Disproportionate policy dynamics in crisis and uncertainty: an international comparative analysis of policy responses to COVID-19[*]

Replication Project

Submitted: 15 April 2024

Can we assess the impacts of policy responses on pandemic outcomes?

Shafi and Mallinson (2023)[1] employ **pooled panel linear regression** using **weekly time series data for 160 countries from 27 Jan to 27 Dec 2020** to explore the dynamics between state responses during the first year of the COVID-19 pandemic and reported cases of and deaths from the virus. The paper aims to fill existing knowledge gaps in the literature around the *dynamics* and the *outcomes* of different policy approaches across countries.

The authors' findings indicate that increased income supports and debt relief are associated with fewer COVID-19 deaths up to four weeks, and potentially beyond, while "broad policy interventions" across all economic, health, and social dimensions are associated with short-term reductions in the weekly rate of deaths.

This paper is an excellent case study for replication as it provides largely thorough and readable data and code for reproducing results in R, alongside documentation for the secondary data used. However, the published paper and provided code include errors of a minor and serious nature, which are highlighted in a dedicated section of this report. Also note that the numbering of figures in this paper does not consistently match those in the original report - the title will remain the same for easier reference between the two documents.

I chose this specific paper for replication as it is thematically relevant to 2024's international policy context, and comparative policy analysis more broadly, as well demonstrates advanced

---

[*]Shafi, S. and Mallinson, D. J. (2022) "Replication Data for: Disproportionate Policy Dynamics in Crisis and Uncertainty: An International Comparative Analysis of Policy Responses to COVID-19", Harvard Dataverse, V1. Available at: https://doi.org/10.7910/DVN/MABPEC (Accessed 15 April 2024)

[1]Shafi, S. and Mallinson, D. J. (2023) "Disproportionate Policy Dynamics in Crisis and Uncertainty: An International Comparative Analysis of Policy Responses to COVID-19", Policy Studies, 44 (1), pp. 09-111. Available at: https://doi.org/10.1080/01442872.2022.2053093 (Accessed 15 April 2024)

1

methodology which is relevant to my own interests in comparative policy analysis and data-informed policy research.

Errors in the Paper and Data

There is a numbering typo error on line 400 which I corrected. Further, the code for Figures 4 and 5 is absent, most likely an accidental exclusion. As they are not essential plots for the overall discussion or for replication purposes, I did not attempt to reproduce these plots.

In the paper and replication code, the first figure erroneously plots the Stringency Index for France in all four graphs, where each plot should show the country's progression across the Containment Health, Government Response, and Economic Support indices respectively. While this does not impact the overall findings of the paper, I show the published and the corrected plots side-by-side in Figure 1 below. Additionally, the original plotting of the results for the GRI models includes an error in the x-axis of the fourth subplot for "New Weekly Deaths". This is also a minor graphical error and I do not correct it when comparing the results from the original paper and the updated models.

More seriously, the replicated results of the Hausman Tests, which assess if the fixed-effect models are more appropriate than the random-effect models, are contrary to what the paper reports - regarding the GRI model presented by the authors for discussion, one result is consistent, while the remaining three are not.

Further, within the replication code, these tests are only conducted for the `Government Response Index`, and not for the `Stringency`, `Economic Support` nor `Containment Health` indices - this is despite the authors explicitly reporting that these tests were conducted and returned overall conclusive results supporting their argument "that a fixed effects model was the most efficient, and thus preferred, model ($p < 0.05$)" (Shafi and Mallinson, 2023, p. 98). Therefore, alongside the contradictory results from what is included in the replication code, the exclusion of these other tests' code is problematic.

However, due to resource constraints, I do not perform these tests on the original models and instead follow the remaining steps of the authors' analysis in order to demonstrate the impacts of my "twist". Note I have included the raw code below as the `stargazer` package in R does not support Heusman Test tables as of the time of writing.

```
1  #Hausman Test to see if RE improve the model
2  #H0: Random Effects model consistent and efficient (preferred)
3  #H1: Fixed effects model consistent and preferred
4  phtest(gri1.fe, gri1.re) #"FE is preferred", and statistically significant
5  phtest(gri2.fe, gri2.re) #"FE is preferred", yet not statistically significant
6  phtest(gri3.fe, gri3.re) #"FE is preferred", yet not statistically significant
7  phtest(gri4.fe, gri4.re) #"RE is preferred", yet statistically significant
```

What is the data used?

For data on COVID-19 policy responses, this paper uses four indices from the *Oxford COVID-19 Government Response Tracker* (OxCGRT)[2], with minor modifications for the purposes of the study:

- Stringency index

- Government response index

- Economic support index

- Containment health index

These are 'nested' indices, meaning each index uses a subset of the same 23 indicators of government policy responses to the COVID-19 pandemic. As illustrated in Figure 1 (see in the next section), this leads to similar trends across all four indices internationally.

The paper also draws from *World Development Indicators* (WDI) datasets from the World Bank[3] as well as the World Health Organisation[4] and the World Meters platform [5] for control variables, including total population, total population aged 65 and over, population density, and GDP per capita.

Once loaded, the authors cleaned and modified/transformed the data for analysis, for example, normalising population measures for better comparability across countries of different sizes, and log transformations on measures of fiscal and emergency investment measures as their distributions skewed to the right. All modifications are specified included within the replication code.

What are the research questions?

While the authors acknowledge the preliminary nature of the analysis, the empirical validity of the paper's results remains questionable - in addition to the aforementioned inconsistent results supporting the methodology, the manuscript does not explicitly state nor test

---

[2]Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow (2021) "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)." Nature Human Behaviour. Available at: https://doi.org/10.1038/s41562-021-01079-8 or https://github.com/OxCGRT/covid-policy-tracker

[3]World Bank (2020) "World Development Indicators", 1 July, World Bank. Available at: https://datacatalog.worldbank.org/dataset/world-development-indicators (Accessed 7 August 2020)

[4]World Health Organization (2020) "WHO Coronavirus Disease (COVID-19) Dashboard", World Health Organzation. Available at: https://covid19.who.int/ (Accessed 11 July 2020)

[5]World Meters (2020) "Coronavirus Cases". Available at: https://www.worldometers.info/coronavirus/coronavirus-cases/ (Accessed 7 October 2020)

hypotheses drawn from the three stated research questions - rather, the paper uses descriptive analysis and theoretical arguments to connect analyses and results to its three research questions. Below I highlight the key features of the paper's analysis:

1. Have countries pursued consistently disproportionate policy responses throughout the pandemic? Is there evidence that countries have pursued intentional or nonintentional disproportionate policymaking?

   - Figure 1 shows significant variation in the policy responses over the first year of the COVID-19 pandemic, highlighting France and the US as examples.
   - Note that the authors take correlation between government response and daily cases as indicative of *intentionality* (or non-) and *pro-activity* (or reactivity) of national-level policymaking, though neither theoretical evidence nor statistical analysis is provided to substantiate this causal claim.

2. Are countries with more stringent policy responses more effective in managing the public health crisis than countries with less stringent policy responses?

   - Figure 2 plots the risk-response ratios for each countries with a smoothed line of best fit. Results are inconclusive as evidence for stringency correlating with COVID-19 outcomes, and the authors point to potential sources of error in

     (a) the relative nature of the risk-response ratio classification process (as an absolute scale may capture the true magnitudes of variation and disparity between policy responses), and

     (b) the model's exclusion of contextual factors such as socio-cultural, political and economic dimensions which uniquely impact individual national contexts.

   - The latter point is especially pertinent, as the power dynamics and pressures from social and political contexts are highly influential on state-level decision-making. While this is a provisional study, neglecting national socio-political contexts undermines any findings from a paper oriented towards exploring the dynamics and outcomes of policymaking in crisis contexts.

3. Which types of policies — containment, economic, or health — have proven most effective in managing the public health crisis?

   - The authors highlight evidence that "economic support policies have proven most effective in managing the public health crisis", naming income support and debt/contract relief as associated with a fall in the rate of deaths for up to four weeks, while "broad" interventions for containment, health, and economic supports reflect reductions in the same for a week from their implementation, but not beyond this time frame (Shafi and Mallinson, 2023, p. 107-8).
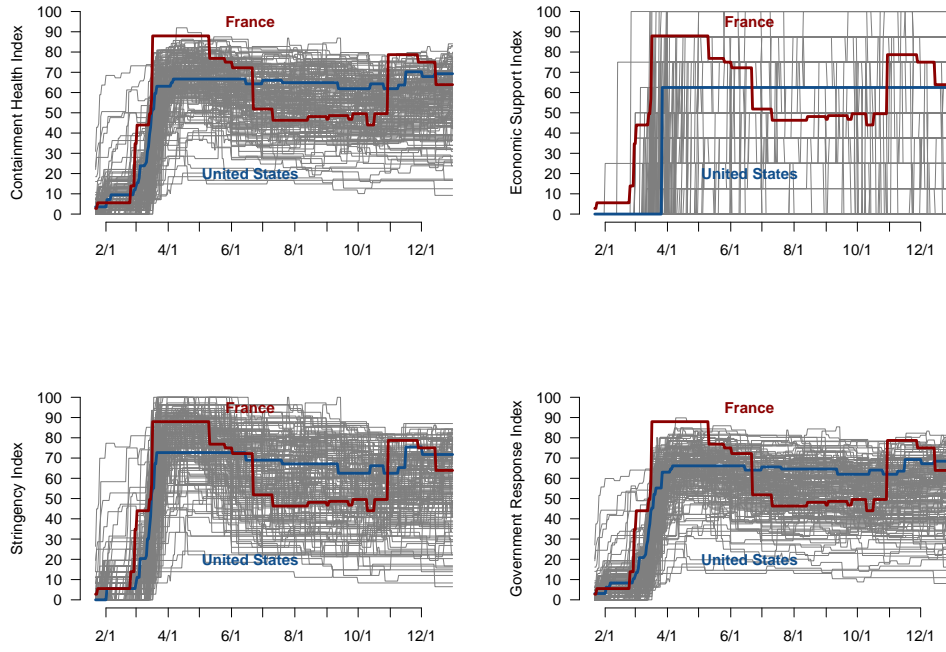
(a) Published Graphic
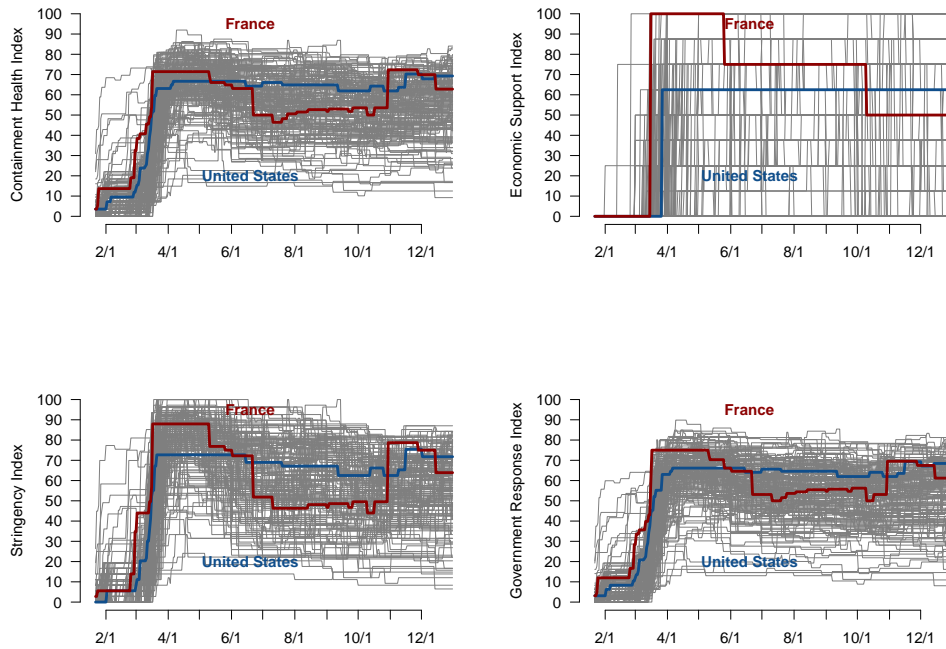
(b) Corrected Graphic

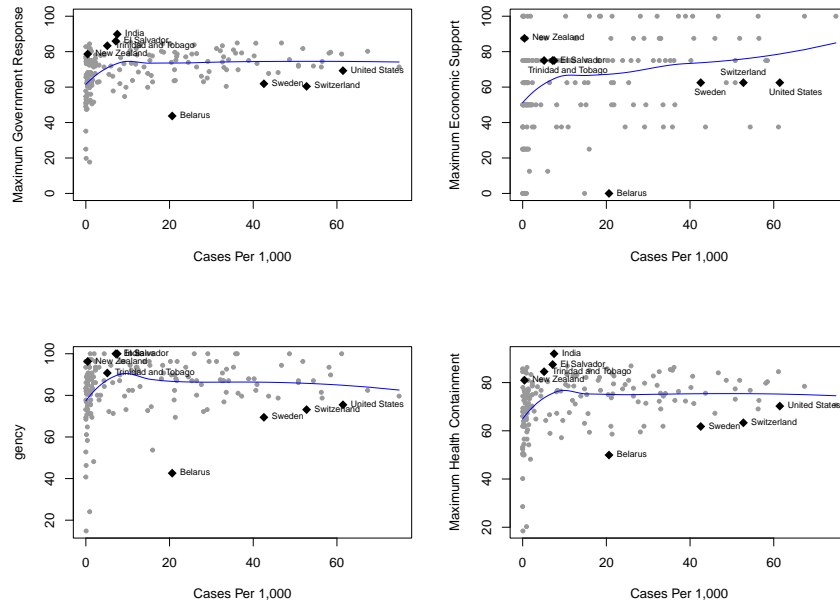Figure 1: COVID-19 Policy Response Indices from January to December 2020

Figure 2: Policy Over- and Under-Response to COVID-19 [Replicated]

- Figure 3 illustrates that the largest log-odds coefficient for `Economic Support Index` on `New Weekly Deaths` lags by 3 weeks, and still holds up to 4 weeks. Results for most other measures are less notable.

- While the aggregated global-level results inform the authors' insights, the use of national-level indicators to identify singular policy responses as generally the 'best' across across the international community is problematic, given the wide variation in social and political contexts, as well the influence of unique national features on policy responses and outcomes, such as the nature of employment and in which sectors, and the variation in actual implementation of particular policy responses. The authors acknowledge that precisely these dimensions are not captured within this analysis, which further undermines any insights which may be drawn from these results.

What is my 'twist'?

A key dimension to the paper's analysis is the nature of Over- and Under-Response to COVID-19, with countries above / to the left of the smoothed regression line classified the former, and those below / to the right as the latter. However, the paper does not directly compare these two groupings of country responses. I argue there are systematic differences between countries which consistently respond 'above' the average rate to those 'below'.

By repeating analyses with subsets of "Over-Responders" and "Under-Responders" (defined

(a) Cases Growth Factor

(b) New Weekly Cases

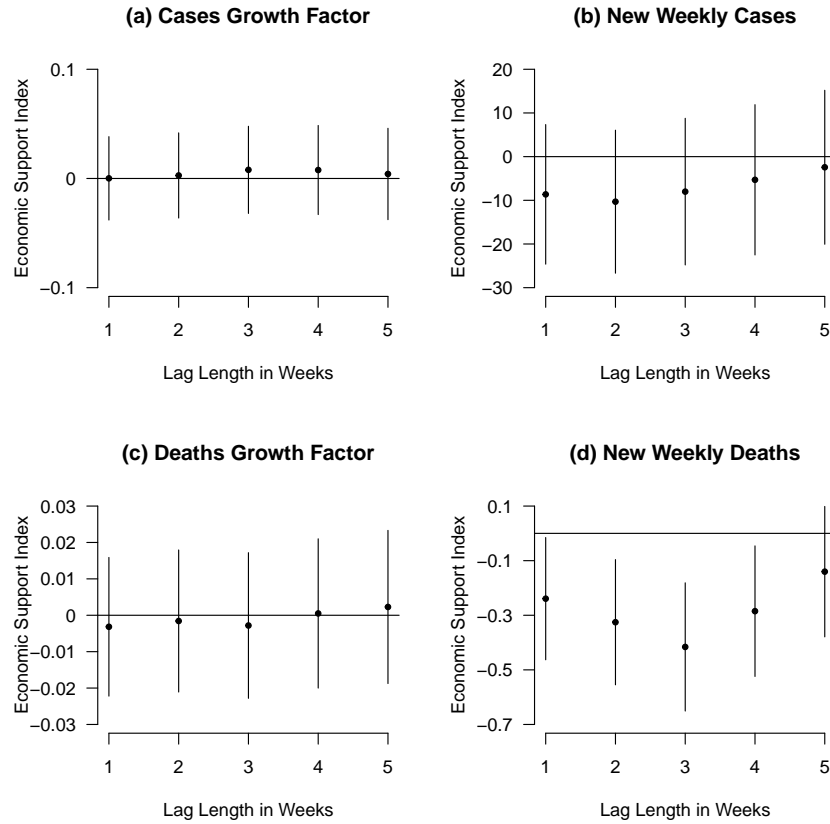**(c) Deaths Growth Factor**

(d) New Weekly Deaths

Figure 3: Economic Support Index Coefficients across Four COVID-19 Outcomes, Lagged 1–5 Weeks [Replicated]

according to relative position to the smoothed regression line, as specified by the authors), I expect to highlight these systematic differences, which later analyses including socio-political dimensions may capture with more precision:

```
1  ## Assigning membership groups of Over− and Under−Response # "Above" and "
      Below"####
2  # Calculate residuals for each point in each plot
3  new_si <− data.frame("country" = max.si$country)
4  new_gri <− data.frame("country" = max.gri$country)
5  new_chi <− data.frame("country" = max.chi$country)
6  new_esi <− data.frame("country" = max.esi$country)
7  new_si$residuals <− resid(loess.ms)
8  new_gri$residuals <− resid(loess.gri)
9  new_chi$residuals <− resid(loess.chi)
10 new_esi$residuals <− resid(loess.esi)
11
12 # Create dummy variables for membership based on residuals
13 new_si$si_membership <− factor(ifelse(new_si$residuals > 0, 1, 0),
14                                 labels = c("Below", "Above"))
```

7

```
15  new_gri$gri_membership <- factor(ifelse(new_gri$residuals > 0, 1, 0),
16                                        labels = c("Below", "Above"))
17  new_chi$chi_membership <- factor(ifelse(new_chi$residuals > 0, 1, 0),
18                                        labels = c("Below", "Above"))
19  new_esi$esi_membership <- factor(ifelse(new_esi$residuals > 0, 1, 0),
20                                        labels = c("Below", "Above"))
```

```
1  # Prepare new data frame for plotting
2  new_data <- data
3  new_data$country <- new_data$CountryName # Rename for merging
4  # Merge membership dummy variables
5  new_data <- merge(new_data, new_si[c("country", "si_membership")], by = "
       country")
6  new_data <- merge(new_data, new_gri[c("country", "gri_membership")], by = "
       country")
7  new_data <- merge(new_data, new_chi[c("country", "chi_membership")], by = "
       country")
8  new_data <- merge(new_data, new_esi[c("country", "esi_membership")], by = "
       country")
```

Next, I reproduce the original paper's first figure (see Figure 4) to illustrate how

    a) countries can be visually grouped by their Over- or Under-responsiveness, though the boundaries between these clusters are not clearly distinguishable, and

    b) the Averages across the two groups diverge from late March/early-April 2020, perhaps by the greatest magnitude across the Economic Support Index (which is especially notable as economic supports are highlighted by the authors as the most significant measures associated with reduced COVID-19 deaths across all countries).

Note, I use the `ggplot2` package due to challenges with representing additional graphic features effectively using Base R's capabilities (see lines `1336 to 1495`, or print the code directly in this paper by modifying the comment below in the given LaTeX file).

Updated Results

In line with the author's analysis, I perform **Hausman Tests** on the updated Fixed-Effect models based on subsets of "Over-Response" and of "Under-Response" classes for each index, with results indicated in Tables 1-3. Notably, the updated fixed-effects GRI models for Over-Response countries are "preferred" for three out of four outcomes, over one of four in the corresponding original models. This is the terminology used by the authors, and so I adopt this for comparing the original and updated approach.

```
1  #Hausmann Test to see if FE improve the model (Above)
2  #H0: Random Effects model consistent and efficient (preferred)
3  #H1: Fixed effects model consistent and preferred
4  phtest(agri1.fe, agri1.re) # FE is preferred, statistically significant
5  phtest(agri2.fe, agri2.re) # RE is preferred, not statistically significant
6  phtest(agri3.fe, agri3.re) # RE is preferred, not statistically significant
7  phtest(agri4.fe, agri4.re) # FE is preferred, statistically significant
```
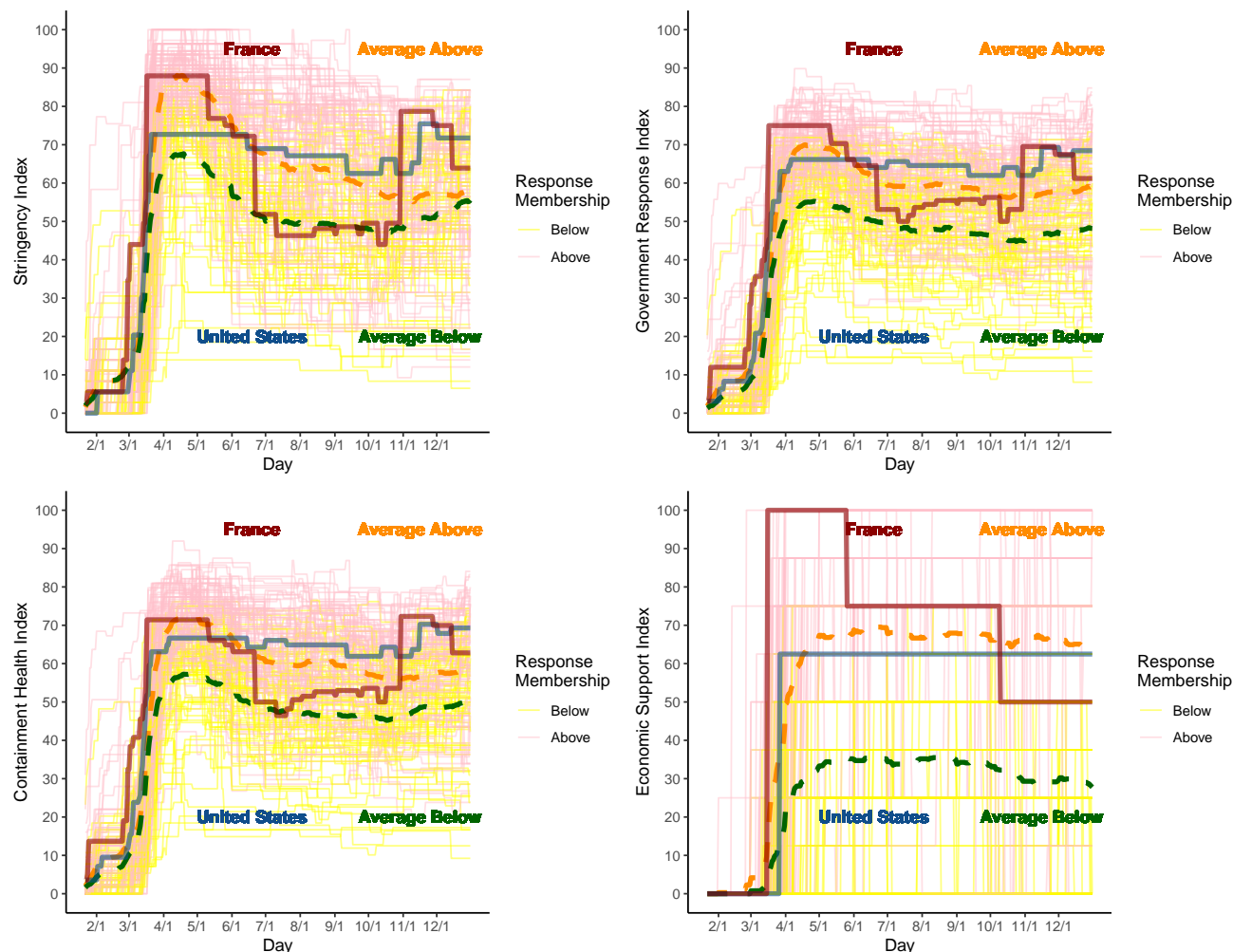
Figure 4: COVID-19 Policy Response Indices from January to December 2020 (Updated)

Meanwhile, the updated fixed-effects GRI models for Under-Responders perform similarly to original models. Overall, these results suggest that, for the GRI index, **grouping countries by degree of policy response gives additional information for estimating the impact of GRI measures on some COVID-19 outcomes**.

```
1 #Hausmann Test to see if FE improve the model (Below)
2 #HO: Random Effects model consistent and efficient (preferred)
3 #H1: Fixed effects model consistent and preferred
4 phtest(bgri1.fe, bgri1.re) # RE is preferred, not statistically significant
5 phtest(bgri2.fe, bgri2.re) # RE is preferred, not statistically significant
6 phtest(bgri3.fe, bgri3.re) # FE is preferred, statistically significant
7 phtest(bgri4.fe, bgri4.re) # FE is preferred, statistically significant
```

I repeat this process for all indices, and similar to the paper's original results, each (updated) FE models are markedly "preferred" for estimating the effects of GR, S, ES, and CH responses on the `New Weekly Deaths` outcome. However, the following discussion will explore how this

Table 1: Hausman Test Results: Government Response Index (Original)

Outcome	Chi-squared	Degrees of Freedom (df)	p-value
Cases Growth Factor	2.1981	4	0.6994
New Weekly Cases	0.69163	4	0.9524
Deaths Growth Factor	4.8157	4	0.3067
New Weekly Deaths	84.324	4	$< 2.2 \times 10^{-16}$

Note: The null hypothesis is that the models are consistent. The alternative hypothesis is that one model is inconsistent.

Table 2: Hausman Test Results: Government Response Index (Over-Response)

Outcome	Chi-squared	Degrees of Freedom (df)	p-value
Cases Growth Factor	559.31	4	$< 2.2 \times 10^{-16}$
New Weekly Cases	0.16971	4	0.9966
Deaths Growth Factor	125.9	4	$< 2.2 \times 10^{-16}$
New Weekly Deaths	490.94	4	$< 2.2 \times 10^{-16}$

Note: The null hypothesis is that the models are consistent. The alternative hypothesis is that one model is inconsistent.

updated approach impacts results and insights.

Firstly, Figure 5 shows the updated results for the Economic Support Index among Over-Responders (see Figure 3 for original results). Interestingly, the updated models indicate that, for Over-Responding countries, Economic Support responses have a statistically significant increasing effect on the log-odds for `Case Growth Factor` up to three weeks, though none before or after. Similar to the original results, there is a decreasing effect on the log-odds `New Weekly Deaths` from the second through to fourth lagged week, though at lower magnitudes to those indicated in the original results. Meanwhile, the results for the Under-Responders subset give no statistically significant findings (these are not printed to save space - see the replication files for details).

The authors also highlight the results from the GRI models in the paper, provided alongside the updated models in Figure 6 below. Note the differences in `Cases Growth Factor` and `Deaths Growth Factor` - the Over-Responders model shows magnitudes-smaller intervals for each, while indicating Government Response has a 2-week lagged effect on cases growth. Additionally, the new models show no statistical significance to the relationship between overall Government Response and `New Weekly Deaths`.

Reflecting on the Stringency Index, while the only statistically significant result from the original model indicates an associated reduction in `New Weekly Deaths` up to the first week, there is no effect among Over-Responders, though over twice the magnitude for Under-

Table 3: Hausman Test Results: Government Response Index (Under-Response)

Outcome	Chi-squared	Degrees of Freedom (df)	p-value
Cases Growth Factor	0.016766	4	1
New Weekly Cases	1.0809	4	0.8973
Deaths Growth Factor	11.689	4	0.01982
New Weekly Deaths	427.15	4	$< 2.2 \times 10^{-16}$

Note: The null hypothesis is that the models are consistent. The alternative hypothesis is that one model is inconsistent.
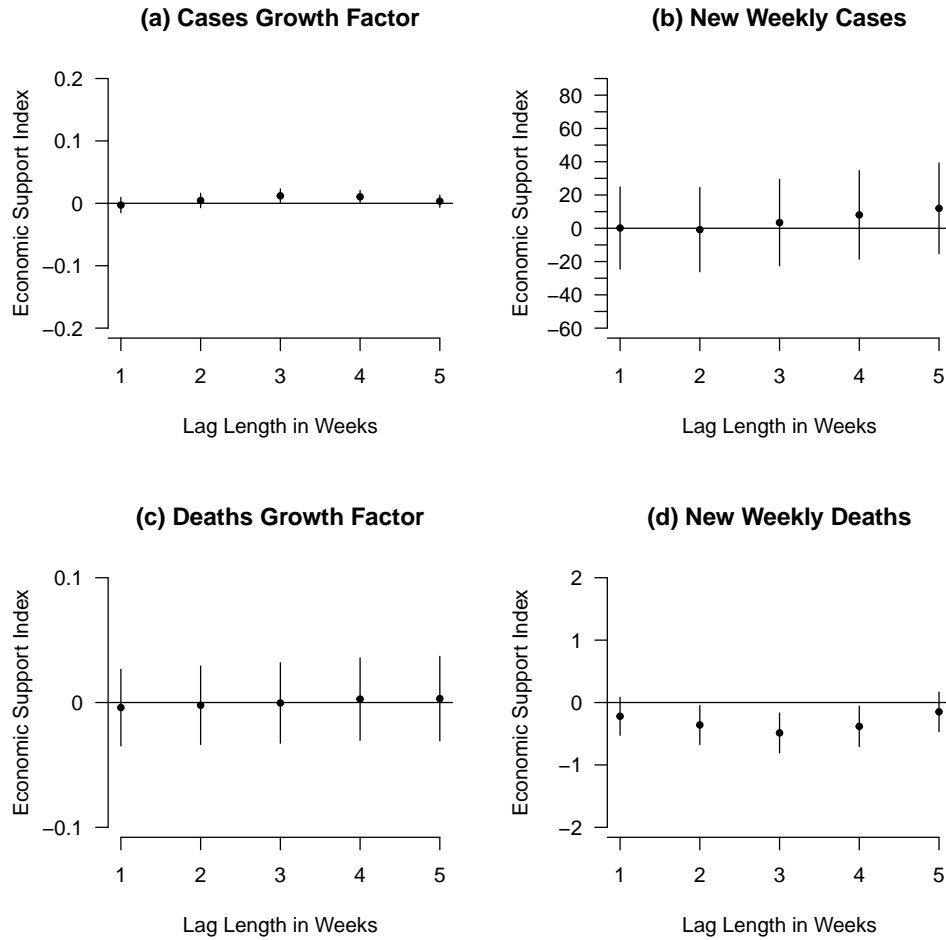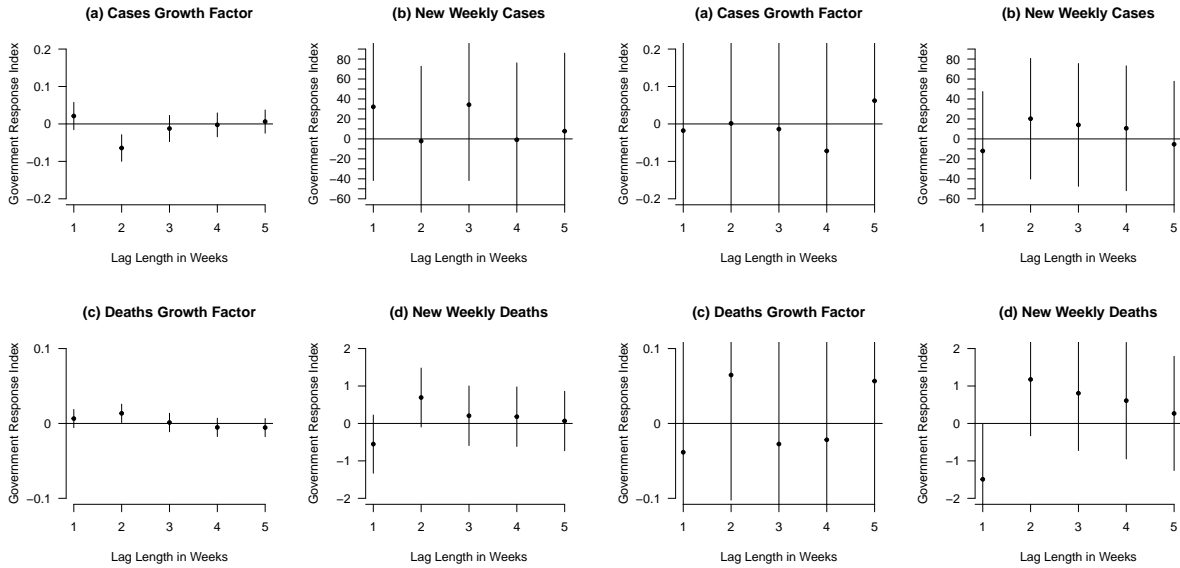

Figure 5: Economic Support Index Coefficients across Four COVID-19 Outcomes, Lagged 1–5 Weeks (Over-Responders)

Responders (Figure 7). However, the following two lagged weeks indicate an associated increase in `New Weekly Deaths`. **This is an important distinction in the dynamics**
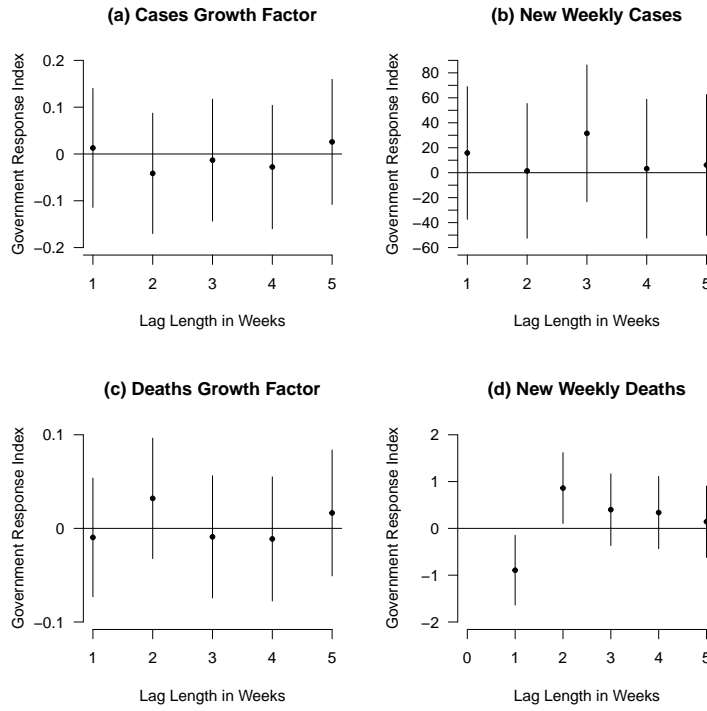
(a) Original

(b) Over-Responders   (c) Under-Responders

Figure 6: Government Response Index Coefficients across Four COVID-19 Outcomes, Lagged 1–5 Weeks

between responses and outcomes for Over- and Under-Responders which is not captured in the original analysis. Meanwhile, **over-responses in stringency appear to have an associated decrease by or up to the second week**, though not before or after, which is also not captured in the original analysis.

Lastly, the updated `Containment Health Index` models (Figure 8) show similar results as for the `Stringency Index`, the magnitude of the decrease in `New Weekly Deaths` up to the first week is greater from under-responses and is absent for over-responses, while over-responses have an indicated lagged effect up to and by the second week. The first point refines the original paper's findings, and the second uncovers a dynamic which the original paper did not find.

Findings and Conclusions

The analyses indicate that, for specific outcomes, the dynamics between particular policy responses and COVID-19 outcomes are conditional upon whether a country Over- or Under-Responded. Further, these results provide evidence supporting the effectiveness of some policy responses which was not found by the original analysis. This therefore indicates further work should be done to parse conditional dynamic effects of policy responses and how these relate to dimensions both included and excluded by the analyses used in the original paper and this replication project - from improving the classification method used in this paper, to including socio-political dimensions and other more localised contextual information.

My twist has added value and insight to the original paper, however it takes for granted problematic assumptions about the methodology used - namely, the validity of fixed-effects models employed in the original research. Their use in this case remains questionable, even based on the Hausman Tests provided within the replication code provided by the authors. Therefore, further investigation into the appropriate use of linear regression models and in comparative policy analysis work is necessary, as well the issue of inconsistency in the replication and/or interpretation of hypothesis test results.
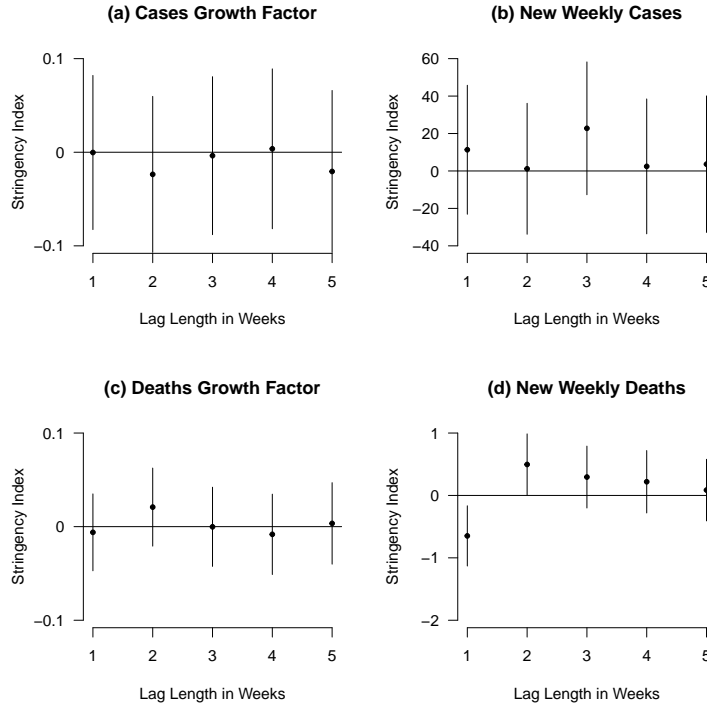
(a) Original
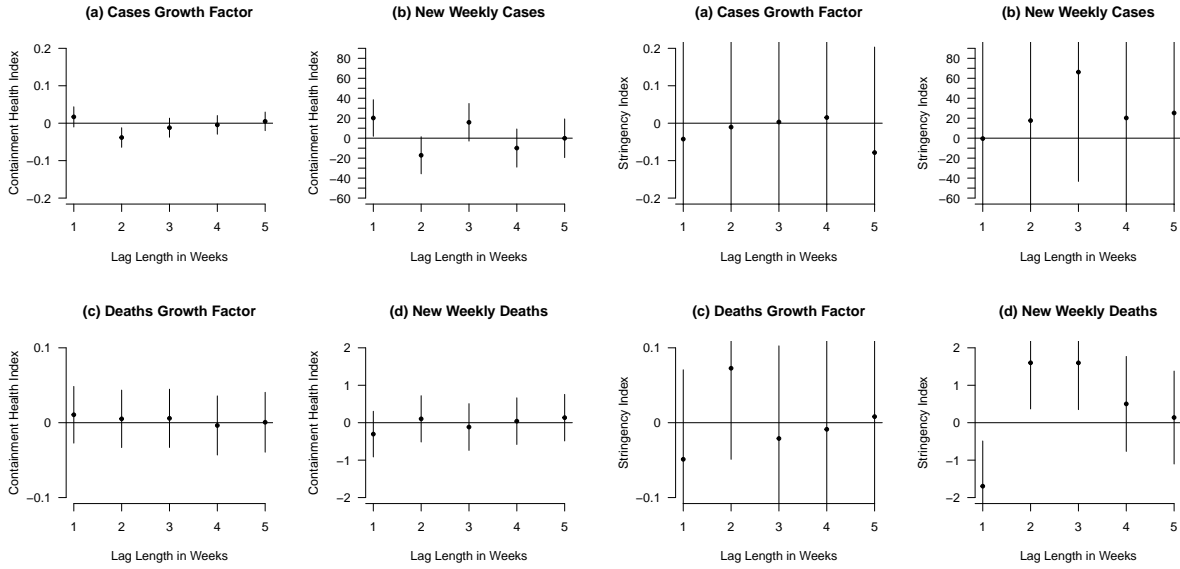
(b) Over-Responders

(c) Under-Responders

Figure 7: Stringency Index Coefficients across Four COVID-19 Outcomes, Lagged 1–5 Weeks
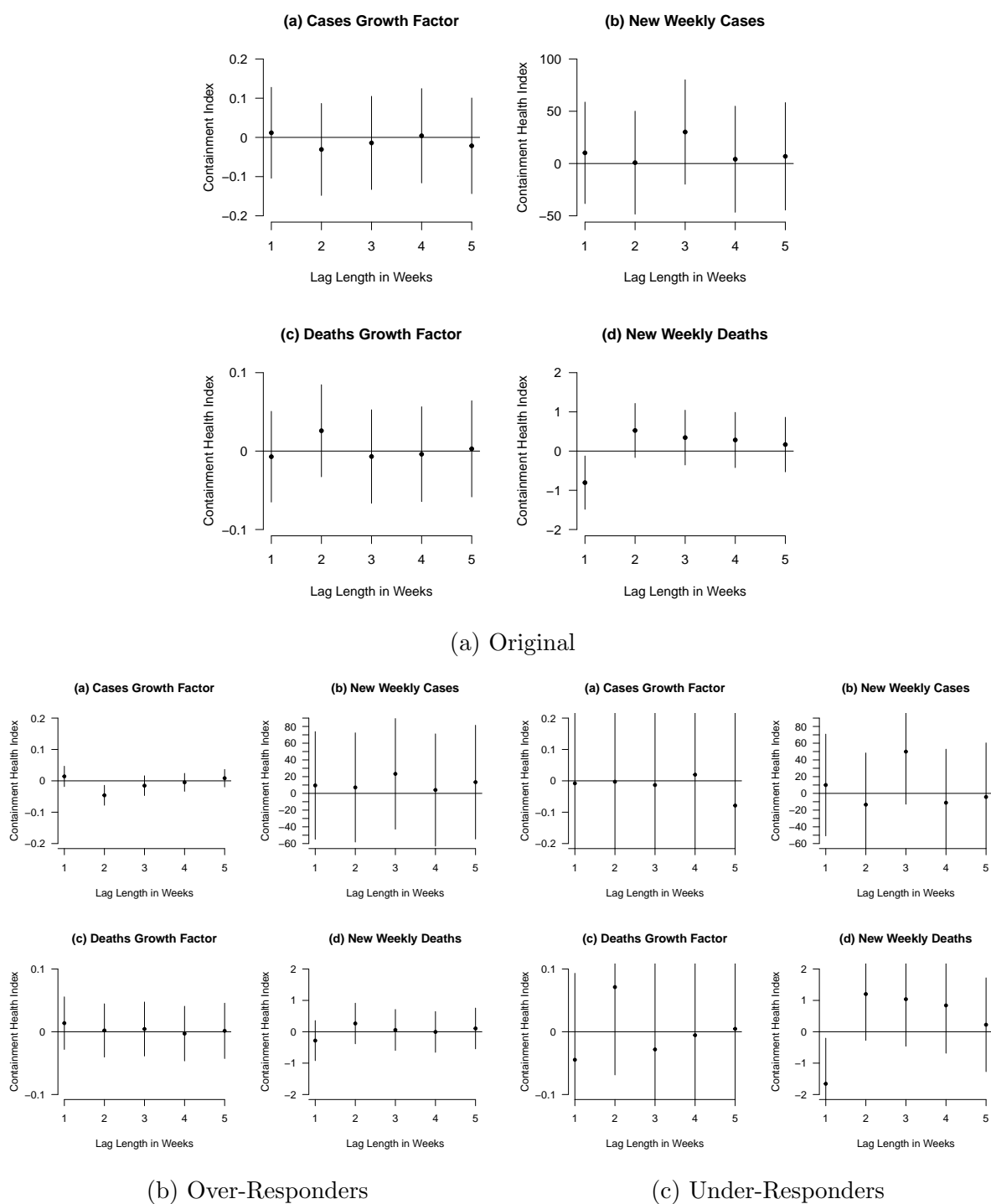
(a) Original

(b) Over-Responders

(c) Under-Responders

Figure 8: Containment Health Index Coefficients across Four COVID-19 Outcomes, Lagged 1–5 Weeks