph:



- Note that this allows the decomposition of the effect of the strict facial-mask policy on the response variable into two channels:
  - Direct effect: $M_{i,t} \longrightarrow Y_{i,t+\ell}$
  - Indirect effect via behaviour: $M_{i,t} \longrightarrow B_{i,t} \longrightarrow Y_{i,t+\ell}$

Causal Assumptions & Estimation I

- In addition to the structural causal model from the last slide, we also impose canton and week-specific effects (α_i and γ_t respectively).
- This implies the following estimating equation:

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

where $\boldsymbol{Z}_{i,t}$ is a valid adjustment set relative to the effect of interest (direct or total).

Causal Assumptions & Estimation I

- In addition to the structural causal model from the last slide, we also impose canton and week-specific effects (α_i and γ_t respectively).
- This implies the following estimating equation:

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

  where $\boldsymbol{Z}_{i,t}$ is a valid adjustment set relative to the effect of interest (direct or total).

Causal Assumptions & Estimation II

- We estimate the parameters $(\theta, \boldsymbol{\beta}) := \boldsymbol{\eta}$ of

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

  by using

  - Fixed effects estimators
  - Debiased fixed effects estimators
  - Random effects estimators
  - Double machine learning

  while putting special consideration on the panel structure of the data while computing $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\eta}})$.

- Note that the different estimation schemes require different assumptions on the error components $(\alpha_i, \gamma_t, \epsilon_{i,t})$, see Hansen [2022].

Debiased Fixed Effects Estimation | Double Machine Learning Estimation | Computing $\widehat{\mathbf{Var}}(\hat{\boldsymbol{\eta}})$

Causal Assumptions & Estimation II

- We estimate the parameters $(\theta, \boldsymbol{\beta}) \coloneqq \boldsymbol{\eta}$ of

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

  by using

  - Fixed effects estimators
  - Debiased fixed effects estimators
  - Random effects estimators
  - Double machine learning

  while putting special consideration on the panel structure of the data while computing $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\eta}})$.
- Note that the different estimation schemes require different assumptions on the error components $(\alpha_i, \gamma_t, \epsilon_{i,t})$, see Hansen [2022].

Debiased Fixed Effects Estimation | Double Machine Learning Estimation | Computing $\widehat{\mathbf{Var}}(\hat{\boldsymbol{\eta}})$

Results

Results I

- We report the treatment effect of the strict facial-mask policy $\hat{\theta}$ along with it's associated 95%–confidence interval
  - For the direct and the total (=indirect+direct) effect
  - For the fixed effects approach, debiased fixed effects approach, random effects approach[3]
  - For both responses of (i) effective reproductive number and (ii) approximated weekly growth rate of new cases
  - For the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$

[3] We report the results from the double machine learning in the sensitivity analysis.

Results I

- We report the treatment effect of the strict facial-mask policy $\hat{\theta}$ along with it's associated 95%–confidence interval
  - For the direct and the total (=indirect+direct) effect
  - For the fixed effects approach, debiased fixed effects approach, random effects approach[3]
  - For both responses of (i) effective reproductive number and (ii) approximated weekly growth rate of new cases
  - For the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$

[3] We report the results from the double machine learning in the sensitivity analysis.

Results I

- We report the treatment effect of the strict facial-mask policy $\hat{\theta}$ along with it‘s associated 95%–confidence interval
  - For the direct and the total (=indirect+direct) effect
  - For the fixed effects approach, debiased fixed effects approach, random effects approach[3]
  - For both responses of (i) effective reproductive number and (ii) approximated weekly growth rate of new cases
  - For the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$

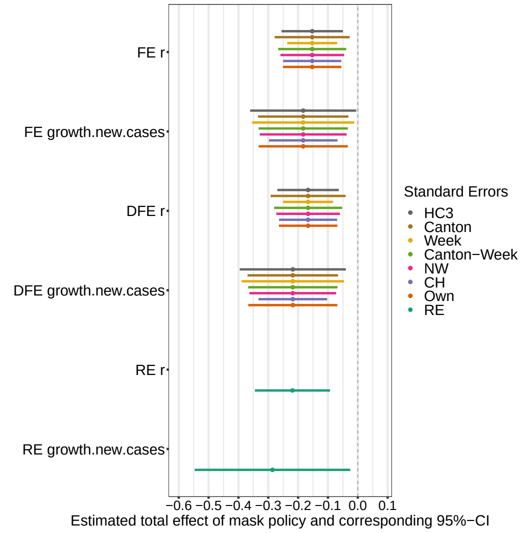
[3] We report the results from the double machine learning in the sensitivity analysis.

Results I

- We report the treatment effect of the strict facial-mask policy $\hat{\theta}$ along with it‘s associated 95%–confidence interval
  - For the direct and the total (=indirect+direct) effect
  - For the fixed effects approach, debiased fixed effects approach, random effects approach[3]
  - For both responses of (i) effective reproductive number and (ii) approximated weekly growth rate of new cases
  - For the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$

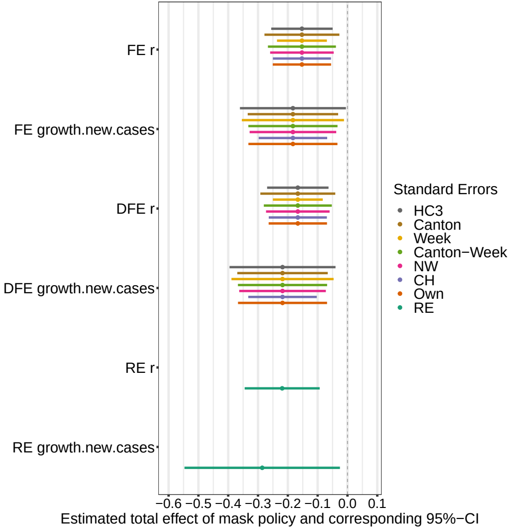
[3] We report the results from the double machine learning in the sensitivity analysis.

Results II: Total Effect



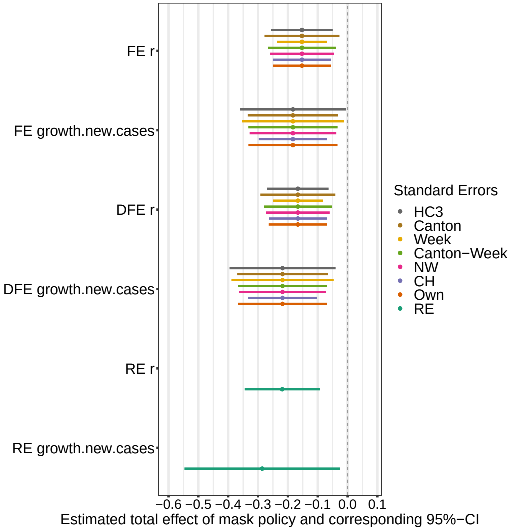
- FE denotes fixed effects, DFE denotes debiased fixed effects and RE denotes random effects.
- The different colours represent the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$.
- The point estimates lie within $[-0.29, -0.15]$.
- θ is deemed significantly different from 0 at $\alpha = 0.05$ in all cases.

Results II: Total Effect



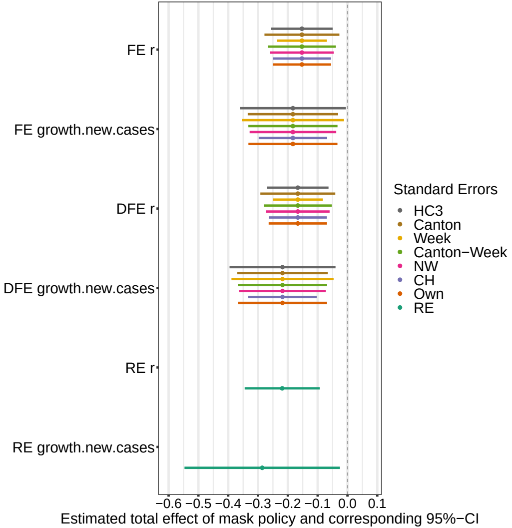
- FE denotes fixed effects, DFE denotes debiased fixed effects and RE denotes random effects.
- The different colours represent the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$.
- The point estimates lie within $[-0.29, -0.15]$.
- θ is deemed significantly different from 0 at $\alpha = 0.05$ in all cases.

Results II: Total Effect



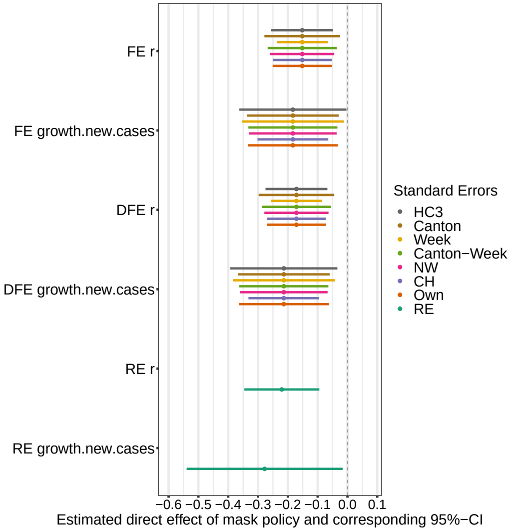
- FE denotes fixed effects, DFE denotes debiased fixed effects and RE denotes random effects.
- The different colours represent the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$.
- The point estimates lie within $[-0.29, -0.15]$.
- θ is deemed significantly different from 0 at $\alpha = 0.05$ in all cases.

Results II: Total Effect



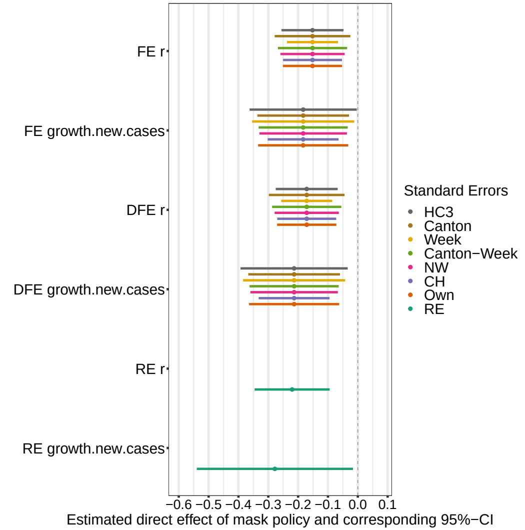
- FE denotes fixed effects, DFE denotes debiased fixed effects and RE denotes random effects.
- The different colours represent the different estimators of $\widehat{\mathrm{Var}}(\hat{\theta})$.
- The point estimates lie within $[-0.29, -0.15]$.
- θ is deemed significantly different from 0 at $\alpha = 0.05$ in all cases.

Results III: Direct Effect



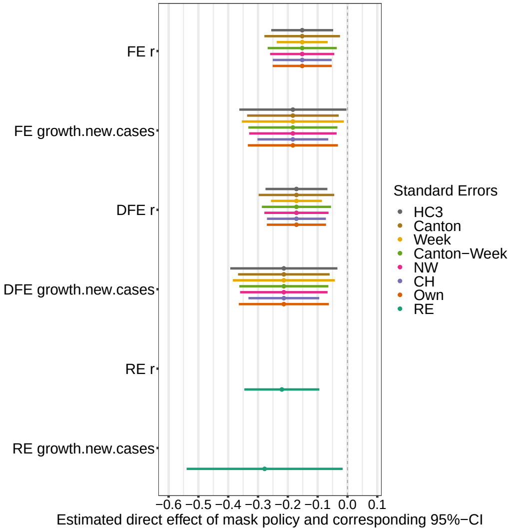
- The point estimates lie within $[-0.28, -0.15]$.
- Again, θ is deemed significantly different from 0 at $\alpha = 0.05$ in all cases.
- The difference between the total and the direct effect is not significant in any of the cases.
  - This implies that we do not find evidence for an indirect effect via behaviour.

Results III: Direct Effect



- The point estimates lie within $[-0.28, -0.15]$.
- Again, θ is deemed significantly different from 0 at $\alpha = 0.05$ in all cases.
- The difference between the total and the direct effect is not significant in any of the cases.
  - This implies that we do not find evidence for an indirect effect via behaviour.

Results III: Direct Effect



- The point estimates lie within $[-0.28, -0.15]$.
- Again, θ is deemed significantly different from 0 at $\alpha = 0.05$ in all cases.
- The difference between the total and the direct effect is not significant in any of the cases.
  - This implies that we do not find evidence for an indirect effect via behaviour.

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Results IV: Sensitivity Analysis

- To check the robustness of our findings with respect to inevitable design choices, we perform the following sensitivity analyses:
  - Further alterations in confidence interval constructions
  - Extended set of information variables
  - Join half-cantons into one respective canton
  - Timing of information variables
  - Outliers
  - Sample period
  - Double machine learning estimation
  - Lag-1 response variable as further predictor
- Again, all the point estimates $\hat{\theta}$ are negative and deemed significantly different from 0 in 25 of the 35 (71%) sensitivity analyses conducted.

Results Sensitivity Analysis

Discussion

Discussion I: Limitations

- Limitation of capturing COVID-19 relevant behaviour using approximated weekly growth rate of household spending $B_{i,t}$.
- As always with observational studies: cautious interpretation of results due to potential unmeasured confounding variables.
- Results hold conditional on the specific time period between July and December 2020 when the Alpha variant of COVID-19 was dominant and no vaccinations were available yet.

Discussion I: Limitations

- Limitation of capturing COVID-19 relevant behaviour using approximated weekly growth rate of household spending $B_{i,t}$.
- As always with observational studies: cautious interpretation of results due to potential unmeasured confounding variables.
- Results hold conditional on the specific time period between July and December 2020 when the Alpha variant of COVID-19 was dominant and no vaccinations were available yet.

Discussion I: Limitations

- Limitation of capturing COVID-19 relevant behaviour using approximated weekly growth rate of household spending $B_{i,t}$.
- As always with observational studies: cautious interpretation of results due to potential unmeasured confounding variables.
- Results hold conditional on the specific time period between July and December 2020 when the Alpha variant of COVID-19 was dominant and no vaccinations were available yet.

Discussion II: Policy Implications

- Our finding agree with previous research on the effectiveness of facial-masks such as Lyu and Wehby [2020], Chernozhukov et al. [2021] or the evidence review study of Howard et al. [2021].
- Considering that the authors deem facial-mask regulations non-invasive compared to other policies[4], we advocate this strategy for potential future pandemics.

[4] E.g. work closings, school closings, restrictions of gatherings, cancellation of events or even (partial) lockdowns.

Discussion II: Policy Implications

- Our finding agree with previous research on the effectiveness of facial-masks such as Lyu and Wehby [2020], Chernozhukov et al. [2021] or the evidence review study of Howard et al. [2021].
- Considering that the authors deem facial-mask regulations non-invasive compared to other policies[4], we advocate this strategy for potential future pandemics.

[4] E.g. work closings, school closings, restrictions of gatherings, cancellation of events or even (partial) lockdowns.

Discussion III: Questions?

- Thank you very much for listening! Are there any open questions?

References

Victor Chernozhukov, Hiroyuki Kasahara, and Paul Schrimpf. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. *Journal of Econometrics*, 220:23–62, 2021. ISSN 18726895.

Jana S. Huisman, Jérémie Scire, Daniel C. Angst, Jinzhou Li, Richard A. Neher, Marloes H. Maathuis, Sebastian Bonhoeffer, and Tanja Stadler. Estimation and worldwide monitoring of the effective reproductive number of sars-cov-2. *eLife*, 11, 2022. ISSN 2050084X. doi: 10.7554/ELIFE.71345.

Regina Pleninger, Sina Streicher, and Jan Egbert Sturm. Do covid-19 containment measures work? evidence from switzerland. *Swiss Journal of Economics and Statistics*, 158, 2022. ISSN 22356282. doi: 10.1186/s41937-022-00083-7.

Maria A. Zoran, Roxana S. Savastru, Dan M. Savastru, and Marina N. Tautan. 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:Article 139825, 2020. ISSN 18791026.

Yongjian Zhu, Jingui Xie, Fengming Huang, and Liqing Cao. Association between short-term exposure to air pollution and covid-19 infection: Evidence from China. *Science of the Total Environment*, 727: Article 138704, 2020. ISSN 18791026.

Daniele Fattorini and Francesco Regoli. Role of the chronic air pollution levels in the covid-19 outbreak risk in italy. *Environmental Pollution*, 264:Article 114732, 2020. ISSN 18736424.

Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. A global panel database of pandemic policies (Oxford covid-19 government response tracker). *Nature Human Behaviour*, 5:529–538, 2021. ISSN 23973374.

Bruce Hansen. *Econometrics*. Princeton University Press, 41 William Street, Princeton, New Jersey, 2022.

W Lyu and G L Wehby. Community use of face masks and covid-19: Evidence from a natural experiment of state mandates in the us: Study examines impact on covid-19 growth rates .... *Health affairs*, 2020.

Jeremy Howard, Austin Huang, Zhiyuan Li, Zeynep Tufekci, Vladimir Zdimal, Helene Mari van der Westhuizen, Arne von Delft, Amy Price, Lex Fridman, Lei Han Tang, Viola Tang, Gregory L. Watson, Christina E. Bax, Reshama Shaikh, Frederik Questier, Danny Hernandez, Larry F. Chu, Christina M. Ramirez, and Anne W. Rimoin. An evidence review of face masks against covid-19. *Proceedings of the National Academy of Sciences of the United States of America*, 118, 2021. ISSN 10916490. doi: 10.1073/pnas.2014564118.

Shuowen Chen, Victor Chernozhukov, and Iván Fernández-Val. Mastering panel metrics: Causal impact of democracy on growth. *AEA Papers and Proceedings*, 109, 2019. ISSN 2574-0768.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21, 2018. ISSN 1368423X.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/1913610`.

Harold D Chiang, Bruce E Hansen, and Yuya Sasaki. Standard errors for two-way clustering with serially correlated time effects. *arXiv preprint arXiv:2201.11304*, 2022.

Fabrizio Colella, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. Inference with arbitrary clustering. *IZA Discussion Paper No. 12584, Available at SSRN: `https://ssrn.com/abstract=3449578`*, 2019.

Data Description

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 $°\mathrm{C}$	$(11.36, 11.89, 7.15)$	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

Back

Sensitivity Analysis

	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	$[-1.54, 0.14]$

Back

Details of Debiased Fixed Effects Estimation i

- Consider estimating the parameters $\boldsymbol{\eta} := (\theta, \boldsymbol{\beta})$ of

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

  using the fixed effects methodology, where $\boldsymbol{Z}_{i,t}$ contains lagged instances of $Y_{i,t+\ell}$, being our information variables $I_{i,t} = Y_{i,t}$.
- Applying the two-way within operator $u_{i,t} \longrightarrow \ddot{u}_{i,t}$ removes the fixed effects yielding

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

  with the issue that $\mathbb{E}[\ddot{Y}_{i,t}\ddot{\epsilon}_{i,t}] \neq 0$ as both are functions of the entire time series.
- This creates a bias of $\mathcal{O}(1/T)$ in estimating $\boldsymbol{\eta}$, which is sometimes deemed negligible for $T \geq 30$ (we have $T = 24$), see Hansen [2022].

Details of Debiased Fixed Effects Estimation ii

- We consider a sample-splitting remedy for this issue, where we repeatedly split the data along:

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

  and

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

  where we permute ($p = 1\ldots, P$) the ordering of the cantons as there is no natural ordering. We then estimate $\hat{\theta}_{S_1,p}, \hat{\theta}_{S_2,p}$ disjointly and combine them via

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

Details of Debiased Fixed Effects Estimation iii

- The final estimate over the $P = 20$ permutations of the cantons is given by

$$\hat{\theta}_{BC} = \frac{1}{P} \sum_{p=1}^{P} \hat{\theta}_{BC,p}.$$

- Intuition for why this works: Given that the bias term is homogeneous across the samples, the combination operation to compute $\hat{\theta}_{BC,p}$ reduces the first-order bias by subtraction, see Chen et al. [2019].

◂ Back

Details of Double Machine Learning i

- We relax the assumption of linearity by imposing the partially linear model given by:

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

  where $m(\cdot)$ and $g(\cdot)$ are potentially non-linear functions that we learn using random forests.
- We perform robust inference for θ using two-way clustered standard errors, see Chernozhukov et al. [2018] for more details.

◂ Back

- Remember that $\hat{\boldsymbol{\eta}} = (\hat{\theta}, \hat{\boldsymbol{\beta}})$ is estimated either via the fixed effects or the debiased fixed effects estimator.
- We obtain $\widehat{\mathrm{Var}}(\hat{\theta})$ as the first diagonal entry in the estimated variance-covariance matrix of $\hat{\boldsymbol{\eta}}$, $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\eta}})$.
- Let in the following

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

  be the observed predictor of canton i and week t, where $P \coloneqq 1 + |\boldsymbol{Z}_{i,t}|$. Then, let $\boldsymbol{X} \in \mathbb{R}^{NT \times P}$ be the stacked predictor matrix.

Variance-Covariance Matrix of Estimated Coefficients ii

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

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

  where

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

  and

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

  with

$$\boldsymbol{\epsilon} := \left(\epsilon_{1,1}, \epsilon_{1,2}, \ldots, \epsilon_{1,T}, \epsilon_{2,1}, \ldots, \epsilon_{2,T}, \ldots, \epsilon_{N,T}\right)^\top.$$

- We denote by $\hat{\boldsymbol{\epsilon}}$ the empirical residuals. The variance of $\hat{\boldsymbol{\eta}}$ is then estimated by plugging in an estimate of $\boldsymbol{\Omega}$ into Equation (1), resulting in $\widehat{\operatorname{Var}}(\hat{\boldsymbol{\eta}}) = \boldsymbol{Q}^{-1} \hat{\boldsymbol{\Omega}} \boldsymbol{Q}^{-1}$.

- To incorporate the panel structure of the data, we consider 7 different estimators of $\mathbf{\Omega}$, corresponding to different assumptions on $\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,s})$ for $i, j = 1, \ldots, N$ and $j, s = 1, \ldots T$. These assumptions reflect the clustered, heteroskedastic and autocorrelated nature of the errors.

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

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

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

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

Variance-Covariance Matrix of Estimated Coefficients iv

(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 $t = s$:

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

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 $i = j$:

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

(5) Newey-West (*NW*), see 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 in $|t - s|$:

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

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

(6) Chiang-Hansen (*CH*), see 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 in $|t - s|$:
The estimator $\hat{\boldsymbol{\Omega}}_6$ is the same as $\hat{\boldsymbol{\Omega}}_5$, where $w(m, M)$ are the triangular weights as in *NW* and M is data driven.

Variance-Covariance Matrix of Estimated Coefficients viii

(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 in $|t - s|$:

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

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

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

and

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

Back

Comparison of Response Variables