Supplementary Information for Cross-National Measures of the Intensity of COVID-19 Public Health Policies

February 16, 2022

Contents

1 Methods **2**

1.1 Data Aggregation 2

1.2 Ideal Point Model 3

1.3 Simulation 9

1.4 Incorporating Measurement Error in Regression Analysis 13

1.5 Evaluation of Predictive Performance 14

2 Application to Estimating COVID-19 Suppression **17**

3 Additional Figures **22**

4 MCMC Convergence **25**

5 CoronaNet Codebook **26**

6 OxCGRT Codebook **40**

7 Composition of Indices **41**

8 References **43**

1 Methods

1.1 Data Aggregation

Employing statistical measurement models from different data sources provides at least two unique benefits. First, it allows us to create indices which contains more information than either data source could provide on its own. The value of combining the CoronaNet and OxCGRT datasets is particularly high given that they rely on disparate taxonomies which otherwise could not be easily combined. Second, and most crucially, by employing measurement methods, our indices are able to move beyond de jure policy indicators to a measure of the overarching goal that is guiding the selection of individual policies. That is, by designing our model to measure a single-dimensional latent trait of policy intensity per policy domain, we can model their causes and consequences at a more conceptual level and we gain uncertainty intervals that describe realistic levels of policy coherence. As we show in this section, doing so is necessary to avoid faulty inferences when multiple policy indicators are used to describe the same policy domain, such as masks or lockdowns.

In creating the indices, we combine indicators from the CoronaNet taxonomy, which itself relies on 20 policy categories, with OxCGRT indicators that cover the same aspect of a policy category. To summarize the composition of our indices, see Table 1 in the supplementary information. It shows how each of the six indices is composed through the policy classification in the CoronaNet and OxCGRT taxonomy by top-level and sub-level variables. We include 154 indicators from the CoronaNet dataset and 10 from OxCGRT. We use more indicators from CoronaNet because of differences in how these two projects code public health policies, which we further explore in the supplemental information. In general, the CoronaNet project takes the approach of separately coding all aspects of COVID-19 related policies, while the OxCGRT produces ordinal scores for each policy type.

One complication is that the CoronaNet dataset contains provincial and even municipal information about policies for all countries in the dataset, though coverage varies. As such, we needed a way to include this information to produce national-level estimates. We employed a simple algorithm in which we assigned a value of +1 for each policy in force at the national level and 0 otherwise. For over-lapping policy coverage at lower levels of government we then add to this baseline score. For each province and city with the same policy in force, we add the proportion of the country's population in the province or city to this baseline score. The resulting score could reach a maximum of 3 if a binary-coded policy was implemented at the national level, among all policies and among all cities in a country.

Of course, we only rarely observe this maximum level of policy enforcement in our indices. For example, if we look at complete business shutdown policies, which were quite prevalent during the pandemic, only 17% of our aggregated observations contain any additional provincial or municipal-level information. Only 4% of records show complete national and provincial level coverage. The main benefit of the aggregation method is to allow highly federal countries, such as the United States, Russia and Switzerland, to have comparable scores to those countries with more unitary enforcement of COVID-19 policies. For most countries in our

dataset, the aggregation method has little effect as COVID-19 policies were primarily decided at the national level.

The additional advantage of our aggregation approach is that it can capture multiple levels of overlapping enforcement, which may reduce loopholes. For example, if a national-level stay at home order is difficult to enforce in rural areas, provincial-level orders may be more effective if national institutions lack sufficient penetration into rural society. As such, the idea that a country's score should be higher if it implements similar policies at multiple levels of government appears to have significant face validity even if we do not always have the correct country-level weights. Again, as we are using a measurement model, we do not assume these scores are without error, nor do we assume that the errors are Normally-distributed. For these reasons, even if the aggregation rule introduces some additional uncertainty by using constant weights for combining policies, this random noise should propagate through to our final measures.

In total we have 189 countries in both datasets with coverage from January 1st, 2020 to April 30th, 2021, resulting in 102,711 country-day observations with all of the 164 pooled indicators. As policy coding remains an ongoing task, we intend to release further periodic updates to our indices as more data is available. A complete codebook of all of the indicators is available in the supplemental information.

1.2 Ideal Point Model

Given the aggregated data, we next want to find a way to infer the combined goal of these policies from the data. Doing so is not straightforward, however, in spite of the granularity of our data. We know that the policies are not conditionally independent, and to treat them as such would warp any observed associations between these policy indicators and outcomes we care about. Instead, we need to understand the causal process which generates the data, and then focus on theoretically defined quantities that are the most relevant to the fight against COVID-19.

To address this challenge, we turn to one of the most widely-used formal frameworks in political science, the ideal point model. The ideal point approach can be understood as a simple formal model in which an actor i must decide between two policy alternatives j and j' (Clinton, Jackman, and Rivers 2004). If we assume that there is a one-dimensional utility score $U(\cdot)$ assigned by actor i to j and j', then actor i will always choose j over j' if $U(j) > U(j')$. In our context, this implies that countries will choose those policies j which have the smallest distance to their ideal point score, denoted x_i. Following Clinton, Jackman, and Rivers (2004), the policymaker i's utility function $U_i(\cdot)$ for a given policy j can be defined as:

$$U_i = \|x_i - N_j\|^2 - \|x_i - Y_j\|^2 \tag{1}$$

where N_j and Y_j represent the No and Yes positions of a given policy choice j in the latent space. For our purposes, the latent space represents a distinct policy outcome that influences demand-side (preventing behaviors that lead to infections) or supply-side (equipping health systems to better treat and detect COVID-

19) factors. We define the six policy goals that we consider concretely in the main text. For example, the mask goal has the aim of increasing mask wearing among the population to prevent infections, and so is a demand-side policy goal. Business restrictions aim to prevent mixing among non-relatives by influencing mobility patterns in workplaces, and are another demand-side policy goal. The policymaker's ideal point, which we term the policy intensity score, reflects their willingness to devote scarce resources, whether they be political, administrative or financial, to accomplish the policy goal. Given the policymakers' level of commitment of resources to accomplish the policy goal, represented by their ideal point score x_i, the policymaker selects those policies j which maximize their utility relative to this level of commitment.

Importantly, the varying No and Yes positions of each policy mean that we do not assume, a priori, that each policy would contribute equally to the policy goal. In fact, by estimating model parameters, we can learn which policies appear to be those that are most cost-effective (from a policymaker perspective) at achieving a given policy goal. Each policymaker's cost function is necessarily heterogeneous and may comprise varying weights attached to political, financial and moral objectives. We can only estimate the policymakers' allocation x_i to a given policy goal, though we can still learn about plausible differences in cost functions by examining conditional associations between x_i and relevant institutional characteristics, such as democratic and authoritarian political institutions. For example, if democratic institutions impose greater costs on policymakers for closing schools, then we might expect that the x_i for schools would be necessarily lower for policymakers who face electoral accountability versus policymakers who do not have to worry as much about the median voter.

For these reasons, estimating both x_i and the yes/no positions of policies in a latent space can help us learn valuable information about the constraints that policymakers face and the ways that they combine policies to achieve a given policy goal. While the ideal point model is generic, it is also a powerful way of framing the policymakers' choice problem and helps us understand the connections between otherwise disparate policies. As we show in the next section, when we simulate data from this model it shows why conditional associations between policies and outcomes can be obscured by the fact that individual policies are only one part of larger policy goals. In fact, regression models using correlated policy indicators are almost certainly biased without taking into account the fact that these policies have similar aims. The utility of the ideal point approach is that we can take indicators we know to be correlated, and pull out of these indicators an estimate of the over-arching policy goal and the relative commitment of policymakers to a given objective, permitting us to estimate meaningful quantities from the disparate policies.

We note that this theoretical approach differs from exploratory methods such as principal component analysis and factor analysis that are commonly used for dimension reduction. It is true that COVID-19 policy indicators are correlated–as we show in our empirical example of Flaxman et al., sometimes highly correlated–but the issue is not simply to produce a less correlated measure. Rather, we have strong prior information about why these indicators are correlated, and we want to capture a higher-level measure that can help us understand how these policies are related to each other and what their main aim is. By learning

how these policies relate to each other, we can gain a much more accurate sense of how these policies are combined to achieve a given policy goal. Purely exploratory methods like PCA and factor analysis would accomplish the dimension reduction objective but would not necessarily map on to our formal model. As we show in the next section, PCA estimates can be correlated with our model's estimates but have significant bias arising from mis-specification.

Having specified the formal model, we now consider inference. The only observed data we have are the presence or absence of different types of policies Y_{ij} that policymakers implement. We cannot directly observe the level of commitment x_i that underlies the selection of a particular set of policies Y_{ij}. As such, we need to be able to gain an estimate of the unobserved x_i conditional on the observed data Y_{ij}. We can do so, again following Clinton, Jackman, and Rivers (2004), by using their equivalence between the ideal point model and the 2-Pl item-response theory specification with ability parameters $\hat{x_i}$ for the unobserved ideal points x_i, discrimination parameters δ_j, difficulty parameters β_j, and a function $g(\cdot)$ that represents an appropriate probability density or mass function for the policy indicator Y_{ij}:

$$\delta_j = \frac{2(Y_j - N_j)}{\sigma_j} \tag{2}$$

$$\beta_j = \frac{Y_j^{`\prime} Y_j - N_j^{'} N_j}{\sigma_j} \tag{3}$$

$$Pr(Y_{ij}) = g(\delta_j^{'} \hat{x_i} - \beta_j) \tag{4}$$

This equivalency is arguably one of the most important results in empirical political science. We can use the item-response theory estimation technique, which has existed for decades in psychometrics, to obtain valid measures of the unobserved policy-maker intensity scores x_i as our parameter $\hat{x_i}$. The primary loss of information relative to the full formal model is that we do not know the actual yes and no positions of each policies. Rather, we have the discrimination parameters δ_j which are equal to the difference between the yes and no positions for a given policy. As that difference increases, a policy is likely to show increasing discrimination, meaning it is more responsive to shifts in the policymaker intensity score $\hat{x_i}$. For this reason, we can interpret the discrimination scores as those relatively high cost policies as only policymakers that commit a high level of resources x_i could afford them.

Given the definition in 2, we can define our IRT model as a directed a-cyclic graph (DAG) with observed and unobserved parameters, which we show in Figure 1. In this figure we have four unobserved variables: the level of commitment (i.e., the policy intensity score) x_i, the unobserved policymaker-specific costs C_i and B_i that undergird the policymakers' choice, and finally the bureaucratic process B_j that turns the policy intensity score x_i into an implemented policy. We cannot observe C_i and B_i as we do not have direct measures of each policymakers' political constraints, though as noted previously, we can infer some associations with these unobservables post-estimation by looking at how factors that plausibly affect costs and benefits, such

as the type of political institutions in the policymaker's country, affect policy intensity scores x_i indirectly via C_i and B_i.

We could in principle observe the bureaucratic implementation process $B_j(\cdot)$, but doing so would require extensive field research in government agencies. Instead, we observe media reports M_{ij} for each policy implemented by the bureaucracy. These media reports are subsequently coded by our research assistants as policy indicators Y_{ij}. As can be seen, there are multiple ways that measurement error can creep into this process as media reporting is an imperfect proxy, research assistants can miscode media reports, and finally policymaker intentions are imperfectly communicated through the bureaucratic policy implementation process. Our model cannot distinguish between these sources of measurement noise, although it can separate the variation in our measured data into components that would appear to be generated by a one-dimensional construct (intensity scores) and those that are due to other sources (various forms of measurement error).

The DAG in Figure 1 also makes clear how the discrimination parameters δ_j relate to the policymakers' intensity score x_i. An increasing value for δ_j will result in a much higher probability of the policy indicator Y_{ij} being imposed conditional on the level of x_i. In other words, the discrimination parameters and the policy intensity scores interact. The discrimination parameters could be also be though of as weights that represent the optimal policy mix to achieve the policy goal, at least as they are perceived by policymakers in the aggregate. It is theoretically possible for the discrimination parameters to be negative–for the Yes and No positions on a given policy to be reversed–but our definition of the latent variable suggests that this should not happen because we are measuring a level of commitment to a given policy area. Reversing the yes and no positions would imply that some policymakers implement more policies as their level of investment of resources x_i decreases, which is of course nonsensical.

However, the ideal point model admits this possibility because it can be used to study latent variables that exhibit polarizing dynamics, such as conservative and liberal bills in the U.S. Congress. In that case, a high value of the ideal point x_i could signal either Yes or No votes on a bill conditional on whether the legislator is conservative or liberal. Because our measure x_i represents instead investment in a given policy area, we can directly test this assumption by estimating whether the discrimination parameters are all positive. If they are, then it provides some helpful evidence that our interpretation of the latent variable x_i as a commitment of resources is credible.

To allow us to identify the model, we fix some of the discrimination parameters to discrete values and also assign priors to all parameters. These priors for the policy indicator parameters are:

$$\delta_j \sim \text{Normal}(0, 3) \tag{5}$$

$$\beta_j \sim \text{Normal}(0, 3) \tag{6}$$

We use a distinct prior for each policymaker, or in the case of our data, country $\hat{x_{it}}$, because we want

Figure 1: Directed Acyclic Graph Representation of Model

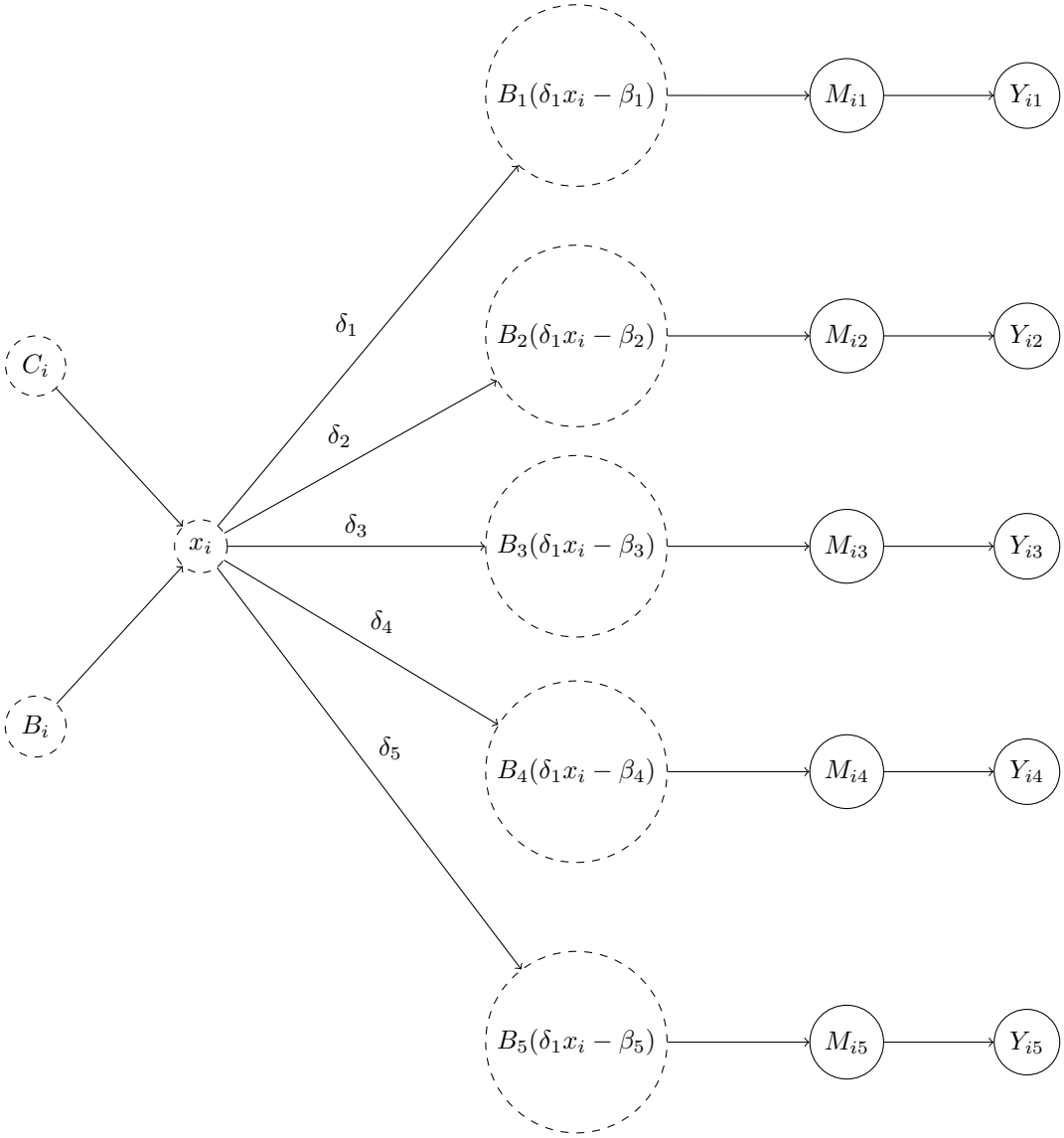


Figure shows the process by which the observed policy indicators Y_{ij} are generated. Given the policymaker i's unobserved costs C_i and benefits B_i, the policymaker chooses an optimal level of investment x_i to the given policy goal. The δ_i determine the allocation of these resources across specific policies to achieve the desired outcome. Bureaucracies $B_j(\cdot)$ take in both the intensity level and policymaker investment and then implement a policy. This policy implementation results in a media report M_i about the policy, which is then subsequently coded in CoronaNet data as an policy indicator $Y_1 i$.

to allow the intensity scores to vary over time. We use a straightforward random walk or Brownian motion model by adding the following prior for $\hat{x_{it}}$ with a static prior for the first time point $t = 1$:

$$\hat{x_{i(t=1)}} \sim \text{Normal}(0, 1) \tag{7}$$

$$\hat{x_{it}} \sim \text{Normal}(\hat{x_{i(t-1)}}, \sigma_i) \tag{8}$$

$$\sigma_i \sim \text{Exponential}(.1) \tag{9}$$

This prior allows us to include time autocorrelation into the model because we know that the prior time period's level of x_i will influence what the policymaker decides in the current day. That is, policymakers decide whether to increase, reduce or maintain the current level of policy; policies are not invented de novo on each day. The random walk process is one of the simpler time processes available and imposes few constraints on the values of $\hat{x_{it}}$ for a given day. The prior will induce some smoothing governed by σ_i, but because we estimate one $\hat{x_{it}}$ for each day, the smoothing in the prior has only minimal influence on the final scores. We include autocorrelation primarily to increase realism in the model; i.e., to acknowledge the fact that policies are not re-imposed each day but rather usually are a part of long-term policy initiatives that often have significant stickiness.

The most important detail about the priors is that they are centered around zero. A priori, without data, the estimated intensity scores $\hat{x_{it}}$ will follow a random walk around the policy space loosely defined by the Normal priors. We know that the limit of Brownian motion results in the Normal distribution (i.e., a Galton board), but the over-time average for all countries will remain mean 0. As such, without data the model will produce uncertainty intervals that are centered at zero, which signifies that we are using conservative prior distributions that create a slight bias in the model estimates towards finding no information. While this slight bias will vanish once we have sufficient data, it is still important to note that the model does not on its own produce informative estimates of $\hat{x_{it}}$ absent informative data, i.e., it does not naturally over-fit.

Given the priors and the likelihood $Pr(Y_{ijt}|\hat{x_{it}}, \delta_j, \beta_j, \sigma_i) = g(\delta_j^{'}\hat{x_{it}} - \beta_j)$, we can then estimate the joint posterior distribution of the parameters using Bayes' rule:

$$Pr(\hat{x_{it}}, \delta_j, \beta_j, \sigma_i|Y_{ijt}) \propto Pr(\hat{x_{it}})Pr(\delta_j)Pr(\beta_j)Pr(\sigma_i)Pr(Y_{ijt}|\hat{x_{it}}, \delta_j, \beta_j, \sigma_i) \tag{10}$$

We omit the normalizing constant in the denominator as is common in applied Bayesian problems (it is not directly used in estimation).

Estimating this model is a difficult challenge given the scope of the data and the diversity in variable types, including binary, continuous, ordinal policy indicators, and the need to incorporate over-time variation. To accomplish this, we employ a Bayesian version of item-response models previously used by Cheng et al. (2020) that are estimated with the Hamiltonian Markov Chain Monte Carlo (MCMC) sampler Stan (Carpenter et al. 2017). To handle the very large nature of the data, we use within-chain parllelization that distributes

the posterior gradient calculations across 128 cores. Even so, estimation of one policy intensity index takes approximately two days even with a dedicated node.

To fit the models, we run four independent MCMC chains with Stan for 600 warmup iterations and 350 sampling iterations. To test convergence, we report split-$\hat{R}$ diagnostic plots in the supplementary information for each index model, which show clear convergence of the chains given these sampling parameters.

1.3 Simulation

As we mentioned, we can then interpret the ideal point score as a measure of the relative level of commitment or intensity in the policy domain, while the position of the policy item corresponds to its relative cost and presumed efficacy to the policy actor. Only those actors who commit a lot of resources are able to implement high-cost policies. To establish the relationship between the ideal point model and the item-response theory 2-Pl formulation, we simulate data from the ideal point model and then estimate the intensity scores x_{it} as the item-response theory parameters $\hat{x_{it}}$.

For our simulation, we use a set of countries $N = 20$, a set of policy indicators $J = 10$, a set of time points $T = 60$, and an average over-time variation in country-level policy intensity scores of $\sigma_i = 0.25$. For each of 2000 simulations, we then draw the relevant parameters using the following distributions following the specification in 1:

$$x_{i(t=1)} \sim \text{Normal}(0, 1) \tag{11}$$

$$x_{it} \sim \text{Normal}(x_{i(t-1)}, 0.25) \tag{12}$$

$$N_j \sim \text{Exponential}(1) \tag{13}$$

$$Y_j \sim N_j + \text{Exponential}(2) \tag{14}$$

We simulate the yes positions Y_j and no positions N_j such that $Y_j > N_j \forall j \in J$. This reflects our interpretation of the latent variable as representing relative levels of policy commitment, not a potentially polarizing latent variable like ideology. It should be noted that this prior differs from that used for discrimination parameters in the IRT 2-Pl specification in 6, which allows for $Y_j < N_j$ with equal probability. As such, the estimation uses the "wrong" prior so that we can determine if the model can pick up on the correct ordering of Y_j and N_j given the observed data, which is a test of whether the model's posterior estimates of the sign of δ_j can be interpreted as showing the validity of our estimation approach vis-a-vis the model.

Given these parameters, we can then calculate the latent utility of the policymaker's choice in country i:

$$U_{ijt} = \|x_{it} - N_j\|^2 - \|x_{it} - Y_j\|^2 \tag{15}$$

Given U_{ijt}, we can then draw our observed policy indicators Y_{ijt}. Half of these policy indicators are

continuous variables drawn from the Normal distribution, and half are binary variables drawn from the Bernoulli distribution as follows:

$$Y_{i(j<6)} \sim \text{Normal}(U_{ijt}, 0.5) \tag{16}$$

$$M_{i(j>5)} \sim \text{Normal}(0, 0.5) \tag{17}$$

$$Y_{i(j>5)} \sim \text{Bernoulli}(q(U_{ijt} + M_{ij})) \tag{18}$$

where $q(\cdot)$ represents the inverse logit function to map the continuous latent utility to the $(0, 1)$ interval. We add Normally-distributed noise to the Bernoulli distribution to reflect additional measurement noise such as coding error; this error propagates to the continuous variables through the standard deviation parameter of the Normal distribution of 0.5. This additional error is included to make the resulting data even more error-prone than the model might expect to ensure it is realistic.

We pass the observed data Y_{ijt} to our Bayesian IRT estimation and fix one policy indicator discrimination δ_j to the true value, i.e. the difference between the Yes vote and the No vote. We select the discrimination parameter with the highest difference between these two values.

In addition to generating the policy indicators Y_{ij}, we can also simulate the effect of a given policy investment x_i on COVID-19 infections, which we denote I_{it}, via the following relationship:

$$I_{it} \sim \text{Binomial}(q(-10 - 2x_{it}), 10000000) \tag{19}$$

where $q(\cdot)$ is again the inverse logit function to handle scaling issues.

This simulation generates a count of COVID-19 infections I_{it} that is negatively related to the level of policy investment in a country with a population of 10 million. This simulation reflects a situation in which devoting resources x_i to a specific policy goal such as mask-wearing has an indirect effect on COVID-19 infections through policy-induced mask-wearing. We can then consider our ability to capture the true policy effect, -2, as a metric of the success of our measurement model.

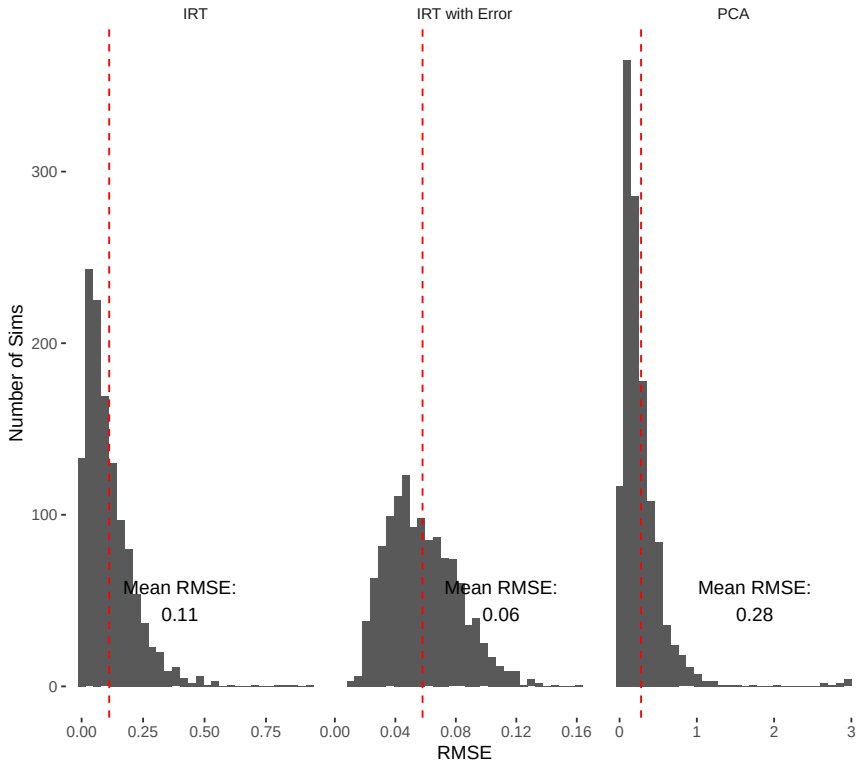
As a comparison, we also employ a dynamic PCA specification (Hörmann, Kidziński, and Hallin 2015) to obtain a different estimate of x_i and an estimate of the policy goal effect. To ensure that arbitrary scaling differences do not affect results, we standardize all estimates of x_i and the true value of x_i before estimating the generalized linear model of simulated COVID-19 case counts. We estimate two models with the policy intensity score measures $\hat{x_{it}}$, one with posterior means as a covariate and the other with Normally-distributed measurement error by passing the posterior standard deviation for each $\hat{x_{it}}$. Finally, we also estimate a GLM with the policy indicators and the simulated COVID-19 case counts. Because Lasso has been used in the literature as a way of addressing policy indicator correlation (Haug et al. 2020), we also estimate a Lasso regression with the policy indicators, selecting the hyper-parameter α with 10-fold CV.

One limitation in the simulation is that automatic identification is not completely possible. Approx-

imately 10% of the draws show sign identification issues in which the discrimination parameters and the policy intensity scores have the wrong sign. For this reason, we exclude any draws with policy regression coefficients that are larger than +1 as these indicate a lack of sign convergence either with our IRT model or with the PCA model. We also exclude draws which our variance-inflation factor test shows have nearly collinear policy indicators, causing the test to fail.

In Figure 2 we report RMSE for the estimated regression coefficients from the COVID-19 infection counts relative to the true coefficient of -2. As can be seen, the PCA model has approximately three times as high RMSE than the IRT models, which indicates that although it finds a latent dimension that is correlated with the ideal point model, it is not in fact the same thing. We note that this RMSE difference remains even after standardizing the latent variables to remove scaling artifacts. Furthermore, the PCA estimation shows greater variance in RMSE, implying that at times the estimation technique ends up with regression coefficients that are quite far from the true value.

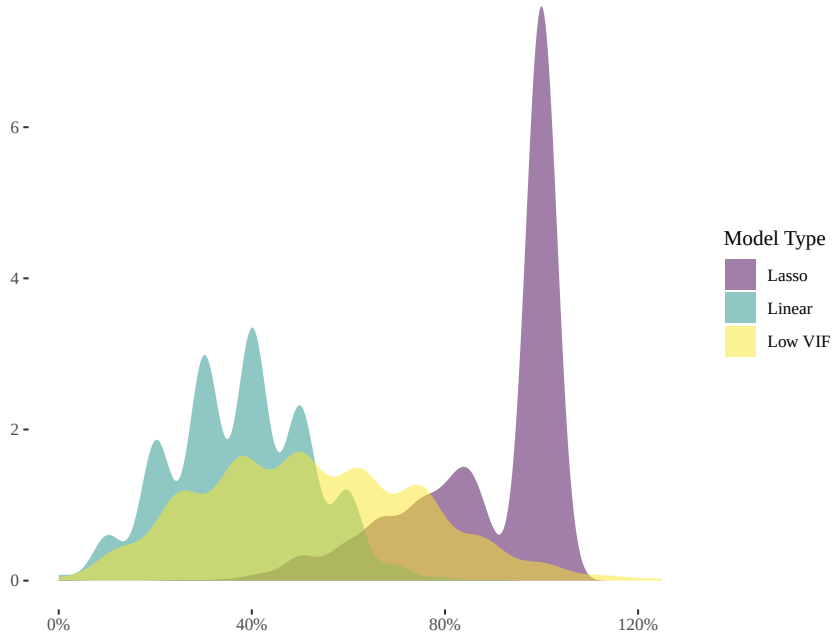
Figure 2: Results of Simulation of Latent Variable Effect on COVID-19 Rates



Plot shows 383 Monte Carlo simulations from an ideal point model (squared distance to policy positions). RMSE is calculated relative to the coefficient of the true value of the latent variable in a regression model with a binomial−distributed outcome. The true value of the regression coefficient is −2. Estimates greater than +1 have been removed as they signify a failure of model identification.

Another very interesting conclusion from Figure 2 is that the IRT model incorporating measurement error in the post-estimation regression has approximately 50% lower RMSE than the model that only uses the posterior mean. This improvement in relation to the true value is due to our robust incorporation of measurement error via Bayesian regression modeling as we discuss in the next section. In essence, the

Figure 3: Ratio of Policy Indicator Coefficients with Wrong Signs



Plot shows the positive proportion of estimated policy indicator coefficients in both a linear model and a Lasso model on simulated COVID−19 case counts from 1434 simulation draws. A distribution of linear model coefficients with the top two highest variance inflation factor scores excluded is also shown.

posterior location of the latent variable is lower variance than the IRT estimates because the model finds those values which make the regression outcome most likely. Because that relationship contains information on the latent variable, the measurement error model is able to find a more accurate representation of x_{it} than just the mean of $\hat{x_{it}}$. As such, we believe this result shows the importance and utility of incorporating measurement error in downstream analyses whenever possible.

Figure 3 shows the number of positive coefficients from our policy indicator regressions, both the standard linear model and the Lasso regression. As can be seen, both types of regressions show substantial numbers of positive coefficients, which would be the opposite sign of the relationship between the latent variable and COVID-19 infections. As an additional check, we show in the figure the distribution for all policy indicators from the linear model after we removed the coefficients with the two highest variance inflation factor (VIF) scores. This distribution though is even more highly skewed towards positive coefficients.

The results in Figure 3 indicate why it is not feasible to simply substitute the estimate of policy intensity scores x_i for individual policy indicators. It is trivial in this case to find coefficients that have the wrong sign, implying that these policies caused COVID-19 cases when in fact their combined effect is negative. Again, because the latent variable x_i can only positively predict policy indicators Y_{ij}, it is an inference error to conclude that these policies are positively associated with COVID-19 cases when they should be negatively associated.

When we probed this result, we noticed that entering each policy indicator into a regression singly

resulted in relationships that are always negative vis-a-vis simulated COVID-19 infections. It is clear that the positive coefficients come about entirely due to assumptions of linear regression that there is no endogenous relationship (i.e., mutually causal) between the regressors. As we discuss in the main text, we can even identify these kinds of non-intuitive results in published work that employs multiple policy indicators in the same model (Haug et al. 2020). Furthermore, we note that Lasso is unable to resolve this problem, and apparently makes it worse, dramatically over-estimating the proportion of positive coefficients compared to a "naive" linear regression. The only way to identify the combined effect of the policies on outcomes is to first model the policy selection process.

1.4 Incorporating Measurement Error in Regression Analysis

We note that there are three potential ways that the measurement error, or uncertainty in the posterior estimates, of our indices can be incorporated in applied analyses. These methods all involve treating the indices as Normally-distributed random variables rather than observed values, where the standard deviation of the Normal distribution for each observation is taken from the standard deviation of the indices' posterior draws. We briefly discuss three methods.

The simplest approach is known as errors-in-variables models and have existed for decades (Durbin 1954). In these models, predictors can be included with a covariance matrix specifying Normally-distributed measurement error. This covariance matrix then inflates the standard errors in ordinary least squares (OLS) regressions. The covariance matrix, however, only specifies a single value (standard deviation) of measurement error for each index variable.

By contrast, two approaches exist that allow measurement error to be included at the observation level, which is potentially important as uncertainty can increase or decrease over time as our previous figures show. The first and most direct method is to include the indices in a Bayesian regression model as Normally-distributed random variables with a mean specified by the index value and a standard deviation from the estimated posterior deviation (Kurz 2020). This estimation is quite powerful in that it completely replaces the predictors with random variables with uncertainty that can vary over time.

An analogue to this approach can be achieved via frequentist methods via bootstrapping (Bizarro, Coppedge, and Pemstein 2016). For each draw of the posterior values of the indices, a model can be fit with the indices as data. The estimates can then be calculated as the average coefficient estimate over these draws, and the confidence interval as the 5% to 95% empirical interval of these estimates.

In the analyses in which we model the association of our indices with reported contact rates, we employ Bayesian specifications to directly incorporate measurement error at the observation level (i.e., including the estimated posterior standard deviation for each country-day index estimate). To look at associations between contact rates and these indices, we include the posterior median values of each index assuming Gaussian measurement error with estimated standard deviation $\hat{\sigma_{it}}$ from the posterior standard deviation of each index estimate $\hat{x_{it}}$ for country i and day t. To do so, we employ the R package `brms` (Bürkner 2017)

with default weakly informative priors.

To incorporate the fact that policymakers are simultaneously implementing the six different types of policy indices we predict, we allow for the errors to be correlated across specifications. To do so, we directly model the estimated posterior variance of the for a set of countries I, days T and policy scores P,

$$\hat{x_{itp}} \sim \text{Normal}(\mu_j, \sigma_j) \tag{20}$$

$$\mathbf{M} \sim \mathbf{MVN}([0, 0, 0, 0, 0, 0]', \Omega) \tag{21}$$

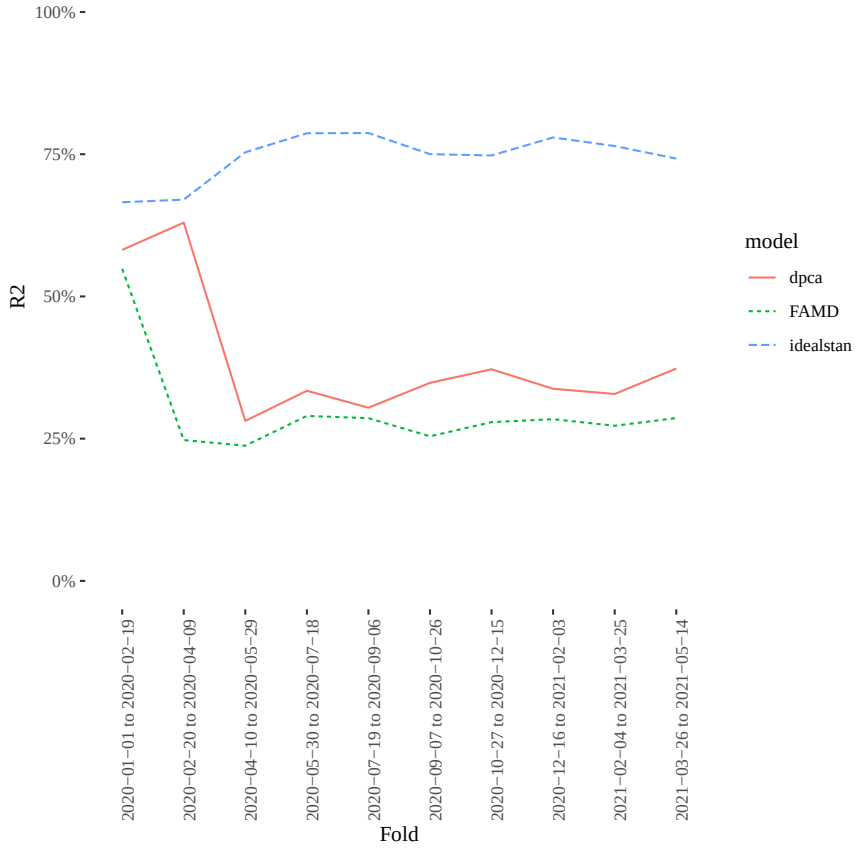
where σ_j is the observed posterior standard deviation of each policy intensity score, $\mathbf{M}$ is a vector with P elements containing the mean of the latent policy intensity scores μ_j and Ω is an estimated IxI covariance matrix of unmodeled (residual) correlation between the means of the latent scores. This correlation matrix can help provide an indication of strategic behavior when policymakers change the policy intensity level, such as whether the different policy types are substitutes (inversely correlated) or complements (positively correlated).

In our replication and extension of Flaxman, Mishra, Gandy, et al. (2020), we employed the same Bayesian susceptible-infected-recovered (SIR) model specification both with and without measurement error in our indices. To do so, we modified their code to allow us to include predictors of R_t, the parameter in their model which measures the reproduction rate of SARS-CoV-2, as Gaussian-distributed variables with pre-specified standard deviations for our model with measurement error. For both models with and without measurement error, we constrain the predictor coefficients to be positive to achieve identification, as the authors did in their original paper. The main difference with the original specification is that we did not use varying effects for the lockdown policy category given that lockdowns do not feature as prominently in our results.

1.5 Evaluation of Predictive Performance

While the simulation shows how our approach is able to parse measurement error, in this section we compare the predictive performance of our model to both the dynamic PCA specification by (Hörmann, Kidziński, and Hallin 2015) and a version of factor analysis for mixed data (i.e., categorical and continuous data) by Pagès (2004). We cannot compare to other IRT specifications as these all assume categorical data of different types. These PCA and factor analysis models are similar in that they are estimating one-dimensional models, though they do not take into account the grouping structure (time series nested within countries). However, they are the best available alternative models. We compare predictive performance using the R^2 statistic (proportion of variance explained). Given that all of the specifications use a single latent variable, this statistic helps capture how much each type of latent variable is able to incorporate the variation in our data. We cannot implement prediction metrics using training and test splits because predicting out of sample requires first predicting values of the latent variable. While this can be implemented

Figure 4: Model Performance Metrics



Plot shows calculated R^2 values from a 10-fold cross-validation of data from our social distancing index for 10 countries. Our model comes from the 'idealstan' package and is compared to dynamic PCA ('dpca') and mixed data factor analysis ('FAMD').

using our specification as we have access to the full range of model coefficients, it unfortunately unavailable from other existing methods. However, the R^2 is still able to give us a reasonable framework for evaluating predictive performance for 1-dimensional latent variable methods.

We estimate R^2 on a subset of 10 countries from our social distancing dataset as this index has the most variation at the policy level. We further implement 10-fold cross-validation by taking sequential sequences of approximately five weeks from the dataset, starting in January 1st, 2020 up until May 1st, 2021. Figure 4 shows the R^2 values on the y axis with the different folds of the cross-validation in the x axis in increasing order. As can be seen, the performance of factor analysis and PCA is modestly less than our model, 'idealstan' for the first 12 weeks of the time series. Beyond that, though, 'idealstan' remains highly predictive with an R^2 of about 0.80 while the other two models struggle to reach an R^2 higher than 0.4. This big gap in performance is likely due to the fact that there is increasing variation in the policy data in later time periods, and because 'idealstan' appropriately handles both over-time dynamics, mixed data types, and the grouped structure of the data, it is much better able to "explain" variance in the observed data.

We believe that this exercise helps to confirm the simulation analysis by showing that our model is also

highly predictive of the data, providing another source of validation for our chosen model.

2 Application to Estimating COVID-19 Suppression

We next turn to illustrating the use of our indices as predictors of COVID-19 mortality outcomes with a replication and extension of Flaxman et al.'s study of government policies' suppression effects (Flaxman, Mishra, Gandy, et al. 2020). In their analysis of the early pandemic in Europe, each of the variables they measured (lockdowns, public events restrictions, school closures, self-isolation policies and public encouragements of social distancing) was coded as either 0 before policy adoption or 1 for every day following policy adoption to the end of the sample period (31 December, 2019 to 4 May, 2020). While the study was remarkably successful at estimating the R_t (time-varying reproduction number) of SARS-CoV-2 spread from limited and biased data, the authors noted that a limitation of their findings came from coding issues with their policies:

> "Most interventions were implemented in rapid succession in many countries, and as such it is difficult to disentangle the individual effect sizes of each intervention (p. 260)."

This difficulty in disentangling effect sizes stems partially from the fact that their five policy variables are highly correlated. In their response (Flaxman, Mishra, Scott, et al. 2020) to a critique of their estimates (Soltesz et al. 2020), the authors again noted that it was impossible to fully resolve these issues without "more granular information on NPIs [that] will in future hopefully give a more nuanced understanding of which measures—whether mandatory or voluntary—contributed most to reductions in transmission."

As can be seen in the bottom half of Figure 5, our policy intensity scores for the same sample period never have a correlation value higher than 0.74 and are often as low as 0.21, indicating that the information in the indices still varies even with the similar country responses during the early period of the pandemic. For this reason, our policy intensity scores help resolve the data reduction problem by eliminating highly correlated variables and substituting a much more informative continuous measure with a clear theoretical basis.

Moreover, the advantage for applied modeling is not only theoretical. Due to these highly correlated measures, Flaxman et al. could not obtain estimates when including all policy indicators as predictors of R_t. Figure 1 in the SI shows their multivariate results, which show that only lockdowns are able to reduce R_t. As the authors note in their supplemental information (Figure 19), when each of their policy indicators are entered separately, they all are associated with very large reductions in R_t. In other words, the very large effect on R_t attributed to lockdowns in SI Figure 1 is partly an artifact of the high correlation of the policy indicators.

To show how our indices provide more information, we report our re-analysis of the Flaxman et al. model with all six of our policy scores in Figure 6. We estimated two specifications; the first with our scores as covariates without measurement error and the second with normally-distributed measurement error as described in the Methods section in the SI.

The low correlation between covariates is a result of the fact that our estimation approach of our indices

Figure 5: Comparison of Correlation of Policy Measures in Flaxman et al. and the Policy Intensity Scores

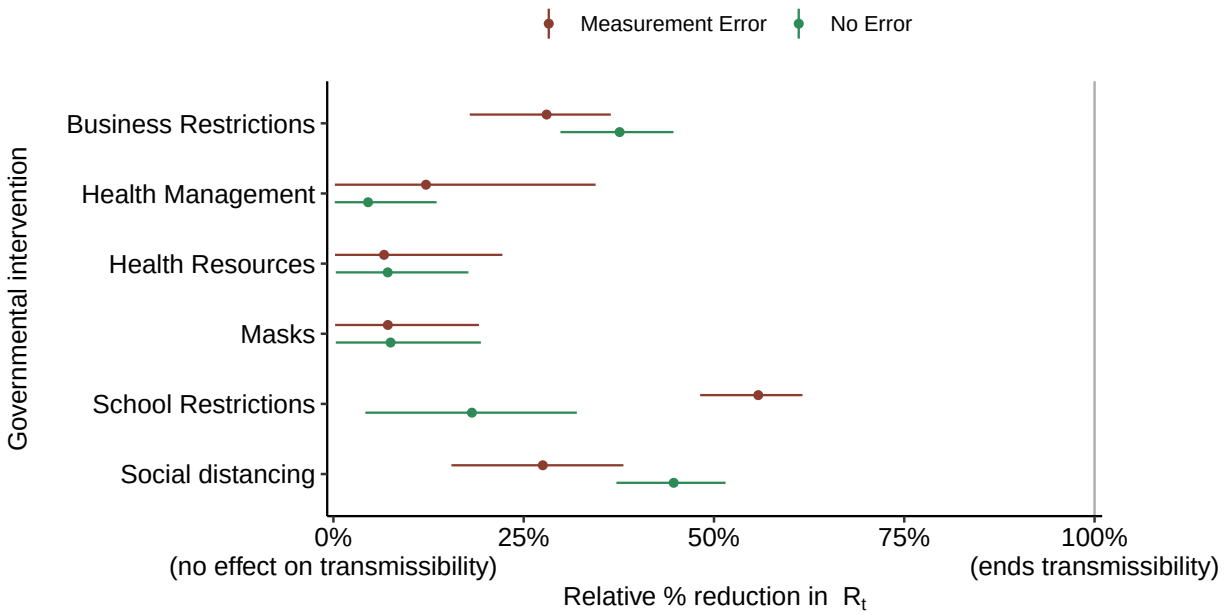
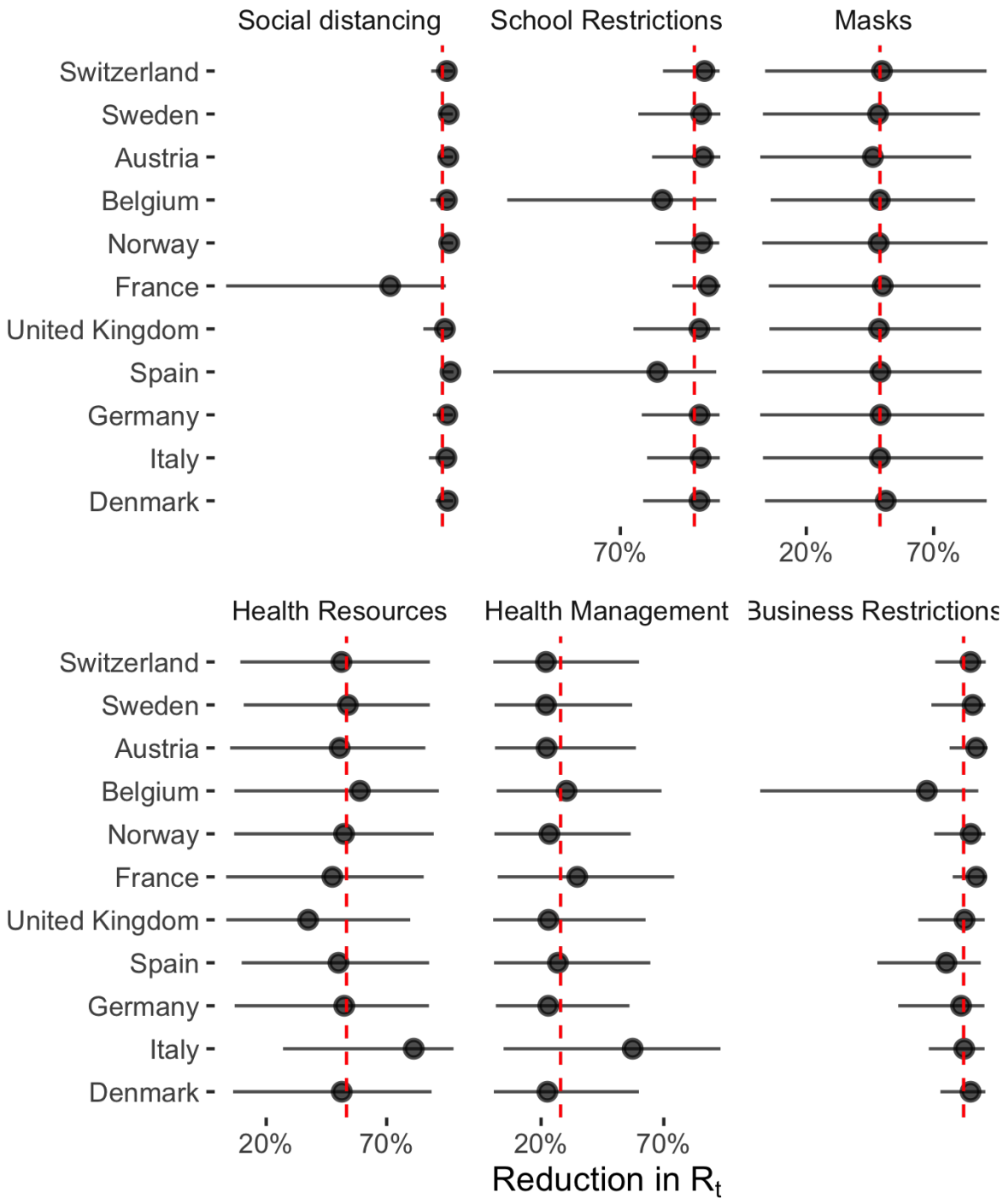


Figure 6: Flaxman et al. Replication Using Policy Index Scores both with and without Measurement Error

employed a diverse variety of data in each policy type. This means that the coefficients in Figure 6 can take on a variety of values. We also note that uncertainty in general is much higher than the original results. On the whole, both models with and without measurement error show that most policies had modest effects on COVID-19 outcomes, while there is disagreement between the models about whether general social distancing or school restrictions had the greatest effect on reducing R_t.

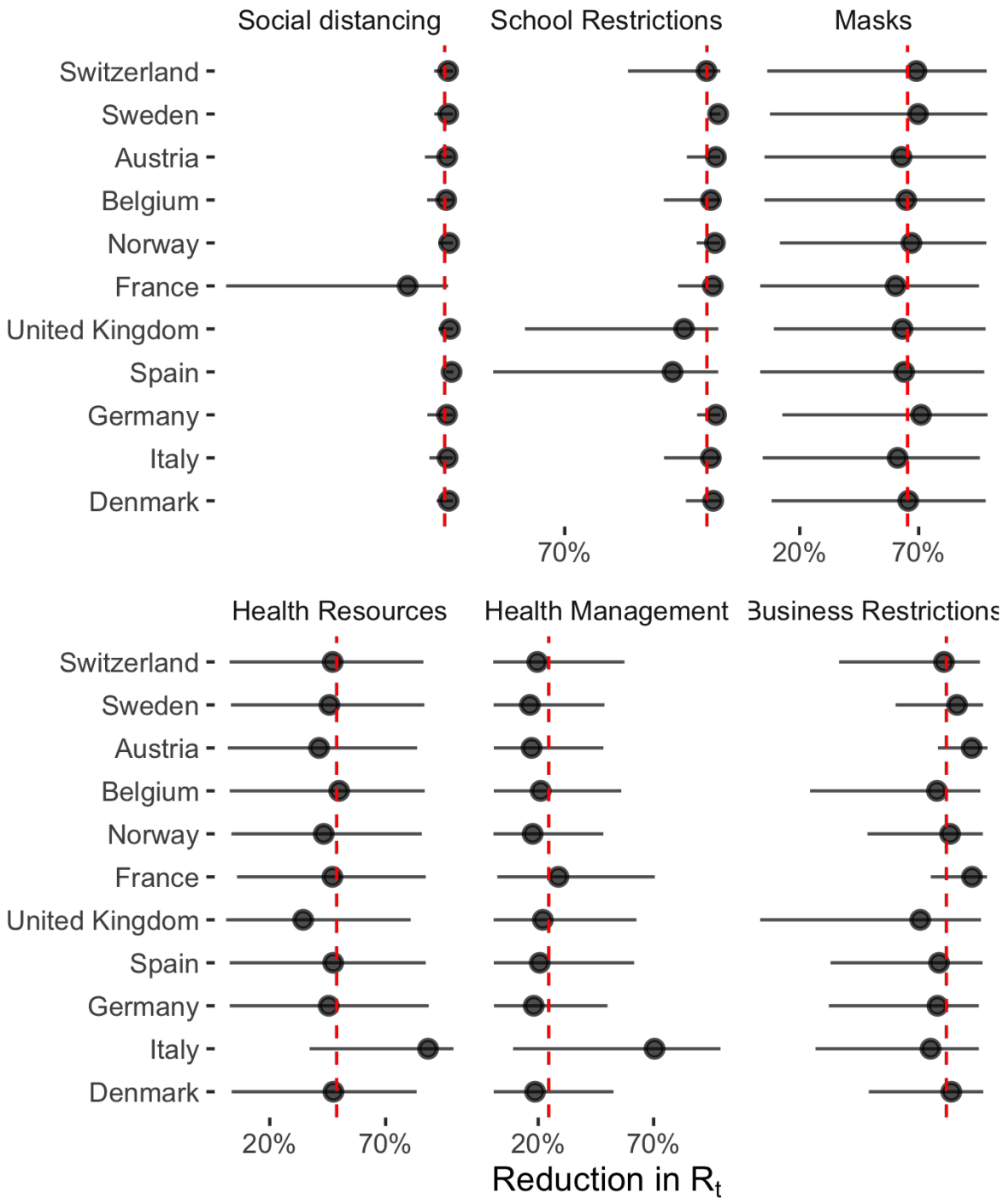
Figure 6 shows that the model with measurement error tends to put more weight on the intensity of school restrictions while the model without measurement error gives a greater role to general social distancing policies. This discrepancy in the results shows why incorporating measurement error does not only inflate uncertainty, but can also change substantive conclusions. In this case, the measurement error propagates through the non-linear model in a way that is difficult to know a priori.

Another statistical advantage to using measures with more variation is that the estimates are less vulnerable to arbitrary changes in the data. In Figures 7 and 8 we show the model's estimates of the NPIs' effects on R_t while leaving out one country at a time. As can be seen, the estimates are quite stable even though the original study had only 11 countries.



Plot shows the Flaxman et al. (2020) SIR model fit when leaving out a country specified on the x axis. The value of the R_t reduction parameter for a given index without measurement error is shown relative to the average parameter estimate as a dotted red line. Intervals are the 5% - 95% quantiles of the empirical posterior.

Figure 7: Leave-one-country-out Models without Measurement Error



Plot shows the Flaxman et al. (2020) SIR model fit when leaving out a country specified on the x axis. The value of the R_t reduction parameter for a given index with measurement error is shown relative to the average parameter estimate as a dotted red line. Intervals are the 5% - 95% quantiles of the empirical posterior.

Figure 8: Leave-one-country-out Models with Measurement Error

3 Additional Figures

Based on the above simulation, we re-examined the existing literature with an eye to these false inferences. One of the first influential studies of non-pharmaceutical interventions (NPIs) by Flaxman, Mishra, Gandy, et al. (2020) made it clear that high correlation between policies would make it difficult to obtain estimates of partial effects, a difficulty the study authors were unable to resolve satisfactorily. We have found further evidence in the literature that studies with linear regression models did at times estimate *increased* COVID-19 infections for some NPIs in, but these results were ignored or even manipulated away. The Haug et al. (2020) study, cited 1,021 times on Google Scholar to date, reports estimates of varying strength in repressing the COVID-19 reproduction number R_t across linear model specifications in their Figure 1 but do not code any results as doing the opposite. In their supplementry material, though, Figure 13 shows that their lasso regression with CoronaNet policy indicators estimated a strongly positive association between R_t and Quarantine policies. On the other hand, when they enter each policy indicator separately into a regression model in supplementary Figure 12, each CoronaNet policy indicator has a negative or zero association with R_t, suggesting that the positive association for Quarantine was related to endogeneity between indicators. For the reader's reference, we include annotated versions of these figures in our SI Additional Figures section.

Similarly, the Sharma et al. (2020) analysis of NPIs in Europe reports only R_t reductions or no associations for NPIs in the main text (see Figure 2 in the main text). However, these NPI effects are an artifact of a highly informative prior they placed on the NPI regression coefficients, biasing them towards reductions in R_t (see p. 7 top of column 2), a prior they also employed in an earlier study of NPIs (Brauner et al. 2020). When the authors employed a weakly informative prior centered on no effect, they instead show opposite associations of increasing R_t for some NPIs, including restricting public gatherings and limiting the number of people who can congregate in a private residence–but this information is only reported in their supplementary Figure 13, and the most constrasting effect is censored from the plot due to axis limits. Again, an annotated figure from their paper is available for reference in our SI Additional Figures section.

Figures 9 and 10 show estimates reported in the SI in Haug et al. (2020). Both plots are based on the CoronaNet data, one of several datasets they employ in the article. None of these estimates are reported in the main text, nor do the authors comment on the peculiar pattern of some policies showing positive associations with the COVID-19 reproduction rate R_t. Figure 9 shows each CoronaNet policy indicator entered separately into a regression model, which show either negative associations with R_t or no clear association. On the other hand, Figure 10 reveals that a multivariate model with all of these estimates shows very precisely estimated positive associations with R_t on the bottom of the plot, especially for Quarantine policies. They estimate that Quarantine policies increase R_t by approximately 15%, which would run very strongly against our substantive knowledge about how the virus spreads through interpersonal contact and airborne transmission.

Figure 11 shows a SI figure from Sharma et al. (2020) with different estimates of the effects of NPIs in their paper with different prior choices. The results in the main text use an informative "assymetric Laplace"

Figure 9: Haug et al. Bivariate Results with Solely Negative Associations with R_t

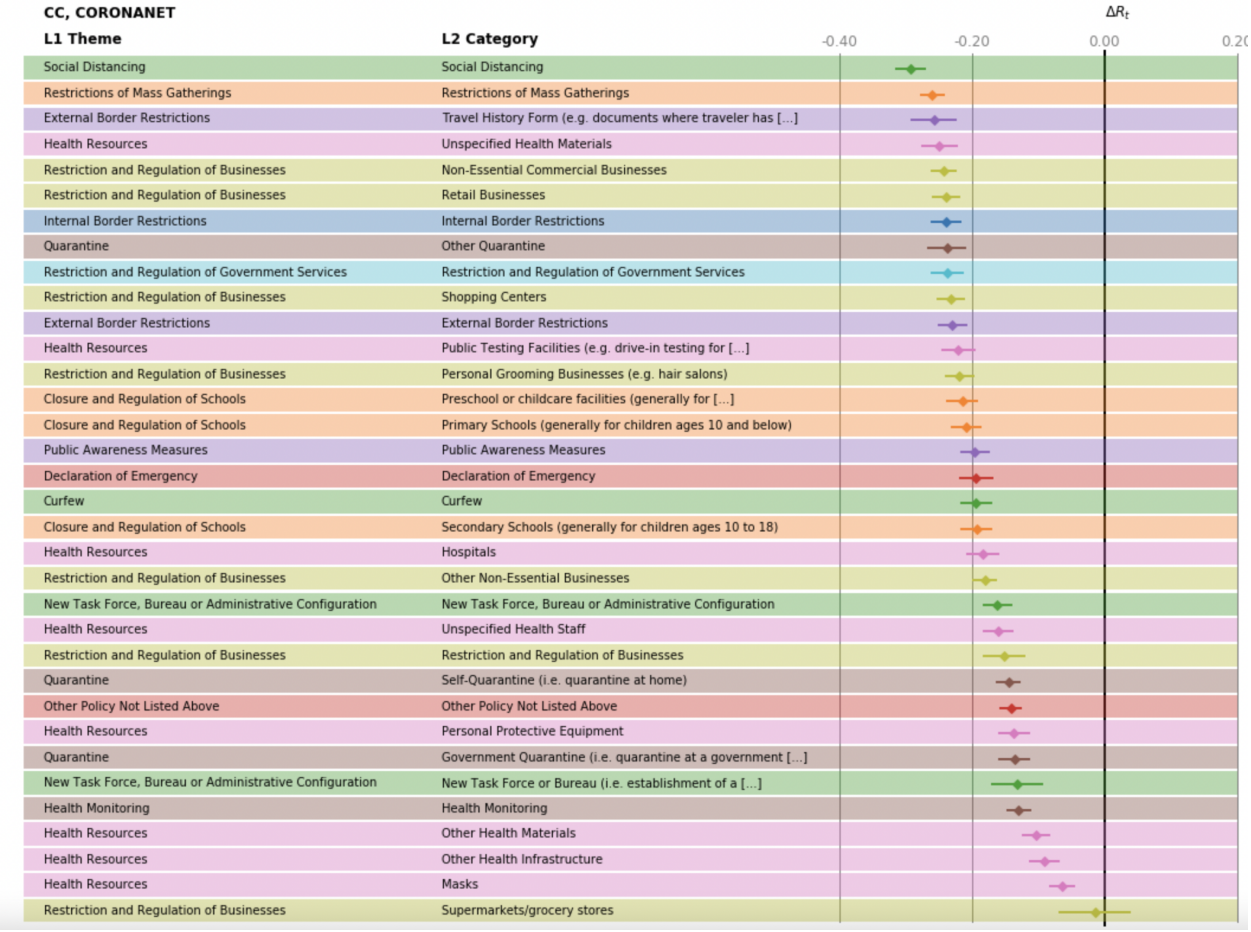


Figure shows SI Figure 12 from Haug et al. (2020). In this model, which they refer to as a case control model, each policy indicator was entered into their hierarchical R_t regression separately, showing that without endogeneity between the indicators, the associations are all negative (i.e. reducing R_t).

prior that concentrates 80% of the prior mass on positive effects of NPIs. This informative prior means that a "null" effect in their study is in fact a positive NPI effect; one consequence of this prior is that adding an NPI that is essentially random noise into their study results in a positive association with reducing R_t.[1] With this informative prior in place in Figure 11, all NPIs have either zero or positive associations with reducing R_t.

However, as is noted in the Figure, when they instead employ a weakly informative Normal prior centered around zero, or no association with R_t, they instead do have NPIs that are negatively associated with reducing R_t (or positively associated with R_t). In fact, one of the NPIs, banning mixing in private households, has an association with reduced R_t that is so negative that the estimate is censored due to the x axis limits. We believe this exclusion to be a typographical error, but nonetheless highlights the sensitivity of the estimates

1. See the E-Letters discussion on the article webpage of Brauner et al. (2020) at https://www.science.org/doi/10.1126/science.abd9338. Dostal, Furst and Vencalek in the E-Letters noted this problem and replicated their work with a random NPI, showing that the random NPI would also reduce R_t. The authors of the original study responded that this was expected behavior: "The result simply reflects that the prior also has a median above zero - a choice that reflects our prior belief that NPIs are more likely to reduce R_t than to increase it."

Figure 10: Haug et al. Multivariate Results with Positive Associations with R_t

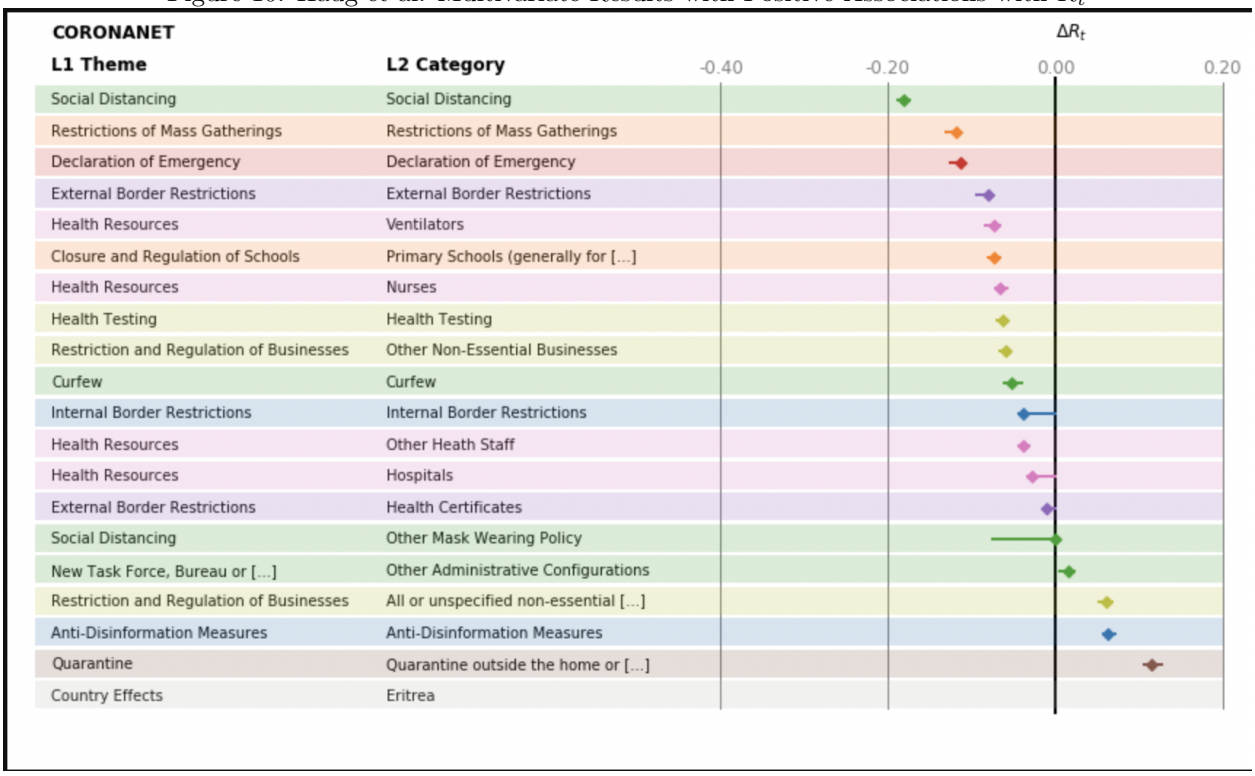


Figure shows SI Figure 13 from Haug et al. (2020). In this model, they employed a Lasso correction in their hierarchical R_t regression while including all policy indicators from the CoronaNet. In this case, the Quarantine policy has a strong positive association with increased R_t, which is at odds with the negative association in their SI Figure 12.

to prior choices that are not explained in the article.

In summary, both of these papers show that there are significant problems with identifying the effects of NPIs with observational data when multiple indicators exist for the same policy. Existing studies took a variety of approaches to address this, sometimes recoding data, sometimes burying inconsistent results in appendices and sometimes even adopting strongly informative priors that eliminated the problem deus ex machina. We believe that these inconsistent results are in fact related to the problem of identifying over-arching policy effects from multiple components of the same policy goal and consequent endogeneity between the indicators.

Figure 11: Sharma et al. Results with Varying Priors and Censored Estimates

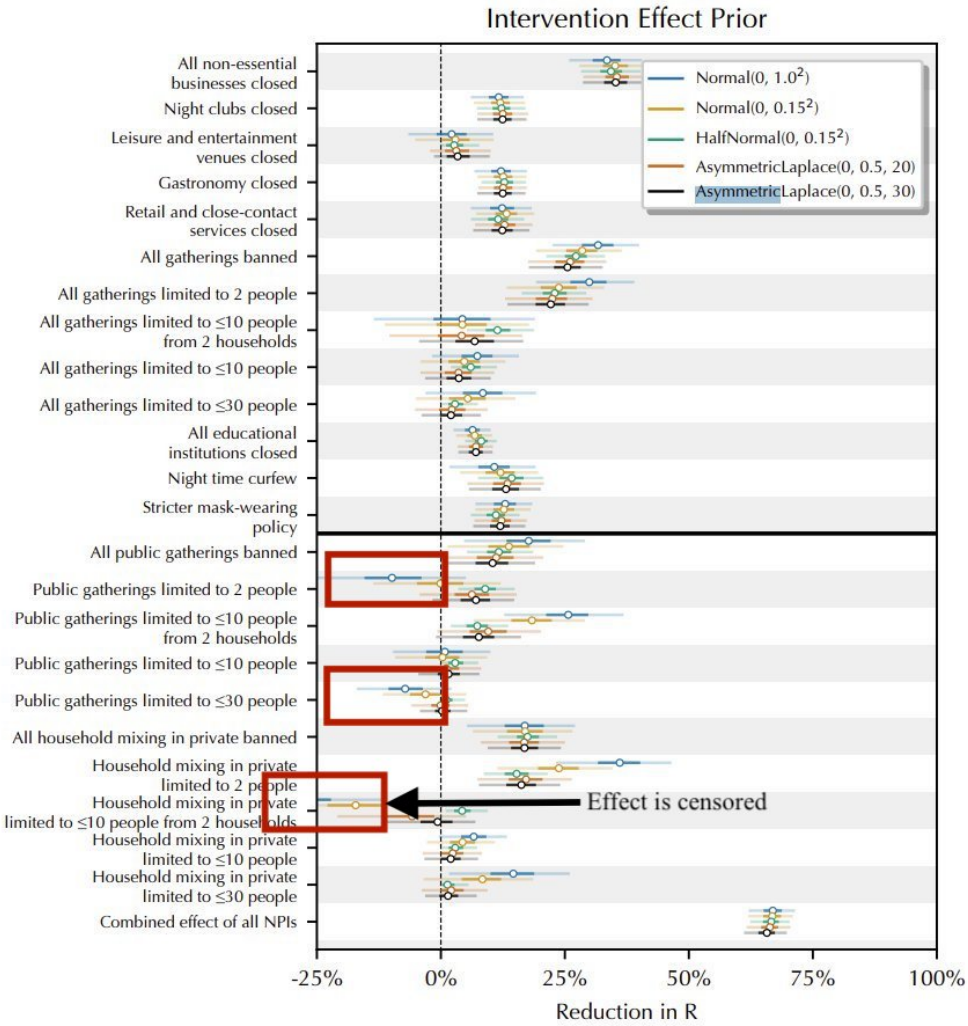
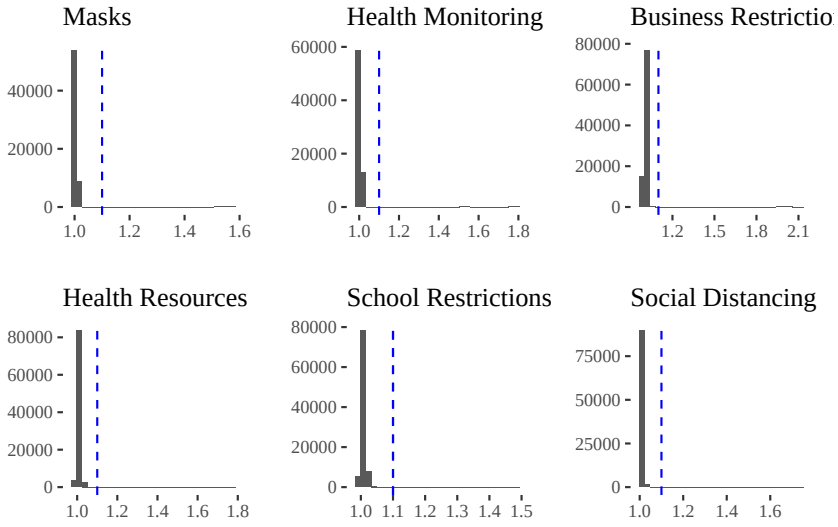


Figure shows SI Figure 14 from Sharma et al. (2020). The figure shows robustness checks in which they varied the prior they placed on the coefficient measuring the association between NPIs and R_t. Positive values indicate an association with *reduced* R_t. As can be seen, when using a weakly informative prior, they do indeed have associations between policy indicators and *increased* R_t.

4 MCMC Convergence

As mentioned in the main text, for each index we estimated four separate MCMC chains. In none of the chains did we observe transitions reaching the maximum treedepth or divergent transitions, which suggests that sampling of the full joint posterior was possible. We report histograms of the split-$\hat{R}$ values for the index parameters of the six indices in Figure 12. The vertical blue line indicates the value of 1.1 above which it is possible that the parameter may not have converged to a stationary distribution. As can be seen, almost all of the parameters are well below this threshold, providing additional evidence that the MCMC chains converged and the uncertainty measures are valid.

Figure 12: Convergence Statistics for MCMC Runs



We note that to improve sampling, we dropped days in which there was no change in policy indicators for some policy goals, especially health monitoring and masks which showed long periods of stasis. We then rolled down estimates from early time periods to complete the time series.

5 CoronaNet Codebook

In the attached document we show the definition of all the indicators from the CoronaNet dataset. Unlike previous versions of CoronaNet, this dataset release is in wide form, i.e., one observation per country per indicator value. To do so we reshaped the existing data from a list of policy records with start and end dates to a set of values for each policy showing either that it was in effect or not in effect.

CoronaNet Wide Data Codebook

This codebook provides variable descriptions for the indicators that are released in our wide data files and also used in the construction of our indices (see Kubinec et al. 2021). By wide data, we mean that every row of this data represents one day of policy records for a given country and each column is a separate policy indicator. Some of these columns/variables are binary while others are ordinal or continuous, details provided below.

The first wide data release covers January 1st, 2020 to January 15th, 2021. Note that there are 201 countries in this released data, though only a subset of those are used in the indices due to relatively few policy records for some countries and also missing data for these countries in the Oxford COVID-19 Government Response Tracker which is used in index creation.

This data release is considered provisional and will be updated as we collect more data over time and add and recode existing policy records. In particular, it may be the case that some policies have ended or will end and we do not have those end dates up to date in the present release. If there are any issues with the data, please post them as issues on our Github page, https://github.com/saudiwin/corona_tscs.

The data are organized by index, and we also include the variable/column in our original, long dataset from which the indicators were created. The original/long dataset generally has more information about the actual policy records from which this data was created.

School Restrictions Index

Original CoronaNet variable: type_school

preschool. Takes a value of 1 if no restrictions are placed on preschool or childcare facilities (generally for children ages 5 and below), a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

primary_school. Takes a value of 1 if no restrictions are placed on primary schools (generally for children ages 10 and below), a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

secondary_school. Takes a value of 1 if no restrictions are placed on secondary schools (generally for children ages 10 to 18), a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

higher_ed. Takes a value of 1 if no restrictions are placed on higher education institutions (i.e. degree granting institutions), a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

Original CoronaNet variable: type_school_var

Conditions under which type_school are allowed to operate.

school_clean. Takes a value of 1 if regular cleaning and sanitary procedures are required as a condition of schools operating.

school_other. Takes a value of 1 if other conditions are required as a condition of schools operating.

school_num. Takes a value of 1 if the number of people on the school premises is limited as a condition of schools operating.

school_type_pers. Takes a value of 1 if the types of people on the school premises are limited as a condition of schools operating.
school_health_q. Takes a value of 1 if health questionnaires are required as a condition of schools operating.

school_special_student. Takes a value of 1 if special provisions exist for all students in a school (e.g. students in primary school do not have to social distance).

school_special_teacher. Takes a value of 1 if special provisions exist for how teaching is done (e.g. teachers must tele-teach).

school_temp. Takes a value of 1 if temperature checks are required as a condition of schools operating.

school_health_monitoring. Takes a value of 1 if other health monitoring measures are required as a condition of schools operating.

Business Restrictions Index

Original CoronaNet variable: type_business

biz_restrict_all. Takes a value of 1 if no restrictions are placed on all or unspecified businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_rest. Takes a value of 1 if no restrictions are placed on restaurants and/or bars, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_comm. Takes a value of 1 if no restrictions are placed on commercial businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_retail. Takes a value of 1 if no restrictions are placed on retail businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_shop. Takes a value of 1 if no restrictions are placed on shopping centers, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_groom. Takes a value of 1 if no restrictions are placed on personal Grooming Businesses (e.g. hair salons), a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_other. Takes a value of 1 if no restrictions are placed on other businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_grocery. Takes a value of 1 if no restrictions are placed on supermarkets/grocery stores, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_telecom. Takes a value of 1 if no restrictions are placed on telecommunications businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_info. Takes a value of 1 if no restrictions are placed on information service activities, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_publish. Takes a value of 1 if no restrictions are placed on publishing activities, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_construct. Takes a value of 1 if no restrictions are placed on construction businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_farm. Takes a value of 1 if no restrictions are placed on agriculture, forestry and fishing industries, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_transport. Takes a value of 1 if no restrictions are placed on transportation (land, water and air), a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_hotel. Takes a value of 1 if no restrictions are placed on paid lodgings (e.g. hotels, motels), a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_warehouse. Takes a value of 1 if no restrictions are placed on warehousing and support activities for transportation, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_health. Takes a value of 1 if no restrictions are placed on private health offices, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_pharmacy. Takes a value of 1 if no restrictions are placed on pharmacies, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_water. Takes a value of 1 if no restrictions are placed on water supply, sewerage, waste management and remediation activities, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_finance. Takes a value of 1 if no restrictions are placed on financial service activities except insurance and pension funding, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_mining. Takes a value of 1 if no restrictions are placed on mining and quarrying, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_restrict_insurance. Takes a value of 1 if no restrictions are placed on insurance; reinsurance, and pension funding except compulsory social security, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_takeout. Takes a value of 1 if businesses are allowed to have take-out ordering.

biz_delivery. Takes a value of 1 if businesses are allowed to have delivery ordering.

biz_nonessential. Takes a value of 1 if no restrictions are placed on non-essential businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

biz_essential. Takes a value of 1 if no restrictions are placed on essential businesses, a value of 2 if these facilities are partly open, and a value of 3 if they are completely closed.

Original CoronaNet variable: type_business_var

conditions under which type_business are allowed to operate

biz_hygiene. Takes a value of 1 if hygiene and sanitation measures are required as a condition of business operating.

biz_hours. Takes a value of 1 if the number of working hours is limited as a condition of business operating.

biz_work_home. Takes a value of 1 if the number of employees and working hours are limited as a condition of business operating.

biz_meeting. Takes a value of 1 if the size of business meetings is limited as a condition of business operating.

biz_social_distance. Takes a value of 1 if keeping a distance of at least 6ft or 1.5 meters apart is required as a condition of business operating.

biz_mask. Takes a value of 1 if mask wearing is required as a condition of business operating.

biz_temperature. Takes a value of 1 if temperature checks are required as a condition of business operating.

biz_health_cert. Takes a value of 1 if health certificates are required as a condition of business operating.

biz_health_q. Takes a value of 1 if health questionnaires are required as a condition of business operating.

biz_num_cust. Takes a value of 1 if the number of customers is limited as a condition of business operating.

biz_store_size. Takes a value of 1 if the size of stores is limited as a condition of business operating.

biz_cont_trace. Takes a value of 1 if contact tracing is required as a condition of business operating.

biz_cond_other. Takes a value of 1 if other conditions are required as a condition of business operating.

Health Resources Index

Original CoronaNet variable: type_health resources

hr_cold_storage. Takes a value of 1 if there is a policy which affects cold storage capacity for COVID-19 vaccines.

hr_doctors. Takes a value of 1 if there is a policy which affects the number and allocation of doctors.

hr_dry_ice. Takes a value of 1 if there is a policy which affects dry ice for COVID-19 vaccine storage.

hr_sanitizer. Takes a value of 1 if there is a policy which affects quantity and allocation of hand sanitizer.

hr_insurance. Takes a value of 1 if there is a health testing policy which affects health insurances.

hr_facilities. Takes a value of 1 if there is a health testing policy which affects health research facilities.

hr_volunteers. Takes a value of 1 if there is a policy which affects quantity and allocation of health volunteers.

hr_hospitals. Takes a value of 1 if there is a health testing policy which affects hospitals.

hr_masks. Takes a value of 1 if there is a policy which affects quantity and allocation of face masks.

hr_drugs. Takes a value of 1 if there is a policy which affects quantity and allocation of pharmaceutical drugs.

hr_nurses. Takes a value of 1 if there is a policy which affects quantity and allocation of nurses.

hr_other_infra. Takes a value of 1 if there is a policy which affects “other” health infrastructure.

hr_other_mat. Takes a value of 1 if there is a policy which affects “other” health material.

hr_other_staff. Takes a value of 1 if there is a policy which affects “other” health staff.

hr_ppe. Takes a value of 1 if there is a policy which affects quantity and allocation of personal protective equipment.

hr_testing. Takes a value of 1 if there is a policy which affects public testing facilities.

hr_syringe. Takes a value of 1 if there is a policy which affects quantity and allocation of syringes.

hr_quarantine. Takes a value of 1 if there is a health testing policy which affects temporary quarantine centers.

hr_pcr. Takes a value of 1 if there is a policy which affects quantity and allocation of thermal cyclers (PCR machines/DNA amplifiers).

hr_ventilator. Takes a value of 1 if there is a policy which affects quantity and allocation of ventilators.

hr_test_kit. Takes a value of 1 if there is a policy which affects quantity and allocation of test kits.

hr_target_staff. Takes a value of 1 if the target of a policy is health staff.

hr_target_supply. Takes a value of 1 if the target of a policy is health-related supplies.

Health Monitoring Index

Original CoronaNet variable: type_health_mon

hm_home_visit. Takes a value of 1 if home visits are used to support the government's health monitoring strategy.

hm_other_mon. Takes a value of 1 if “other” human health monitoring strategy is used to monitor citizens’ health status.

hm_telephone. Takes a value of 1 if telephone calls are used to support the government's health monitoring strategy.

hm_loc_nursing. Takes a value of 1 if health monitoring takes place in nursing homes.

hm_loc_other. Takes a value of 1 if health monitoring takes place in “other” locations.

hm_loc_subway. Takes a value of 1 if health monitoring takes place in subways and trams.

hm_loc_buses. Takes a value of 1 if health monitoring takes place in buses.

hm_loc_trains. Takes a value of 1 if health monitoring takes place in trains.

hm_loc_nursing. Takes a value of 1 if health monitoring takes place in nursing homes.

hm_cert. Takes a value of 1 if health certificates are used to monitor a person's health at a given point in time.

hm_q. Takes a value of 1 health if questionnaires are used to monitor a person's health at a given point in time.

hm_snap_other. Takes a value of 1 if an "other" type of a snapshot is used to monitor a person's health at a given point in time.

hm_snap_temp. Takes a value of 1 if temperature checks are used to monitor a person's health at a given point in time.

hm_stra_contact_human. Takes a value of 1 if contact tracing through human teams is used to monitor citizens' health status.

hm_stra_contact_phone. Takes a value of 1 if contact tracing through smart phones (e.g. apps)is used to monitor citizens' health status.

hm_stra_other. Takes a value of 1 if "other" strategy is used to monitor citizens' health status.

hm_stra_wearable. Takes a value of 1 if wearable technology (e.g. bracelets, anklets, beacons)is used to monitor citizens' health status.

hm_tech_bluetooth. Takes a value of 1 if bluetooth is used to monitor citizens' health status.

hm_tech_gps. Takes a value of 1 if GPS is used to monitor citizens' health status.

hm_tech_qr. Takes a value of 1 if QR is used to monitor citizens' health status.

hm_tech_other. Takes a value of 1 if "other" technology is used to monitor citizens' health status.

Health Testing Index

Note: the health testing indicators were merged into the Health Monitoring index.

Original CoronaNet variable: type_health_testing

ht_door2door. Takes a value of 1 if the government employed door-to-door testing.

ht_drivein. Takes a value of 1 if the government employed drive-in testing.

ht_fixed. Takes a value of 1 if the government employed fixed health testing stations.

ht_drivein. Takes a value of 1 if the government employed drive-in testing.

ht_entire_pop. Takes a value of 1 if the government employed health testing of the entire population under the government's jurisdiction.

ht_mobile. Takes a value of 1 if the government employed mobile health testing stations.

ht_other. Takes a value of 1 if the government employed "other" health testings.

ht_type_antibody. Takes a value of 1 if the government employed antibody/serological tests.

ht_type_antigen. Takes a value of 1 if the government employed antigen tests.

ht_type_other. Takes a value of 1 if the government employed "other" test type.

ht_type_pcr. Takes a value of 1 if the government employed PCR tests.

ht_portal_email. Takes a value of 1 if email is used to inform people of their COVID-19 test results.

ht_portal_sms. Takes a value of 1 if mobile text is used to inform people of their COVID-19 test results.

ht_portal_app. Takes a value of 1 if an app or website is used to inform people of their COVID-19 test results.

ht_portal_other. Takes a value of 1 if "other" or not specified means of communication was used to inform people of their COVID-19 test results.

ht_portal_paper. Takes a value of 1 if paper results picked up in person are used to inform people of their COVID-19 test results.

ht_portal_phone. Takes a value of 1 if phone calls are used to inform people of their COVID-19 test results.

ht_cost_free_all. Takes a value of 1 if testing on COVID-19 is free for all individuals.

ht_cost_free_subset. Takes a value of 1 if testing on COVID-19 is free for a subset of the population.

ht_cost_partly_free. Takes a value of 1 if testing on COVID-19 is partially subsidized by the government for all individuals.

ht_cost_other. Takes a value of 1 if there is no information provided regarding costs of COVID-19 testing.

ht_cost_all_pay. Takes a value of 1 if all individuals must pay full cost of their COVID-19 test.

ht_cost_symptomatic. Takes a value of 1 if only symptomatic people are eligible for a test.

ht_loc_clinic. Takes a value of 1 if health clinics are used as health testing sites.

ht_loc_private. Takes a value of 1 if private doctors offices are used as health testing sites.

ht_loc_hospital. Takes a value of 1 if hospitals are used as health testing sites.

ht_loc_other. Takes a value of 1 if “other” location is used as a health testing site.

ht_loc_pharmacy. Takes a value of 1 if pharmacies are used as health testing sites.

Social Distancing Index

Original CoronaNet variable: type_social_distancing

social_distance. Takes a value of 1 if there is a policy for keeping a distance of at least 6 feet or 1.5 meters apart.

distance_other. Takes a value of 1 if there is a policy for keeping a distance of some other distance not listed above.

buses. Takes a value of 1 if there are restrictions on ridership of buses.

other_transport. Takes a value of 1 if there are restrictions on ridership of other forms of public transportation.

private_transport. Takes a value of 1 if there are restrictions on ridership private vehicles in public circulation.

subways. Takes a value of 1 if there are restrictions on ridership of subways and trams.

trains. Takes a value of 1 if there are restrictions on ridership of trains.

Original CoronaNet variable: type_mass_gatherings

curfew_length. Takes a value of 1-24 (hours).

number_mass. Takes a value of 1 if there is the most stringent mass gathering restriction and 0 if there is no mass gathering restriction.

cancel_annual_event. Takes a value of 1 if there is a cancellation of a recreational or commercial event.

prison_pop. Takes a value of 1 if there is a policy regarding reduction of prison population.

postpone_ann_event. Takes a value of 1 if there is a postponement of a recreational or commercial event.

postpone_rec_event. Takes a value of 1 if there is a postponement of an annually recurring event.

private_event. Takes a value of 1 if events at private residencies are restricted.

event_no_audience. Takes a value of 1 if events are allowed to occur without an audience.

Border Restriction Index

Original CoronaNet variable: travel_mechanism

int_restrict_flights. Takes a value of 1 if flights are restricted.

int_restrict_border. Takes a value of 1 if travel through land borders is restricted.

int_restrict_all. Takes a value of 1 if all kinds of transport across borders are restricted.

int_restrict_NA. Takes a value of 1 if the type of border restriction is not specified

int_restrict_cruises. Takes a value of 1 if cruise ships are restricted.

int_restrict_ferries. Takes a value of 1 if ferries are restricted.

int_restrict_ports. Takes a value of 1 if travel through seaports is restricted.

int_restrict_trains. Takes a value of 1 if trains are restricted.

int_restrict_buses. Takes a value of 1 if buses are restricted.

Mask Index

Variable group: type_mask

mask_public. Takes a value of 1 if there is a policy for wearing masks inside public buildings.

mask_everywhere. Takes a value of 1 if there is a policy for wearing masks in all public spaces.

mask_business. Takes a value of 1 if there is a policy for wearing masks inside private businesses (e.g. supermarkets).

mask_primary _school. Takes a value of 1 if there is a policy for wearing masks inside primary schools (generally for children ages 10 and below).

mask_sec_school. Takes a value of 1 if there is a policy for wearing masks inside secondary schools (generally for children ages 10 and to 18).

mask_transport. Takes a value of 1 if there is a policy for wearing masks inside public transportation.

mask_unspec. Takes a value of 1 if there is an unspecified mask wearing policy.

mask_preschool. Takes a value of 1 if there is a policy for wearing masks inside preschools or childcare facilities(generally for children ages 5 and below).

mask_higher _ed. Takes a value of 1 if there is a policy for wearing masks inside higher education institutions (i.e. degree granting institutions).

6 OxCGRT Codebook

Below we summarize the ten indicators we employed from the OxCGRT index and also identify which index they were assigned to. For more information about the OxCGRT dataset, we refer the reader to (Hale et al. 2021).

- Social Distancing Index
  - **C4_Restrictions on gatherings** – Level of restrictions on mass gatherings. Ordinal indicator with five categories.
  - **C5_Close public transport** – Closing of public transport. Ordinal indicator with 3 categories.
  - **C3_Cancel public events** – Banning of events open to the public. Ordinal indicator with 3 categories.
  - **C6_Stay at home requirements** – Policies requiring residents to remain at home. Ordinal indicator with 4 categories.
  - **C7_Restrictions on internal movement** – Policies preventing residents from traveling to areas within the country. Ordinal indicator with 3 categories.
- School Restrictions Index
  - **C1_School closing** – degree to which schools are closed. Ordinal indicator with 4 categories.
- Business Restrictions Index
  - **C2_Workplace closing** – degree to which workplaces are closed. Ordinal indicator with 4 categories.
- Health Monitoring Index
  - **H2_Testing policy** – extent of government involvement in creating access to COVID-19 tests. Ordinal indicator with 4 categories.
- Health Resources Index
  - **H4_Emergency investment in healthcare** – dollar amount of resources invested in health care-related expenses to fight the COVID-19 pandemic. Continuous variable. We convert this variable to a 30-day rolling cumulative sum expressed as dollars per capita for each country.
- Mask Index
  - **H6_Facial Coverings** – Extent to which face masks are mandated in the country. Ordinal indicator with 5 categories.

7 Composition of Indices

Hundreds of research projects have sprung up over the course of the pandemic with the aim of documenting the diverse ways that governments have implemented policies to slow the spread of COVID-19. These data projects themselves take a variety of approaches in documenting these policies. While some focus on a single type of policy (*UNDP and UN Women COVID-19 Global Gender Response Tracker* 2021; Elgin, Basbug, and Yalaman 2020) or a particular region of the world (Naqvi 2021; Adolph et al. 2021), others aim to collect data at world-wide scale across a range of indicators (*COVID-19 Government Measures Dataset* 2020; Porcher 2020; Grundy, Quinn, and Todowede 2021; Suryanarayanan et al. 2021). Of this latter set, the datasets with the widest coverage and most detailed indicators include CoronaNet, OxCGRT, the Complexity Science Hub COVID-19 Control Strategies List (CCCSL) (Haug et al. 2020; Desvars-Larrive et al. 2020) and Health Intervention Tracking for COVID-19 (HIT-COVID) (Zheng et al. 2020). The CCCSL and HIT-COVID datasets are similar in approach to the CoronaNet dataset in that they try to measure as many discrete attributes of policies as possible. The OxCGRT dataset, on the other hand, releases qualitative summary ordinal indices instead of disaggregated indicators that place country responses to COVID-19 on ordinal scales.

CoronaNet taxonomy relies on 20 policy categories and some of our indices directly correspond to its policy categories, merging numerous subcategories into fewer indicators. Concretely, our indices rely on the following indices:

Business Restrictions Index: Matches policy category Restriction and Regulation of Businesses in CoronaNet, which serves to track government policies that regulate private, commercial activity which can include closing down commercial establishments completely, allowing commercial establishments to open without conditions or allowing commercial establishments to open with conditions. The Business Restrictions Index uses the latter variable. We add the "Workplace Closing" variable (C2) from OxCGRT´s containment and closure policies.

Health Monitoring, and Health Resources Indices: Correspond to eponymous policy categories in CoronaNet. We add the "Testing policy" variable (H2) respectively the "Emergency investment in healthcare" (H4) variable from OxCGRT´s health system policies.

Mask Index: Applies indicators from Social Distancing, Restriction and Regulation of Businesses and Closure and Regulations of Schools in CoronaNet. We add the "Facial Coverings" variable (H6) from OxCGRT´s health system policies.

School Restriction Index: Matches policy category Closure and Regulations of Schools in CoronaNet, which serves to track government policies that regulate educational establishments in a country such as: closing an educational institution completely, allowing an educational institution to stay open without conditions and allowing an educational institution to open with certain conditions. Schools Restriction Index uses the latter variable, merging 19 CoronaNet subcategories into 9 indicators. We add the "School Closing" variable (C1) from OxCGRT´s containment and closure policies.

Table 1: Composition of Indices

Index	CoronaNet Indicators	OxCGRT Indicators
Business Restrictions	**Restriction and Regulation of Businesses** *type_business* *type_business_var*	**C - containment and closure policies** *C2_Workplace closing*
Health Monitoring	**Health Testing** *type_health_testing*	**H - health system policies** *H2_Testing policy*
Health Resources	**Health Resources** *type_health resources*	**H - health system policies** *H4_Emergency investment in healthcare*
Mask	**Social distancing** *Mask_public* *Mask_everywhere_var* *Mask_transport_var* *Mask_unspec_var* **Restriction and Regulation of Businesses** *Mask_business* **Closure and regulations of schools** *Mask_primary_school* *Mask_sec_school* *Mask_preschool* *Mask_higher_ed*	**H - health system policies** *H6_Facial Coverings*
School Restriction	**Closure and Regulations of Schools** *type_school* *type_school_var*	**C - containment and closure policies** *C1_School closing*
Social Distancing	**Closure and Regulations of Schools** *type_social_distancing* **Restrictions of mass gatherings** *type_mass_gatherings* **Curfew** *type_curfew*	**C - containment and closure policies** *C3_Cancel public events* *C4_Restrictions on gatherings* *C5_Close public transport* *C6_Stay at home requirements* *C7_Restrictions on internal movement*

Note: Policy mapping based on two datasets presented by **top level policy type** and *sub level variable*

Social Distancing Index: Merges subcategories from Social distancing, Restrictions of mass gatherings and Curfew policy category in CoronaNet. We add the "Cancel public events" (C3), "Restrictions on gatherings" (C4), "Close public transport" (C5), "Stay at home requirements" (C6), and "Restrictions on internal movement" (C7) variables from OxCGRT´s containment and closure policies.

8 References

References

Adolph, Christopher, Kenya Amano, Bree Bang-Jensen, Nancy Fullman, and John Wilkerson. 2021. "Pandemic Politics: Timing State-Level Social Distancing Responses to COVID-19." *Journal of Health Politics, Policy and Law* 46, no. 2 (April): 211–233. ISSN: 0361-6878, accessed April 19, 2021. https://doi.org/10.1215/03616878-8802162. https://doi.org/10.1215/03616878-8802162.

Bizarro, Fernando, Michael Coppedge, and Daniel Pemstein. 2016. *Incorporating v-dem's uncertainty estimates in regression analysis.* Working paper. April. Accessed August 11, 2016. https://kellogg.nd.edu/projects/vdem/files/2016/tutorial.pdf.

Brauner, Jan M., Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B. Stephenson, et al. 2020. "Inferring the effectiveness of government interventions against COVID-19" [in en]. *Science* (December). ISSN: 0036-8075, 1095-9203, accessed January 29, 2021. https://doi.org/10.1126/science.abd9338. https://science.sciencemag.org/content/early/2020/12/15/science.abd9338.

Bürkner, Paul-Christian. 2017. "brms: An R package for Bayesian multilevel models using Stan." *Journal of Statistical Software* 80 (1): 1–28. https://doi.org/10.18637/jss.v080.i01.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A probabilistic programming language." *Journal of Statistical Software* 76 (1).

Cheng, Cindy, Joan Barcelo, Allison Spencer Hartnett, Robert Kubinec, and Luca Messerschmidt. 2020. "COVID-19 government response event dataset (CoronaNet v.1.0)." *Nature Human Behavior,* https://doi.org/https://doi.org/10.1038/s41562-020-0909-7.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The statistical analysis of rollcall data." *American Political Science Review* 98 (2): 355–370.

COVID-19 Government Measures Dataset [in en]. 2020. Technical report. ACAPS, March. Accessed April 19, 2021. https://www.acaps.org/covid-19-government-measures-dataset.

Desvars-Larrive, Amélie, Elma Dervic, Nils Haug, Thomas Niederkrotenthaler, Jiaying Chen, Anna Di Natale, Jana Lasser, et al. 2020. "A structured open dataset of government interventions in response to COVID-19" [in en]. *Scientific Data* 7, no. 1 (August): 285. ISSN: 2052-4463, accessed March 31, 2021. https://doi.org/10.1038/s41597-020-00609-9. https://www.nature.com/articles/s41597-020-00609-9.

Durbin, J. 1954. "Errors in variables." *Review of the International Statistical Institute* 1 (22): 23–32.

Elgin, Ceyhun, Gokce Basbug, and Abdullah Yalaman. 2020. "Economic policy responses to a pandemic: Developing the COVID-19 economic stimulus index." *Covid Economics* 1 (3): 40–53.

Flaxman, Seth, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, et al. 2020. "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe." *Nature* 584, no. 7820 (August): 257–261. https://doi.org/10.1038/s41586-020-2405-7.

Flaxman, Seth, Swapnil Mishra, James Scott, Neil Ferguson, Axel Gandy, and Samir Bhatt. 2020. "Reply to: The effect of interventions on COVID-19" [in en]. *Nature* 588, no. 7839 (December): E29–E32. https://doi.org/10.1038/s41586-020-3026-x.

Grundy, Chris, Orlagh Quinn, and Sewedo Todowede. 2021. *Global Dataset of Public Health and Social Measures* [in en]. Technical report. World Health Organization. Accessed April 19, 2021. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, et al. 2021. "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)" [in en]. *Nature Human Behaviour* (March): 1–10. ISSN: 2397-3374, accessed March 31, 2021. https://doi.org/10.1038/s41562-021-01079-8. https://www.nature.com/articles/s41562-021-01079-8.

Haug, Nils, Lukas Geyrhofer, Alessandro Londei, Elma Dervic, Amélie Desvars-Larrive, Vittorio Loreto, Beate Pinior, Stefan Thurner, and Peter Klimek. 2020. "Ranking the effectiveness of worldwide COVID-19 government interventions" [in en]. *Nature Human Behaviour* 4, no. 12 (December): 1303–1312. ISSN: 2397-3374, accessed January 29, 2021. https://doi.org/10.1038/s41562-020-01009-0. https://www.nature.com/articles/s41562-020-01009-0.

Hörmann, Siegfried, Łukasz Kidziński, and Marc Hallin. 2015. "Dynamic functional principal components." Publisher: [Royal Statistical Society, Wiley], *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 77 (2): 319–348. ISSN: 1369-7412, accessed December 10, 2022. https://www.jstor.org/stable/24774739.

Kurz, A. Solomon. 2020. "Missing Data and Other Opportunities." In *Statistical Rethinking with brms, ggplot2, and the tidyverse.* Accessed August 11, 2021. https://bookdown.org/ajkurz/Statistical_Rethinking_recoded/missing-data-and-other-opportunities.html.

Naqvi, Asjad. 2021. "COVID-19 European regional tracker" [in en]. *medRxiv* (February): 2021.02.15.21251788. Accessed April 19, 2021. https://doi.org/10.1101/2021.02.15.21251788. https://www.medrxiv.org/content/10.1101/2021.02.15.21251788v2.

Pagès, Jérôme. 2004. "Analyse factorielle de donnees mixtes: principe et exemple d'application." *Revue de statistique appliquée* 52 (4): 93–111.

Porcher, Simon. 2020. "Response2covid19, a dataset of governments' responses to COVID-19 all around the world" [in en]. *Scientific Data* 7, no. 1 (November): 423. ISSN: 2052-4463, accessed April 19, 2021. https://doi.org/10.1038/s41597-020-00757-y. https://www.nature.com/articles/s41597-020-00757-y.

Sharma, Mrinank, Sören Mindermann, Jan Markus Brauner, Gavin Leech, Anna B. Stephenson, Tomáš Gavenčiak, Jan Kulveit, Yee Whye Teh, Leonid Chindelevitch, and Yarin Gal. 2020. "How Robust are the Estimated Effects of Nonpharmaceutical Interventions against COVID-19?" *arXiv:2007.13454 [cs, q-bio, stat]* (December). Accessed January 29, 2021. http://arxiv.org/abs/2007.13454.

Soltesz, Kristian, Fredrik Gustafsson, Toomas Timpka, Joakim Jaldén, Carl Jidling, Albin Heimerson, Thomas B. Schön, et al. 2020. "The effect of interventions on COVID-19." *Nature* 588, no. 7839 (December): E26–E28. https://doi.org/10.1038/s41586-020-3025-y.

Suryanarayanan, Parthasarathy, Ching-Huei Tsou, Ananya Poddar, Diwakar Mahajan, Bharath Dandala, Piyush Madan, Anshul Agrawal, et al. 2021. "AI-assisted tracking of worldwide non-pharmaceutical interventions for COVID-19." *Scientific Data* 8, no. 1 (March): 94. ISSN: 2052-4463. https://doi.org/10.1038/s41597-021-00878-y. https://doi.org/10.1038/s41597-021-00878-y.

UNDP and UN Women COVID-19 Global Gender Response Tracker. 2021. Dataset. United Nations Development Programme. https://data.undp.org/wp-content/uploads/2021/03/COVID-19-Global-Tracker-Methodological-Note-v_March2021.pdf.

Zheng, Qulu, Forrest K. Jones, Sarah V. Leavitt, Lawson Ung, Alain B. Labrique, David H. Peters, Elizabeth C. Lee, and Andrew S. Azman. 2020. "HIT-COVID, a global database tracking public health interventions to COVID-19" [in en]. *Scientific Data* 7, no. 1 (August): 286. ISSN: 2052-4463, accessed April 19, 2021. https://doi.org/10.1038/s41597-020-00610-2. https://www.nature.com/articles/s41597-020-00610-2.