Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Business Cycle and Health Dynamics during the COVID-19 Pandemic: A Scandinavian Perspective

---

CAMA Working Paper 19/2024
March 2024

Hilde C. Bjørnland
BI Norwegian Business School
Centre for Applied Macroeconomic Analysis, ANU

Malin C. Jensen
BI Norwegian Business School

Leif Anders Thorsrud
BI Norwegian Business School

Abstract

We use a unique measure of daily economic activity and manually audited non-pharmaceutical intervention indexes for Norway and Sweden to model the dynamics between COVID-19, policy, health, and business cycles within a SVAR framework. Our analysis documents large measurement errors in commonly used containment policy measures, significant endogeneity between the model's variables, and a strong health-economy trade-off following both policy shocks and precautionary actions. We further document that a large share of the variation in containment policies is driven by news innovations and quantify via counterfactual simulations the output cost per life saved from following stricter versus softer policies.

Keywords

COVID-19, simultaneity, expectations, business cycles

JEL Classification

E32, H51, I12, I15, I18

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

The Centre for Applied Macroeconomic Analysis in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Keywords

COVID-19, simultaneity, expectations, business cycles

JEL Classification

E32, H51, I12, I15, I18

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

1 Introduction

'There was a growing fear in the population about the new virus. Many responded by taking their kids out of kindergarden and school...We thought it was important to take control over the situation and establish some common rules...'
(Norwegian Prime Minister, Erna Solberg, NRK 26/05/2020)

'I think we have failed. We have a large number who have died and that is terrible...' (Swedish King Carl XVI Gustaf, BBC 17/12/2020)

When the COVID-19 pandemic emerged in the spring of 2020, policymakers and health authorities faced a dilemma: implement strict non-pharmaceutical interventions (NPIs) to curb the virus's spread but risk severe economic outcomes or opt for less invasive mitigation strategies with larger health risks but potentially less severe economic consequences. Countries around the world, with widely different institutional, cultural, and socioeconomic characteristics, chose differently. So did the neighboring countries in Scandinavia. While Norway and Denmark implemented strict NPI policies in mid March, such as lockdowns, travel restrictions, and social distancing laws, the initial policy response in Sweden was more lax. Indeed, Sweden's initial strategy relied more on voluntary guidelines and precautionary behaviour to curb the spread of the virus. Schools, businesses, and public spaces remained to a large extent open, and the government focused on recommendations for social distancing and hygiene rather than imposing strict restrictions on movement. As a result, it is commonly believed, the mortality rates in Sweden increased sharply relative to in Norway and Denmark. And, as the economies in these two countries grew by roughly one to two percent in 2020, it became easy to deem the Swedish response a failure. But, is this true?

Despite extensive research efforts, there is surprisingly little consensus regarding the joint health and economic effects of NPI policies, the potential independent role of precautionary behavior, and the intricate dynamics between these factors during the COVID-19 pandemic. Various studies report contradictory findings and conclusions. For instance, initial studies using purely epidemiological models, such as the influential study by Ferguson et al. (2020), predicted substantial reductions in COVID-19 mortality through strict NPIs. Subsequent research, incorporating the influence of endogenous and voluntary precautionary behavior among the affected population, as seen in the meta-analysis by Herby et al. (2022), has instead suggested that individual actions may have a greater impact than publicly enforced NPI policies on reducing mortality. Similarly, studies on the economic effects of NPIs have documented conflicting results. Some find, like Deb et al. (2020), that strict NPI policies have had a significant negative impact on economic activity. In contrast, Sheridan et al. (2020) document that a substantial economic contraction would have occurred irrespective of social distancing laws, highlighting the complex relationship between NPIs and economic outcomes.[1]

In this article we argue that at least three aspects of the pandemic, and the associated research, have contributed to the contrasting evidence. We then propose an empirical model that speaks to these aspects, fit it to Norwegian data, and perform counterfactual simulations to cast light on different NPI strategies and the Norwegian versus Swedish case in particular. As a first argument, we claim that to properly identify causal effects of, e.g., NPI policies on economic and health variables during the pandemic, one needs to take a general equilibrium perspective where policy, behavioral, and economic changes potentially affect each other simultaneously.

A second related argument speaks to the speed of events during the pandemic, which in turn dictates the usage of high-frequency data. Figure 1 provides an illustration. On March 12. 2020 the Norwegian government enforced a national lock-down. However, already five days earlier mobility, a often used proxy for measuring behavioral change, began to fall. Close to half of the fall in mobility observed in this period happened prior to March 12. Moreover, already in late February the business cycle contracted, and this contraction was amplified in the days after the lock-down. Clearly, the timing of many of these changes would have been impossible to detect using lower frequency data, and identifying the independent role of NPI policies relative to say, behavioral changes and news happening independent of policy, would be difficult. And, although this is a fairly uncontroversial claim, very few empirical studies have addressed it, likely due to the scarcity of high-frequency macroeconomic data.

A third issue is that of data quality. Measurement discrepancies (across counties) and under-reporting of important health outcomes, are well known. Additionally, widely used NPI indexes, such as the Oxford COVID-19 Government Response Tracker (OxCGRT), are hand-coded and may be subject to errors and biases. If such measurement issues are severe, they can lead to biased conclusions drawn from their use. Indeed, as we document later, there seems to be important biases in the original containment indexes from OxCGRT for both Norway and Sweden.

To address these issues we first perform three independent manual audits to construct revised containment indexes for Norway and Sweden. Next, we use the Structural Vector Autoregressive (SVAR) model proposed by Korobilis (2022) to estimate Norwegian general equilibrium dynamics at a daily frequency during the period early 2020 to mid 2021, i.e., before a large share of the population became vaccinated. In the model the revised containment policy index and a large set of health, mobility, and economic variables are treated as endogenous to each other. A horseshoe prior (Carvalho et al., 2010) is used to

[1] See, e.g., Brodeur et al. (2021) and Padhan and Prabheesh (2021) for broader literature reviews.

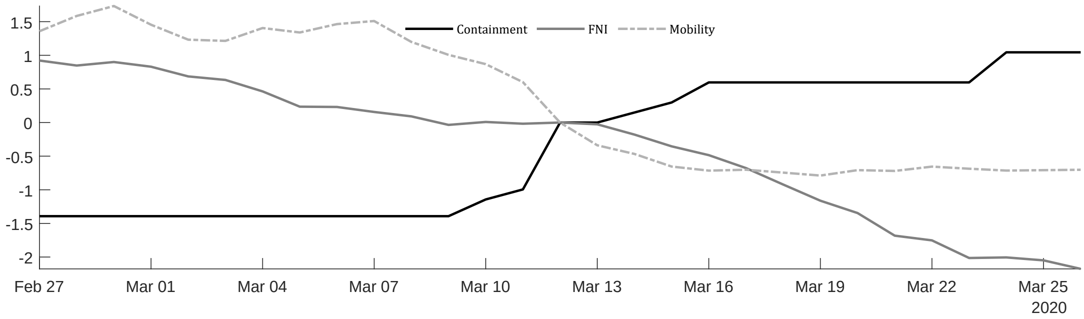

Figure 1. Timing. The graph reports the FNI, a daily business cycle measure for Norway (see Section 2.1), a measure of the stringency of NPI policies (see Section 2.2), and public transit mobility from Apple (see Section 2.3). All series are normalized to zero on March 12. 2020, when a national lock-down policy was enforced. In Norway, the first COVID-19 infected was reported on February 26. 2020.

allow for a rich set of lags, e.g., capturing the epidemiological relationship between being infected and potential death, and enforce sparsity on the system. Similarly, we assume that the variation in the underlying data can be approximated by four main shocks during the first three waves of the pandemic, and use sign restrictions to identify surprises in infection rates, more restrictive (or lax) containment measures, and two precautionary actions. The first is related to anticipatory precautionary actions, driven by, say, news about infection rates and economic sentiment abroad, capturing, e.g., mobility declines ahead of an observed rise in domestic infection rates. The second reflects instantaneous precautionary behavior, due to, say, businesses and employers quickly shifting to remote work arrangements or avoidance of public transport due to fear of crowded places leading mobility and infection rates to move in the same direction contemporaneously. This reduced rank treatment of the stochastic elements of the model not only facilitates identification, but also incorporates as a model feature measurement errors in the underlying data.

Estimation and identification of high-frequency macroeconomic effects is facilitated by using the daily business cycle indicator proposed in Thorsrud (2018). This measure is derived using a large corpus of business news from Norway's main news outlets, and have been documented to historically track the Norwegian business cycle very well. In line with this, we show that its correlations with quarterly GDP growth is highly significant, whereas other often used high-frequency activity proxies, such as electricity consumption and pollution, fit Norwegian GDP growth poorly.

With this data and model at hand, we conduct a number of experiments. First, we analyze impulse response functions and variance decompositions. In general, our results suggest significant endogeneity between the model's variables. More specifically, we find that sudden changes in containment policies reduce the severity of the pandemic and explain a large share of the variation in health variables, such as the number of newly infected, hospitalizations and mortality. However, non-mandatory precautionary actions are far from unimportant, and also help to gradually curb the pandemic. While this has been documented in numerous other studies as well, a more novel finding is that there are relatively large differences in how the business cycle and mobility statistics respond to the shocks in the model, suggesting that using the latter as a proxy for economic activity might be unwarranted, at least in the case of Norway. Finally, our results clearly indicate that containment policies are forward-looking and also respond to anticipated precationary shocks (i.e., news), a fact that is often over-looked in other studies.

Second, we use the strong correlation between quarterly GDP growth and the daily business cycle measure to conduct a simple cost-benefit calculation. Using official guidelines for quantifying the value of a statistical life we find that a containment policy shock that reduces both GDP and mortality ends up of having a net negative monetary value. However, in terms of interpretation, it is important to understand that this finding only applies to the effect of unexpected NPI fluctuations, and does not imply that systematic NPI policies in general have a net negative effect.

Third, to take into account the effect of systematic NPI policies we run a series of counterfactual experiments asking how the early phases of the pandemic would have evolved without such restrictive policies, and how it would have evolved in Norway if Norway had adopted the less restrictive Swedish policies. To run these experiments we use the SVAR output and estimate counterfactual time-varying reproduction numbers. These are then fed into standard SIR compartment model simulations. The results from this analysis show that when we also turn off the systematic containment policy response completely, we get adverse health outcomes that cannot plausibly be compensated by better economic outcomes. We also document that the typical "do nothing" alternative used to motivate policy interventions in the beginning of the pandemic most likely grossly over-estimates the effect this strategy would have had on the pandemic curve because it overlooks the effects of precautionary actions. At the same time, we show that the effect of systematic relative to unexpected NPI policies are stronger for health outcomes than economic outcomes. And, as a result, we find that the output cost per life saved in Norway from following the Norwegian strategy relative to the Swedish one is around 10 million NOK.

To the best of our knowledge, this is the first study to estimate a system where the dynamics of health and economic outcomes are modeled jointly using daily data. Early influential papers in the economic literature used theoretical models to cast light on the important endogeneity issue (Eichenbaum et al., 2021; McKibbin and Fernando, 2021; Alvarez et al., 2020; Berger et al., 2020; Jones et al., 2020), but empirical estimates have been lacking.

The studies by Camehl and Rieth (2023) and Arias et al. (2023) are perhaps the most closely related to ours. Camehl and Rieth (2023) take a global perspective and estimate a Bayesian Panel SVAR with data for 44 countries, while Arias et al. (2023) use a SVAR to assess the causal impact of lock-downs on health and macroeconomic outcomes in Belgium. Since modeling the joint dynamics between health and macroeconomic developments during the COVID-19 pandemic is challenging, we dedicate a separate section to discuss more extensively how our results are well in line with these two studies along some dimensions, but different along others.

Our study also speaks to the more epidemiological oriented literature and studies using compartment models and counterfactual modeling. In particular, we relate to studies such as Mishra et al. (2021) who use compartment models and counterfactual modeling to quantify how infection rates and mortality would have evolved in, e.g., Sweden and Denmark, if each country had adopted the policy of the others' policy. However, in contrast to these type of studies, we are able to decompose the time-varying reproduction number into contributions stemming from either NPI policies or behavioral actions. Although the former dominates, the latter is far from unimportant.

A separate contribution of our work is the manual audit and revised containment indexes for Norway and Sweden. For other research projects using these types of data, this might be of independent interest. Relatedly, this study contributes to studies that have focused on COVID-19 and its distinct health and economic consequences in Scandinavia. See, e.g., Andersen et al. (2022) and the references therein for an overview.

The rest of this article is organized as follows. In Section 2 we describe the data, and in particular the FNI and the revised containment indexes. In Sections 3 and 4 we describe the empirical model and present its main results. Section 5 describes the counterfactual simulations while Section 6 presents robustness experiments alongside a general discussion of our main findings. Section 7 concludes.

2 Measurement and data

Key to our analysis is the usage of a high frequency business cycle indicator for Norway and quantitative measures of governmental NPI policies for Norway and Sweden. Below we shortly describe the business cycle measure we use and document that it provides a better description of aggregate economic activity in Norway than other high-frequency economic activity approximations sometimes used in the COVID-literature. We then illustrate that the widely used OxCGRT containment indexes for Norway and Sweden are at odds with conventional wisdom, and propose an alternative coding scheme which provides more reasonable containment indexes for both of these countries. Finally, we

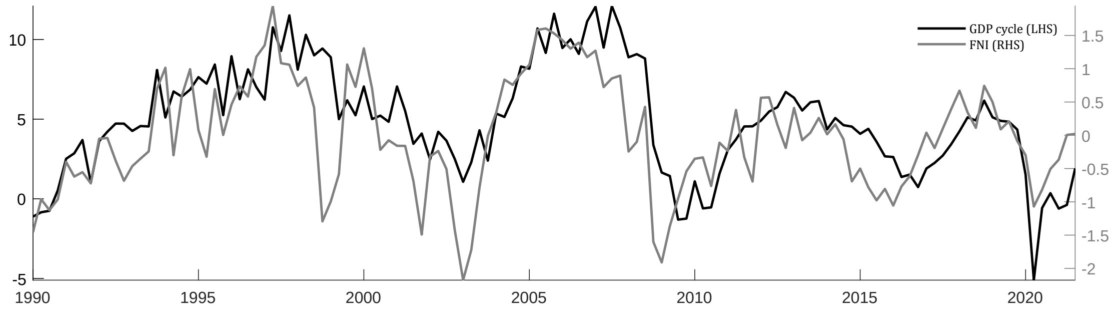

Figure 2. FNI and GDP cycle. The GDP cycle is computed using the eight quarter difference in GDP, i.e., $GDP_t - GDP_{t-8}$, which in practice is very similar to the trend-cycle filter suggested by Hamilton (2018).

present the remaining data series used in the analysis.

2.1 A daily business cycle measure

The daily business cycle index is labeled FNI and builds on Thorsrud (2018). It is sourced from https://www.retriever.no/fni/, which provides an updated index approximately one time per month. The index is constructed by first decomposing Norwegian business news, sampled from Norway's major news outlets, into topics using a Latent Dirichlet Allocation (LDA) model. Next, time series representing the prevalence of each topic is constructed and included, together with quarterly GDP growth, in a mixed-frequency time-varying Dynamic Factor Model, where the common factor is the FNI.

Thorsrud (2018) shows that the derived index is more timely and accurate than other commonly used business cycle indicators in Norway, while Larsen and Thorsrud (2019) use the same methodology together with international business cycle news to show that the model also performs well in the European and U.S. context. In Figure 2 the FNI is aggregated to quarterly frequency and displayed together with the Norwegian business cycle. As the graph clearly shows, there is a strong correlation between the FNI and the Norwegian business cycle.

In lack of good high-frequency variables capturing the state of the aggregate economy, the COVID-literature has often used alternative variables as proxies. Column one to four in Table C.1, in Appendix C, reports parameter estimates from regressing quarterly changes in GDP on the popular alternatives electricity consumption and emissions (Deb et al. (2021), Camehl and Rieth (2023)). The historical availability of these alternative series differ, and the number of observations in each regression is therefore different. Still, we find little evidence in support of using electricity consumption and emissions as good proxies for GDP growth in Norway. In column five to eight we redo the same type of regressions, but now use business cycle frequency changes of GDP as dependent variable. In this specification we find some support for both electricity consumption and emissions, but their marginal contribution relative to the FNI is minor.

Thus, while we do not deny that there might be a relationship between electricity consumption and emissions and GDP growth, the evidence for Norway seems week. In any case, as Table C.1 and Figure 2 documents, the FNI index approximates aggregate economic activity well both before and during the pandemic and is superior to the two alternatives considered here.[2]

2.2 Revised NPI indexes

To measure the effects of NPI policies on health and economic outcomes during COVID-19, the most widely used data sets is the Oxford Covid-19 Government Response Tracker (OxCGRT). OxCGRT is constructed and updated by a team of researchers and volunteers at the University of Oxford (Hale et al., 2021), and contains an impressive and systematic set of 23 daily response indicators, for more than 180 countries, starting January 1. 2020. Here we use the containment and health index (CON), which is an aggregated measure of 14 indicators encompassing containment and closure policies, such as school closures and travel restrictions, and health system policies, such as facial coverings or testing policies.

However, despite a well documented code-book the OxCGRT indicators are hand-coded by researchers and volunteers, and errors and revisions can happen. The broken lines in Figure 3 report the CON indexes for Norway and Sweden as provided by OxCGRT in their dataset released in August 2022. Although it is commonly believed that Sweden had much less stringent containment policies than, e.g., Norway, in the beginning of the pandemic, this is not the impression displayed in the figure. This impression also stands in stark contrast to a large body of research conducted in the earlier phase of the pandemic, concluding that Sweden followed a more lax policy than its Scandinavian neighbors (Yarmol-Matusiak et al., 2021; Juranek and Zoutman, 2020; Esaiasson et al., 2021; Andersson and Aylott, 2020).

Motivated by these facts and the importance of having good quantitative measures of NPI policies, we have manually re-constructed the CON indexes for both Norway and Sweden. To do so we have followed the OxCGRT methodology and critically re-evaluated every policy change for each of the 14 indicators. We have also supplemented our analysis with national data (from press-releases and the media) on policy changes and implementations to ensure that relevant events and policies are not overlooked. A detailed description of how we have done this, together with a short overview of the

[2] The work by Sheridan et al. (2020) illustrates a new line of research using credit card transaction data for high-frequency macroeconomic monitoring. However, this type of data mainly reflect (a fraction) of household consumption developments, and not the aggregate business cycle, and is, at least in Norway, only available on a weekly frequency prior to 2019.

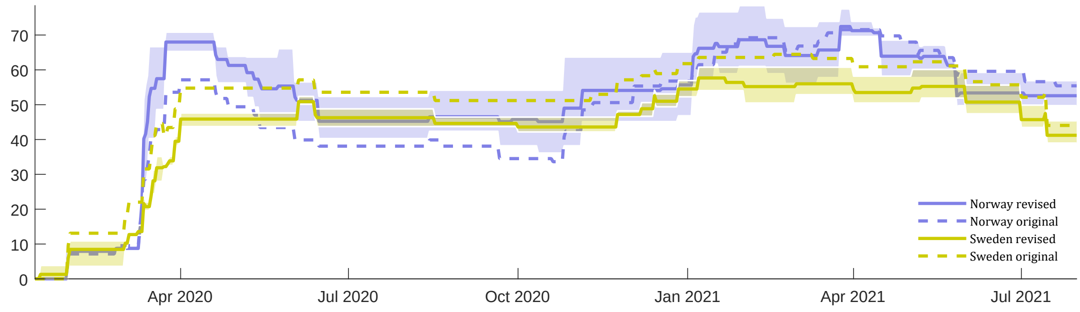

Figure 3. Stringency of NPI policies. The graph reports the original OxCGRT Containment and Health Index and our revised numbers. The shaded areas represent the maximum and minimum value assigned by us or two external coders at any time point.

OxCGRT code-book, is provide in Appendix A.1 and A.2.

To further ensure that the revisions we potentially impose are robust, we have also asked two external coders to independently perform the same manual re-construction. The external coders have been given access to the same underlying code-book and data as we used when constructing the revised indexes, and have then been asked to re-evaluate every observation. While doing this they were not given any information about our own revised numbers or the underlying issue with the original indexes. A more detailed description of the manual audit is given in Appendix A.3.

The solid lines in Figure 3 report the average of the revised CON indexes for Norway and Sweden. The shaded areas represent the maximum and minimum value assigned by us or the two external coders at any time point. As seen in the figure, there is a large degree of agreement on NPI events in the early phase of the pandemic for both Norway and Sweden. This continue to hold for the revised Swedish numbers throughout the sample, but for Norway the numbers vary to a larger degree after April/May 2020.

Despite some uncertainty, however, both indexes have clearly changed following our revised interpretation of the underlying data. And, the revised indexes clearly illustrate that the initial NPI response was softer in Sweden than in Norway. The level of the containment indexes during the summer and autumn of 2020 have also been revised up for Norway and down for Sweden, making the two series very similar during this period. For the reminding part of the sample the differences between the revised and original numbers are smaller, although the level of the revised Swedish index is somewhat lower than in the original series. Thus, the largest differences between the original and revised series are associated with the spring events in 2020. In sum, going forward, we use the average of the revised containment indexes.[3]

[3]The revised indexes can be downloaded from https://www.bjornland.no/filer/dokumentarkiv/21122023_oxford_containment_rev_webdec2023.xlsx.

2.3 Data and transformations

In addition to the indexes described above, we include a number of daily health, mobility and financial variables in the analysis. Below we provide a short description of the variables and data transformations.

The health variables we consider are the daily numbers of new confirmed infected (NEWINF), new hospitalizations (NEWHOSP), new confirmed deaths (MORT), and the share of positive COVID-19 tests relative to the total number of tests (POS). All series are important for understanding the evolution of the pandemic and its severity. While the NEWHOSP and MORT statistics are relatively accurately measured in Norway, the NEWINF and POS series are subject to potential measurement errors. However, as discussed in Section 6, the degree of under-reporting is likely smaller in Norway than in many other countries. All series are sourced from the FHI.

The economic variables we include are foremost related to financial markets, and are the short-term interest rate (NIBOR), the spread between 10-year government bonds and the short-rate (SPREAD), the stock market index (OSEBX), the Dollar/NOK exchange rate (EXCH), and implied realized volatility (VOL) computed using a 5-day moving average of squared returns on the OSEBX index. While these variables are far from perfectly correlated with the aggregate business cycle, they all capture important forward-looking information embedded in current prices. To capture fluctuations potentially more directly associated with the state of the macroeconomy, we also use daily unemployment benefit allowances collected from the Norwegian Labour and Welfare Adminisitration (NAV), an indicator of the governmental economic relief packages (G), and the price of oil (OIL). For a large petroleum exporter such as Norway, the price of oil is an exogenous but highly important determinant of growth (Bjørnland and Thorsrud, 2016). The G variable is constructed using the timeline of news and press releases from the Norwegian Ministries during the pandemic and cross-checked against other similar indexes. Appendix B.1 provides a more detailed description of this variable.

We measure mobility patterns using domestic and international flight data (FLIGHT) from Avinor and mobility statistics based on cellphone data provided by Apple. The flight data captures the cumulative number of domestic and international flights per day from all Avinor-operated airports. Apple's mobility data provides daily relative volume of direction requests per geographical entity compared to a baseline volume on January 13. 2020, and covers public transit (TRANSIT), driving (DRIVING), and walking (WALKING). While the data only represents Apple's cellphone users, it has been cross-checked against some of Google's mobility measures and found to be highly correlated. However, the Apple data is available for a longer time span for Norway and therefore used here.

Prior to estimation we difference all data, except those variables that are already in difference and CON and G, and remove seasonality. The former transformation is done to ensure that we estimate a stationary model. We have experimented with estimating the model, described in Section 3, using all variables in levels. Doing so we experience that it is difficult to obtain draws from a stationary posterior. In terms of seasonality, many of the raw series we use display clear day-of-the-week effects, which magnitude vary across the sample. For example, the number of confirmed infected is typically much higher following weekends than on other days of the week. For this reason, we apply a simple seasonal filter on this type of data prior to estimation. Here, any remaining drifts are removed from the differenced series using a symmetric moving average filter. Then, seasonal indexes are created for each day of the week, and a 7-term symmetric moving average is applied to data within each of these indexes. Finally, the seasonal component is centered around zero and subtracted from the original series.

CON and G are treated differently. Both variables are clearly trending, with irregular shifts. Differentiating these variables creates something close to indicator variables without much persistence and time series dynamics. In our benchmark specification we therefore model these variables as deviations from trend using the filter suggested by Hamilton (2018), which simply involves regressing the variable of interest at date $t + h$ on the four most recent values as of date t and using the residuals from this regression as the cyclical estimate. As documented in Hamilton (2018) this filter accommodates a large range of underlying non-stationary data generating processes and provide an essentially assumption-free summary of the data. Here we use $h = 28$ to capture the fact that new policy actions typically was implemented under the expectation that they needed to last 2-4 weeks before one could observe changes in health outcomes (partly due to the epidemiological characteristics of the virus).

In the interest of preserving space, the raw and transformed data for are presented in Figures C.1 and C.2 in Appendix C.

3 A dynamic macro-health model

To analyze the joint health and economic dynamics during the pandemic we include the variables described in Section 2 in a SVAR. With $n = 18$ variables, this is a large system. The size helps facilitate identification and interpretation, but is also a challenge for estimation. Here we build on the approach proposed in Korobilis (2022), which identifies common shocks and adapts a hierarchical horseshoe prior to handle the large number of parameters. Section 3.1 describes the model in greater detail. Section 3.2 describes the identification assumptions, while Section 3.3 outlays the estimation algorithm.

3.1 The model

Following Korobilis (2022), the reduced-form vector autoregressive (VAR) model we consider can be written as:

$$\boldsymbol{y}_t = \boldsymbol{\Phi} \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \tag{1}$$

with

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{v}_t \tag{2}$$

where $\boldsymbol{y}_t$ is a $(n \times 1)$ vector of observed variables, $\boldsymbol{x}_t = (1, \boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})'$ a $(k \times 1)$ vector (with $k = np + 1$) containing a constant and p lags of $\boldsymbol{y}$, $\boldsymbol{\Phi}$ is a $(n \times k)$ matrix of coefficients, and $\boldsymbol{\varepsilon}_t$ a $(n \times 1)$ vector of disturbances decomposed using the factor model specification in (2). Here $\boldsymbol{\Lambda}$ is a $(n \times r)$ matrix of factor loadings, $\boldsymbol{f}_t$ is a $(r \times 1)$ vector of common factors, and $\boldsymbol{v}_t$ is a $(n \times 1)$ vector of idiosyncratic shocks. We further assume that $\boldsymbol{v}_t \sim i.i.d.N(0_{n\times1}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ being diagonal, $\boldsymbol{f}_t \sim N(0_{r\times1}, \boldsymbol{I}_r)$, and

$$cov(\boldsymbol{\varepsilon}_t | \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) = \boldsymbol{\Omega} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma} \tag{3}$$

The reduced-rank SVAR representation of this model is obtained by combining (1) and (2), and multiplying the reduced-from VAR with the generalized inverse of $\boldsymbol{\Lambda}$. This gives

$$\boldsymbol{A}_1 \boldsymbol{y}_t \approx \boldsymbol{B}_1 \boldsymbol{x}_t + \boldsymbol{f}_t \tag{4}$$

where the matrix $\boldsymbol{A}_1$ equals the generalized inverse of $\boldsymbol{\Lambda}$, and $\boldsymbol{v}_t$ is treated as a residual or noise shock with no structural interpretation.[4]

In the current context, the reduced-rank SVAR specification has two particularly attractive features. First, it allows us to project n shocks into r structural shocks. As n needs to be relatively large to capture the joint dynamics between the endogenous variables, and $r < n$, this dimensionality reduction helps in terms of specifying the identification restrictions as well as estimating the system. Second, the model incorporates as a model feature potential measurement errors in the underlying data through the idiosyncratic error term $\boldsymbol{v}_t$.

Assuming invertability, and using the vector moving average representation of the VAR, it is easy to show that the structural impulse response function on impact is

$$\frac{\partial \boldsymbol{y}_t}{\partial \boldsymbol{f}_t} = \boldsymbol{\Lambda} \tag{5}$$

Thus, parametric restrictions on $\boldsymbol{\Lambda}$ correspond to structural restrictions on the impact impulse response functions, as in conventionally used SVAR models. These restrictions are discussed in Section 3.2.

[4]To see this, notice that $(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}\boldsymbol{y}_t = (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{x}_t + \boldsymbol{f}_t + (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}\boldsymbol{v}_t$. Letting $\boldsymbol{A}_1 = (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}$, we get $\boldsymbol{A}_1\boldsymbol{y}_t = \boldsymbol{B}_1\boldsymbol{x}_t + \boldsymbol{f}_t + (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}\boldsymbol{v}_t$. Accordingly, as $n \to \infty$ the residual noise interpretation is justified by $(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}\boldsymbol{v}_t \to 0$ due to the Central Limit Theorem in Bai (2003).

As in all factor models there are also identification issues related to the common component in (2). First, as $\boldsymbol{\Omega}$ contains $n(n+1)/2$ free elements, while the right-hand side of (2) has $nr+n$ free parameters, the number of structural shocks is constrained by $r \leq (n-1)/2$. Second, an additional $r(r-1)/2$ restrictions are needed in order to deal with the well known rotation problem. However, the assumption that $\boldsymbol{f}_t \sim N(0_{r\times 1}, \boldsymbol{I}_r)$ together with the restrictions on $\boldsymbol{\Lambda}$ ensures both unique estimation of the factors and identification of the structural shocks.

3.2 Identification

As alluded to in the introduction, we postulate that the unpredictable fluctuations in the data during the period we analyze can be captured by four common shocks: an infection shock capturing unexpected changes in underlying infection pressure, due to, e.g., changes in infectivity or mutations; a containment shock capturing whether NPI actions were stronger or weaker than expected; and two precautionary actions: The first is an anticipatory precautionary shock, indicating that agents in the economy act upon unexpected information received today about future economic prospects or Covid-related developments. This could, for example, be purely economic related (international) news, but also negative information about the evolution of the pandemic.[5] The second precautionary shock reflects other developments, due to, say, instantaneous adherence to remote work arrangements or avoidance of public transport due to fear of crowded places. These shocks are orthogonal to the anticipatory shocks, leading mobility and infection rates to move in the same direction contemporaneously. For notational simplicity, we will label the two precautionary shocks as simply news and mobility shocks.

Following Korobilis (2022), we identify these shocks by putting sign-restrictions on $\boldsymbol{\Lambda}$.[6] Table 1 reports the restrictions, which all apply to the contemporaneous effect of the different shocks. The variables we entertain naturally groups into five distinct blocks. We label these block of variables as *Fast moving health*, *Policy*, *Economy*, *Mobility*, and *Slow moving*, and enforce (more or less) the same restrictions for variables within each group

[5] Identification of this type of shock can suffer from the so called non-fundamentalness problem, see, e.g., Blanchard et al. (2013). However, as argued in studies such as Sims and Zha (2006), Forni et al. (2017), and Beaudry and Portier (2014), this potential problem can be alleviated by adding forward-looking variables in the information set considered, which is also what we do in our model.

[6] Sign-restrictions are increasingly used to identify shocks in Vector Autoregressive models. Other popular sing-restriction algorithms include those proposed by Rubio-Ramirez et al. (2009) and Arias et al. (2018). Korobilis (2022) describes why putting restrictions directly on $\boldsymbol{\Lambda}$ have a number of benefits related to inference and efficiency. For example, while the traditional approaches yield set identification, the model and algorithm proposed by Korobilis (2022) puts restrictions directly on the impact matrix, allowing the researcher to use standard techniques to estimate the posterior distributions.

Table 1. Sign restrictions. A + and a − indicate a positive and negative restriction, respectively. Zero restrictions are denoted by a 0, while unrestricted coefficients are denoted with a *na*.

		Shocks			
		Covid-19	Policy	Precautionary	
	Variables	Infection	Containment	News (Anticipatory)	Mobility (Instantaneous)
Fast moving health	New Covid-19 cases	+	-	0	-
	Share positive	+	-	0	-
Policy	Containment Health index	0	+	0	0
	Economic support index	0	na	0	0
Economy	FNI	-	-	-	na
	Interest rate	-	-	-	na
	Stock market index	-	-	-	na
	Spread	-	-	-	na
	Exchange rate	+	+	+	na
	Realized volatility	+	na	+	na
	Oil price	0	0	na	0
Mobility	Transit mobility index	-	-	-	-
	Dom. and int. flights	-	-	-	-
	Walking mobility index	-	-	-	-
	Driving mobility index	-	-	-	-
Slow moving	New hospitalizations	0	0	0	0
	New dead	0	0	0	0
	New unemp. benefit app.	0	0	0	0

to sharpen the identification of the factors, i.e., the shocks.

The infection shock is restricted to increase the fast-moving health variables on impact. Likewise, the (forward looking) domestic economic variables and mobility statistics respond to new information about infections, and move in the same negative direction. Note here that an increase in the exchange rate corresponds to a depreciation and we assume that uncertainty increases following a positive infection shock. Notice also that we use the fast-moving health variables to measure the underlying infection pressure in the economy. To the extent that these types of variables are riddled with measurement errors and delayed reporting we assume that this is well captured by the idiosyncratic errors in the model, and that shocks in the model relate to the underlying infection pressure.

The containment shock is identified such that it increases the stringency level. This alleviates the infection pressure, but has a negative effect on the domestic economy and mobility. Despite having these adverse consequences, we do not put any restrictions on the realized volatility response. I.e., we allow for the possibility that this shock can be looked upon as positive in terms of coordinating the response to the pandemic and thus reducing the uncertainty. Moreover, only the containment shock can affect the NPI variable contemporaneously. This is consistent with the fact that policy interventions need to be coordinated and planned before announced and implemented, hence the other shocks will affect the NPI with a lag. Related to this, we do not restrict the sign of the economic support index following a containment shock, but allow it to respond to capture the empirical regularity that containment measures often were announced together with economic relief packages.

An anticipatory precautionary (i.e., news) shock has a negative effect on the domestic economic variables, mobility, and increases the level of uncertainty. In line with the interpretation of the news shock described above, this is also the only shock in the model that is allowed to be contemporaneously correlated with the price of oil. To be able to fully isolate the effects of a news shock from a shock to infection rates and the instantaneous precautionary shock, we further assume that the fast-moving health variables are contemporaneously uncorrelated with the anticipatory shock. Thus, a prime example of this type of shock is in the beginning of the pandemic in Europe, when there was little, if any, recorded cases in Norway, but massive media coverage on outbreaks and adverse economic consequences abroad triggering behavioral changes at home.

Finally, an instantaneous precautionary (i.e., mobility) shock is identified as a shock that moves the underlying infection pressure and mobility statistics in the same direction without having a contemporaneous effect on the policy variables. Since such changes might, or might not, have adverse domestic economic consequences, we leave these responses unrestricted. In many ways, this shock can be thought upon as a residual precautionary shock capturing instantaneous changes in mobility that are unrelated to any shocks to the observed infection rates, to changes in government enforced containment measures, or to anticipatory precautionary behavior.

Together these assumptions uniquely identifies the four shocks. In the model we also include a set of auxiliary, or slow-moving, variables. These variables do not respond to any of the shocks on impact. Still, their dynamic response is of interest and also allows us to empirically validate the model. E.g., in a properly specified model the response of mortality to an infection shock should be substantially delayed since the epidemiological characteristics of the virus suggests an average time lag of roughly three weeks from infection to potential mortality (Linton et al. (2020), World Health Organization (2020)). Similarly, unemployment statistics are highly unlikely to respond contemporaneously to any of the shocks in the model, but can have dynamics that are informative for the overall VAR and therefore included in the model.

3.3 Specification and estimation

To estimate the model we consider the sample February 10. 2020 to June 30. 2021. This includes the time period immediately before infections started to rise in Norway and covers the three infection waves in the spring of 2020, the late autumn of 2020, and the spring of 2021.[7] We allow for a rich lag structure to capture the dynamic correlations and set $p = 21$. As the sample length is 510 days and $n = 18$ this results in a highly parameterized model, with $np + 2n = 414$ parameters in $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ plus the non-zero elements in $\boldsymbol{\Lambda}$, with relatively low degrees of freedom. To perform inference we therefore follow the Bayesian estimation routine suggested in Korobilis (2022), which adopts the hierarchical horseshoe prior of Carvalho et al. (2010) to penalize the likelihood and shrink the (autoregressive) coefficients towards zero.

The prior specification takes the form:

$$
\begin{aligned}
\boldsymbol{\phi}_i \equiv vec(\boldsymbol{\Phi}_i) \sim \; & N_k(0, \underline{\boldsymbol{V}}_i) \\
\boldsymbol{f}_t \sim \; & N_r(0, I) \\
\boldsymbol{\Lambda}_{ij} \sim \; & \begin{cases} N(0, \underline{h}_{ij}) I(\Lambda_{ij} > 0), & \text{if } S_{ij} = 1, \\ N(0, \underline{h}_{ij}) I(\Lambda_{ij} < 0), & \text{if } S_{ij} = -1, \\ \delta_0(\Lambda_{ij}), & \text{if } S_{ij} = 0, \\ N(0, \underline{h}_{ij}), & \text{otherwise} \end{cases} \\
\sigma_i^2 \sim \; & invGamma(\underline{\rho}_i, \underline{\kappa}_i)
\end{aligned} \tag{6}
$$

for $i = 1, \ldots, n$, $j = 1, \ldots, r$, where $\boldsymbol{\Phi}_i$ is the i^{th} row of $\boldsymbol{\Phi}$, σ_i^2 is the i^{th} diagonal element of $\boldsymbol{\Sigma}$, and $\delta_0(\Lambda_{ij})$ is the Dirac delta function for Λ_{ij} at zero. We set a diffuse prior for $\boldsymbol{\Lambda}_{ij}$, letting $\underline{h}_{ij} = 100$, and $\underline{\rho}_i = 1$ and $\underline{\kappa}_i = 0.01$ implying a prior mean of 0.16 and large variance in the diagonal elements of $\boldsymbol{\Sigma}$. Thus, our prior specification reflect the belief that the scale of the idiosyncratic variances should be small relative to the disturbances associated with the common component $\boldsymbol{\Lambda f}_t$. Finally, $\underline{\boldsymbol{V}}_i = \sigma_i^2 \tau_i^2 \psi_{i,j}^2$, represents the horseshoe prior, where $\psi_{i,j} \sim Cauchy^+(0, 1)$ and $\tau_i \sim Cauchy^+(0, 1)$. As such, compared to other popular shrinkage priors, such as, e.g., the Minnesota prior, it follows that the horseshoe prior is tuning free as the local and global shrinkage parameters have their own distribution and only updated by information in the data.

The posterior distribution of the parameters $p(\{\boldsymbol{\phi}_i\}_{i=1}^n, \{\boldsymbol{\Lambda}_i\}_{i=1}^n, \{\boldsymbol{f}_t\}_{t=1}^T, \{\boldsymbol{\sigma}_i\}_{i=1}^n | \boldsymbol{y})$ is estimated by sequentially sampling from the conditional posterior distributions. Korobilis (2022) provides a detailed description of the Gibbs sampler used. In all our results we run the sampler for 100000 iterations and thin the Gibbs chain by a factor of 10. Additional 20000 iterations are used as burn-in. All draws of the autoregressive parameteres are restricted to be within the stationary region of the model, and we initialize the sampler using OLS estimated parameter values and factor loadings based on a principal components decomposition of the implied reduced form residuals.

[7]In Norway the vaccination program started (slowly) in December 2020. By the end of the spring in 2021 only roughly 10 percent of the population were fully vaccinated.

4 Results

In the following we use the estimated model, and its output, to conduct a number of experiments. First, in Section 4.1 we present model evaluation statistics to show that the model fits the data reasonably well. Next, in Section 4.2 impulse responses and variance decompositions are presented to show the effect of the different shocks and how they transmit across economic and health related variables. Finally, in Section 4.3 we use the insights from this analysis, together with the mapping between GDP growth and the FNI, to conduct a simple cost-benefit analysis.

4.1 Evaluation

We evaluate the model fit along two dimensions. First, we use the Deviance Information Criterion (DIC) of Spiegelhalter et al. (2002) to evaluate whether the data is best represented by fewer or more common shocks than what we propose in our benchmark model. Second, we conduct a simple forecasting experiment to analyze to what extent the model is able to deliver reasonable predictions for hospitalizations. This variable is relatively precisely recorded in the data and was also a variable the health authorities actively tracked during the pandemic. Thus, we can compare the model forecasts for this variable to predictions made by the authorities during the period we study.

Table C.2, in Appendix C, shows that the data clearly favors a model with many rather than few common shocks. A model specification with five factors is preferred. For this specification we apply the same restrictions as in Table 1 on the first four factors, but let the fifth shock by unrestricted. I.e., it is not given any economic interpretation. Still, although this five-factor model is preferred on statistical grounds, we show in Section 6 that including a fifth factor does not qualitatively affect our main conclusions.

In terms of forecasting, the results reported in Figure 4 suggest that the proposed model does a reasonably good job at describing the evolution of the health outcomes in the model. The figure reports the predictions for new hospitalization during the first half of 2021 for three different forecasting horizons. Starting in the end of 2020, the estimated model is fed the observed data at the time and then used to predict new hospitalization 21 days into the future. Next, we add one week to the dataset, and update the predictions for the next three weeks. This process is repeated until the end of the summer in 2021.

To forecasting horizons and time period is chosen to make the predictions comparable to those produced by FHI during the pandemic. For comparison, these predictions are also illustrated in the figure. As clearly seen in the plots, the predictions produced by the VAR model used here are of similar quality, if not more accurate, than those produced by the health authorities. Note, however, that the FHI predictions were produced in

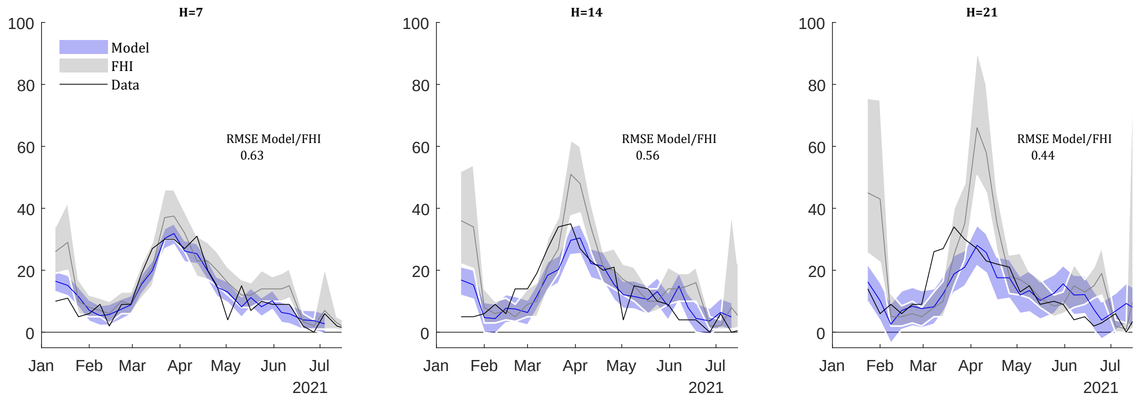

Figure 4. Predictive evaluation. The graphs report the predicted number of new hospitalizations (due to COVID-19) for three different forecasting horizons; one, two, and three weeks. The point estimates are reported together with 68 percent posterior probability bands. The alternative predictions were produced in real-time by the FHI.

real-time. The model-based predictions are constructed with the benefit of hindsight. Thus, the predictive results should be interpreted with care, and are more suggestive of a reasonable model fit than better or worse forecasting power per se.[8]

4.2 Impulse responses and variance decompositions

Although the model delivers a rich set of impulse responses, we focus in Figures 5 and 6 on eight key variables which illustrate the overall dynamics well; The number of infected, hospitalizations, mortality, the FNI, containment, mobility, the stock market, and the exchange rate.[9] The two-by-two quadrant in the left panel of each figure reports the responses of a given variable to the four different shocks. The panels to the right report the full distribution of the corresponding variance decompositions for two different horizons.

Four main observations stand out. First, although the impact responses are constrained by the restrictions we impose, their dynamics across response horizons are not. Still, the responses display plausible dynamics. E.g., an infection shock leads to a gradual increase in the number of infected, and as a response, a sharp decline in mobility.[10] A containment shock lowers mobility immediately, and as a consequence, infection rates gradually fall. Likewise, and consistent with the discussion in Section 3.2, since we allow the anticipatory precautionary (news) shock to be associated with negative information

[8] Doing a full out-of-sample forecasting experiment is not possible because truncation of the sample length results in a training sample with too few observations to obtain meaningful parameter estimates.

[9] Except for the CON index, the numbers reported are cumulative changes. Since the model is estimated on standardized data, all responses are re-scaled by the variable's original standard deviation. The mobility response is computed as the average response across the mobility statistics. The results for the remaining variables are reported in Figures C.3 and C.4 in Appendix C.

[10] The persistent negative effect on the level of mobility is due to the fact that we estimate the model in difference form and because mobility was below its average during most of the sample we study.

about the evolution of the pandemic it leads to both decreased economic activity and more severe health outcomes (i.e., increased hospitalizations). Moreover, the model captures well the epidemiological characteristics of the virus. That is, for all shocks in the model, it takes roughly one week before we observe significant changes in the number of hospitalizations and three weeks before the mortality statistic responds.

Second, there is a clear trade-off between implementing NPI policies which curb the pandemic and economic activity. A one standard deviation containment shock has a negative and significant effect on infection rates, hospitalizations and mortality, but also lowers the FNI significantly. However, the anticipatory precautionary (news) shock also works in the same direction, although with a lag: Negative news drives down economic activity and makes individuals more precautionary by reducing mobility. At first, infection rats and hospitalizations are pushed up, but gradually the decline in mobility helps to ease the negative health outcomes. In contrast, the instantaneous precautionary shock reduces mobility and helps curb the pandemic, but has an insignificant effect on the business cycle. However, with few exceptions, this is the least important shock in terms of the variance decomposition. Related to this, since mobility statistics and the business cycle indicator respond very differently to many of the shocks in the model, e.g., the infection shock, our results also provide a warning against using mobility statistics as a proxy for economic activity (as also alluded to in Section 2.1).[11]

Third, in terms of the relative importance of the shocks, it is clear that the economic variables such as stock prices and the exchange rate to a large degree are driven by the precautionary shocks (news and mobility), while for the FNI index, the containment shock is also very important. The mobility variables respond strongly to infection shocks, followed by the containment and the anticipatory precautionary shocks. For the health variables the picture is somewhat more nuanced. If we focus on the number of infected, the containment shock dominates. However, it is easy to argue that the hospitalization and mortality series are more accurate statistics, and for these variables both the infection and precautionary shocks contribute substantially. E.g., for the hospitalization statistics these shocks explain approximately 20 percent each of the short-run variance. Thus, although containment policies undoubtedly play a significant role for the health outcomes in our model, exogenous variation in people's behavior and infection surprises are far from

[11]Studies such as Aksoy et al. (2022) suggest that employees were favorably surprised by their workers' working-from-home productivity during the pandemic. For the Scandinavian countries, this effect is potentially even stronger because particularly high digitization made us better prepared to work from home (Andersen et al., 2022). Thus, a change in mobility might not be a reliable signal of changes in economic activity. In line with this, Gamtkitsulashvili and Plekhanov (2023) conclude, based on an empirical analysis covering 53 countries, that the correlation between economic activity and mobility deteriorated during the pandemic.

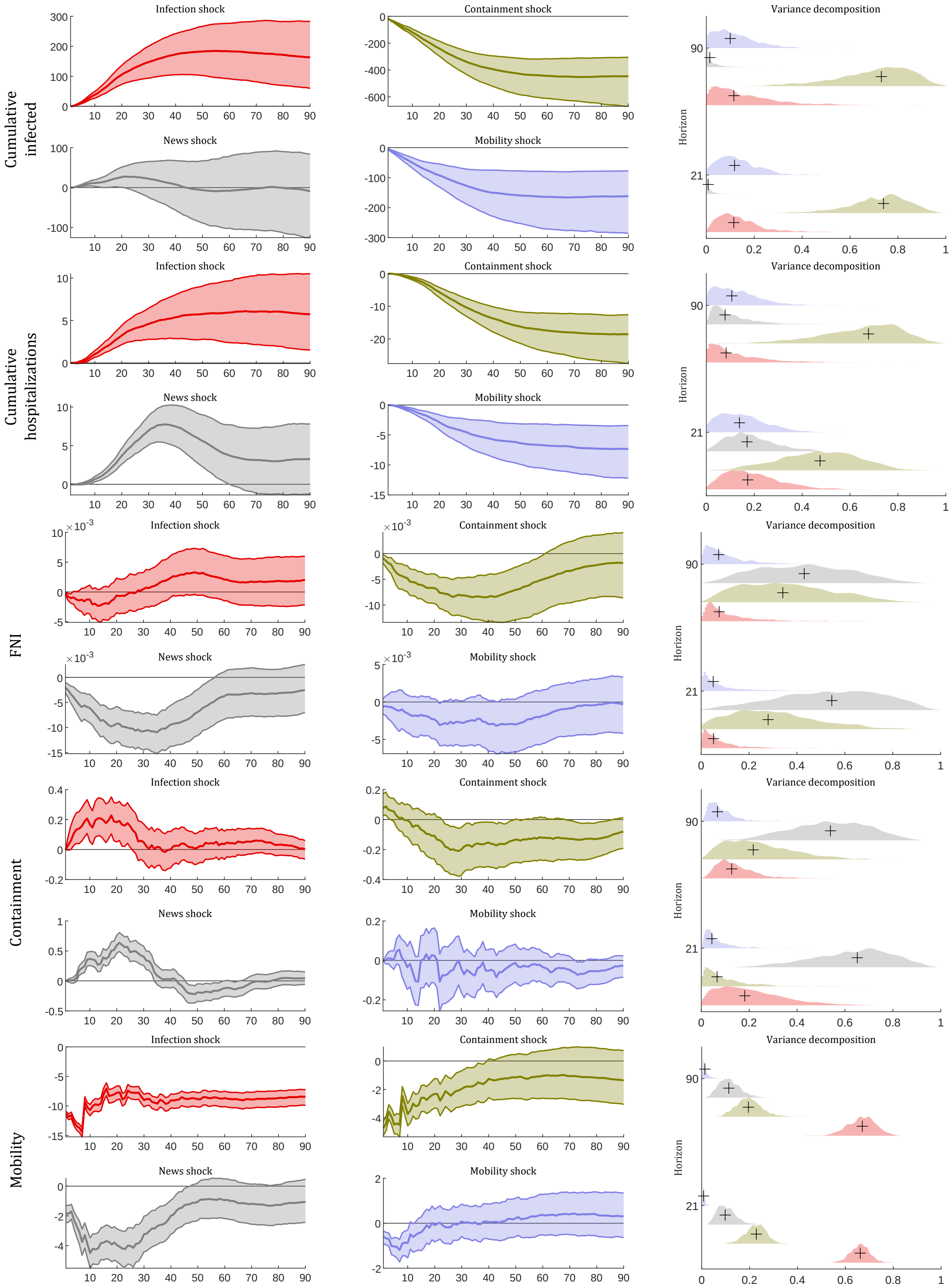

Figure 5. Impulse responses and variance decompositions. Median response to a one standard deviation shock. Shaded areas equal 68 percent posterior probability bands. The variance decompositions report the full distribution, together with the median (+), for two different horizons. For notational simplicity, the anticipatory and instantaneous precautionary shocks are labeled news and mobility, respectively.

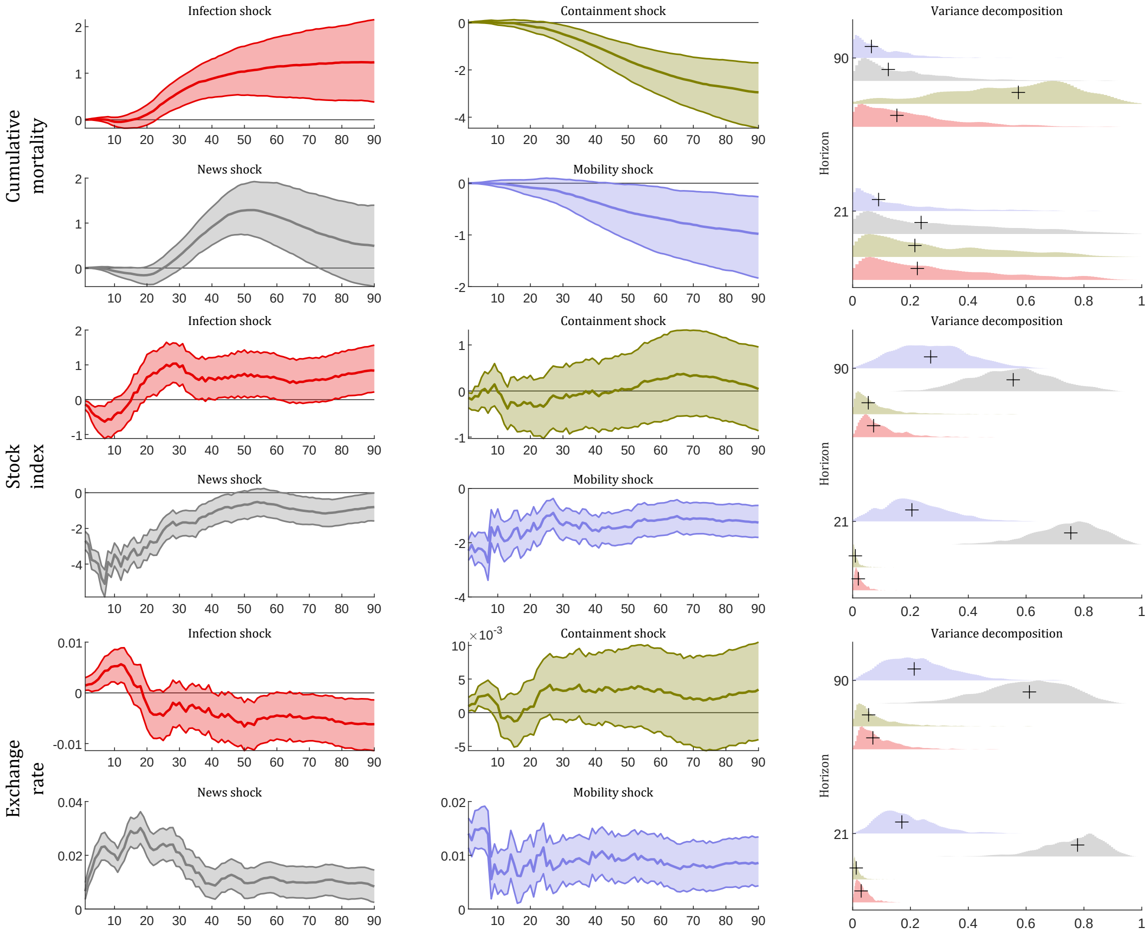

Figure 6. Impulse responses and variance decompositions. See Figure 5 for explanations.

unimportant.

Fourth, containment policies are highly endogenous and to a large degree driven by news about the future (i.e., anticipation). That is, there is a strong systematic response in containment policies following both infection and anticipated precautionary behavior shocks.

4.3 Cost-benefit calculations

At the center of the policy debate during the pandemic was the nexus between saving lives and avoiding adverse economic outcomes. Here we use the estimated impulse responses from the previous section together with the tight relationship between GDP growth and the FNI to provide a simple cost-benefit analysis of unexpected containment policies.

We assume a one standard deviation innovation to containment policies, and then compare the business cycle response to the cumulative number of new hospitalizations and mortality. To translate the health responses to monetary values, we use the assumptions advocated by the public authorities. The Value of a Statistical Life (VSL) and the cost of

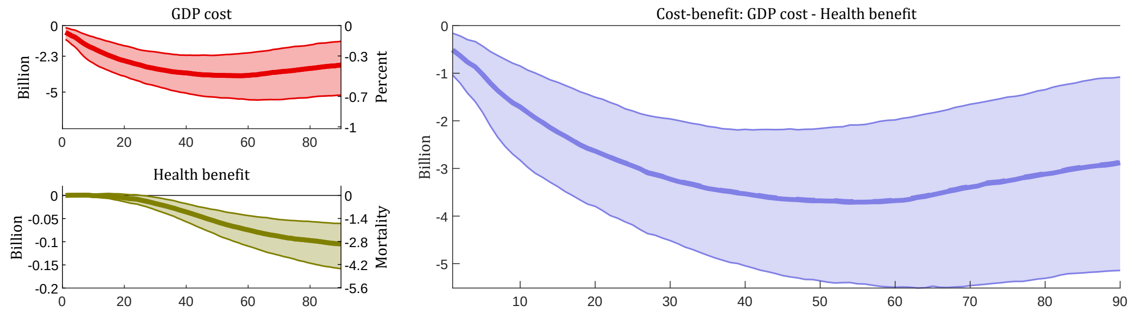

Figure 7. Cost-benefit analysis. Responses to a one standard deviation containment shock. The GDP response is computed using the mapping between GDP growth and the FNI response in Figure 5. The mortality response is translated to a monetary value by assuming a VSL of 35 million and a cost of hospitalization of roughly 1 million. See the text for further explanations.

one new hospitalization is set to 35.5 million NOK and 0.8 million NOK, respectively.[12] The cumulative responses for mortality and hospitalizations are multiplied by these values and added together. To translate the business cycle response to monetary values, we use the regression reported in column 4) in Table C.1, and then compute $y_h = \frac{\hat{\beta}}{h}\sum_1^h x_i$ for $h = 1, 2, \ldots, 90$, where x_i is the FNI response at horizon i and $\hat{\beta}$ is the estimated mapping between quarterly percentage GDP growth and the FNI. I.e., y_h gives the implied percentage change in GDP at horizon h.

Figure 7 reports the results. A one standard deviation containment policy shock reduces quarterly GDP by roughly 0.4 percent, equaling a loss of approximately 3 billion. In turn, the shock reduces mortality by three persons, which amounts to roughly 0.1 billion NOK. In sum, the net benefit is negative.

To put this finding into context Appendix B.4 provides more details about how the Norwegian authorities responded to the crisis and presents some of the policy strategies considered. As is evident from that discussion, the results presented above are well aligned with an interpretation where the containment shocks are seen as random fluctuations around the systematic policy function actually followed during the early phases of the pandemic. However, what's not well captured by this type of analysis is the effect of following one strategy versus another, i.e., the role of the systematic policy response function. We turn to this next.

[12]For VSL we use the guidelines described by The Norwegian Agency for Public and Financial Management (DFØ) and the average value of VSL in 2020 and 2021. The hospitalization costs, taking into account the duration of hospitalization, for one COVID-19 patient was estimated by the Norwegian COVID-19 expert group to be 0.8 million if the patient needed intensive care, and 0.14 million if not. In the calculations reported here we, for simplicity, assume that all hospitalizations are intensive care patients.

5 Counterfactual simulations

In this section we analyze the effect of different policy strategies, i.e., the systematic component of containment policies. To do so we feed a standard SIR model with different time-varying reproduction series. In particular, we use estimates of the time-varying reproduction number (RN_t) based on Arroyo-Marioli et al. (2021), and ask how the three first months of the pandemic would have evolved if we; i) remove the effect of containment shocks from the RN_t estimate; ii) remove the effect of both the containment shocks and systematic containment policies from the RN_t estimate; iii) use a RN_t series reflecting how it would have been in the Norwegian case if Norway had followed the softer NPI policies adopted in Sweden.

In general, for all these SIR simulations, we assume that the so-called generation interval is 5 days and that 50 persons are infectious at the start of the simulations. The former assumption is consistent with many epidemiological studies on COVID-19 (Johansson et al., 2021), while the latter number is set such that the cumulative number of infected persons at the end of the simulation roughly matches the data when using RN_t. To facilitate interpretation, we normalize all the counterfactual scenarios such that the implied reproduction number equals the original RN_t on March 10., i.e., the start of the simulation period and two days before the Norwegian government enforced a lock-down. Finally, to infer the effect on mortality, we assume an Infection Fatality Rate (IFR) of one percent, which was a fairly common estimate for the middle-aged population in the beginning of the pandemic (Levin et al., 2020).

All of these assumptions are based on uncertain estimates, and in general, counterfactual modeling of this type should be interpreted more as suggestive than conclusive. We would still argue that it is useful because it highlights the difference between unexpected and systematic NPI strategies, and the nexus between the Norwegian and Swedish experience in particular.

5.1 Do nothing

The black line in the left graph in Figure 8 shows the RN_t for Norway, as estimated based on Arroyo-Marioli et al. (2021), during the three first months of the pandemic (Figure 9 reports the same statistic for the whole sample). The two other lines in the left graph in Figure 8 reports what this RN_t estimate would have been if not affected by containment policy shocks or containment policies in general. To derive the former series we simply regress RN_t (for the whole sample) on 21 days of lagged and contemporaneous containment shocks and use the residual from this regression as a counterfactual time series. Although not reported, the R^2 from this regression is roughly 0.1 and most of the

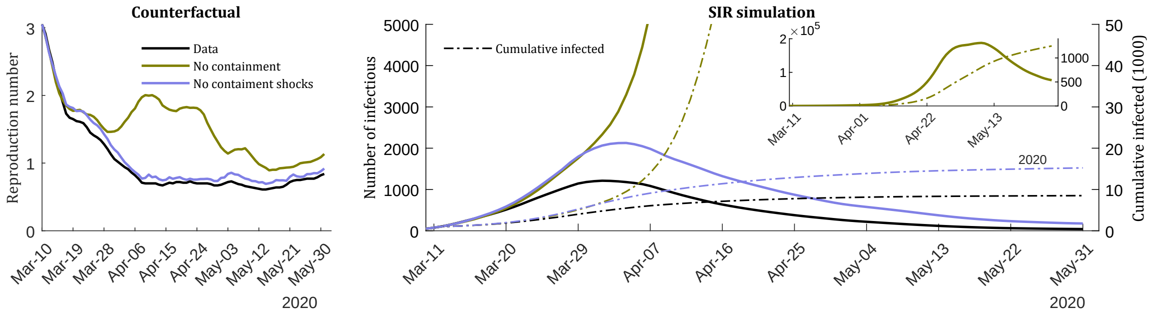

Figure 8. SIR simulations. The graph to the left reports the reproduction number in Norway (from Arroyo-Marioli et al. (2021)) together with two counterfactual estimates; Without any containment shocks and without any systematic containment policies. The graph to the right reports the resulting pandemic curves and cumulative number of infected when a standard SIR model is simulated using these time-varying reproduction numbers. See the text for further explanations.

explanatory variables are significant. To further clean the RN_t time series for containment policy effects, we redo this type of regression, but augment it with lagged values of the CON variable. The R^2 of this augmented regression is roughly 0.4. Letting RN_t^c denote the fitted value, the cleaned measure is simply $RN_t^{nc} = RN_t - RN_t^c$.

The right graph reports the prevalence and cumulative number of infected if we feed the time-varying reproduction series from above into a standard SIR model. By construction, the simulation using RN_t gives close to 10 thousand infected persons by the end of May, which is also what we observe in the data. Removing the effect of containment shocks increases this number to 15 thousand. Under the maintained assumption of an IFR of one percent (Levin et al., 2020), these numbers imply differences in mortality of around 50 persons, and squares reasonably well with the implications of containment shocks discussed in Section 4.3. Indeed, the sum of the containment shocks that according to our model realized during this time period is close to 10 (standardized units).

Removing the effect of containment policies from the RN_t statistics results in an exploding infectious curve, and the cumulative number of infected starts to level-off at around 1.2 million people only at the end of the simulation period. I.e., this "do nothing" scenario creates a much more severe pandemic than what we observe in the data. At the same time, however, this scenario is much less severe than what was typically characterized as a "do nothing" scenario by government authorities and experts in the early phases of the pandemic, see, e.g., Ferguson et al. (2020) and Table B.1. The reason is that the time-varying reproduction number falls, and converge towards one, also after removing the effect of containment. And, one reason for this is that people take voluntary actions to avoid being infected and potentially die. As shown in Section 4.2, there is a strong endogenous mobility response to infection shocks, and independent and unexpected behavioral actions drive down the number of infected (without necessarily having strong negative economic consequences). Figure C.5, in Appendix C, shows that this reasoning carries over to the SIR simulations done here. That is, when we also remove the effect of precautionary behavior from RN_t, the number of infected people grows to over three million by the end of the simulation period.

To gauge the economic costs associated with the no contaiment scenario we have also cleaned the FNI variable for the effect of contaiment policies in the same manner as above, and then compared the level of GDP implied by the original FNI series to the counterfactural no contaiment scenario using the mapping described in Section 4.3. According to these estimates the lagged CON variable and containment shocks explain approximately 20 percent of the variation in the FNI, and in the alternative scenario without containment effects, GDP would have been 12 billion NOK higher than in the benchmark case during the three month period we look at.

One important implication of these results is that when evaluating the potential costs and benefits of different NPI policies, the relevant benchmark of no response will most likely not result in a scenario where everyone get infected, as behavioral changes will prevent infection rates from exploding. Not taking this into account might easily overestimate the benefits of policy interventions relative to the economic costs.

Another insight is that the systematic NPI policy response likely matters more for domestic health outcomes than for the economy. I.e., relative to the analysis in Section 4.3, the change in mortality stemming from turning off containment policies is much larger than the (positive) change in GDP. On reason for this is that Norway is a small and open economy, and GDP would have contracted due to international conditions regardless of the NPI strategy.

5.2 The Scandinavian perspective

We now turn to the Norway versus Sweden comparison, and provide one answer to the question: “How would the pandemic have evolved if Norway had done as Sweden?”. These two countries are both small and open economies, with very similar governmental institutions, culture, and socioeconomic characteristics. However, as illustrated in Figure 3, they implemented very different NPI policies, at least in the early phases of the pandemic. Indeed, Sweden became internationally renowned for their rather soft NPI policies, and had according to many as a result 10 times as many dead as Norway on a per capita basis throughout much of the sample we look at. Still, also the Swedish authorities did something rather than “do nothing”.

To provide an estimate of the counterfactual question alluded to above, we proceed in two steps. First, we estimate how much of the difference in containment policies between Norway and Sweden could be explained by RN^c_t, defined in Section 5.1. I.e., we estimate

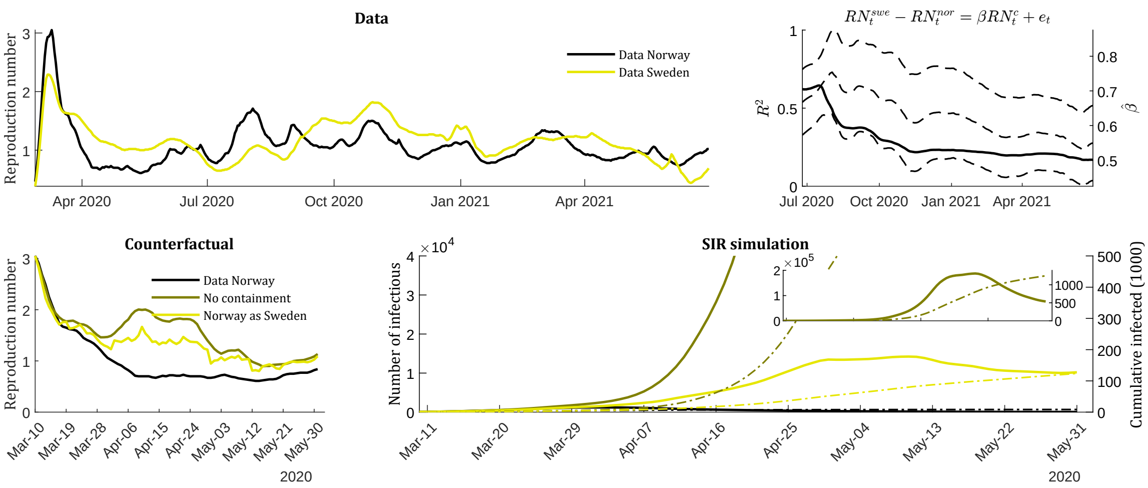

Figure 9. Norway as Sweden. The upper left graph reports the reproduction number in Norway and Sweden (from Arroyo-Marioli et al. (2021)). The upper right graph reports the coefficient estimate and R^2 from regressing the difference in the reproduction number in Norway and Sweden on RN_t^c (defined in the text). The lower left graph reports the reproduction number in Norway together with two counterfactual estimates; Without any systematic containment policies and if Norway had adopted Swedish policies (i.e., the same level of containment). The lower right graph reports the resulting pandemic curves and cumulative number of infected when a standard SIR model is simulated using these time-varying reproduction numbers. See the text for further explanations.

the equation $RN_t^{swe} - RN_t^{nor} = \beta RN_t^c + e_t$. The upper left graph in Figure 9 reports the RN_t^{swe} and RN_t^{nor}, while the upper right graph reports the R^2 and $\hat{\beta}$ when the sample length used to estimate the equation is continuously expanded from a starting point of 100 days. The parameter estimate is in the range 0.6-0.7 and significant.[13] The R^2 estimate starts at close to 0.7 and then falls towards 0.2, consistent with the general belief that the difference between Swedish and Norwegian NPI policies were particularly large in the early phases of the pandemic. Second, we compute $RN_t^{nor,swe} = RN_t^{nor} - \hat{\beta} RN_t^c$ as the counterfactual measure of the reproduction rate Norway would have had if it had followed the same NPI policies as Sweden.

The lower left graph in Figure 9 reports the $RN_t^{nor,swe}$ counterfactual together with the original reproduction number and RN_t^{nc}. In line with the fact that also the Swedish authorities implemented NPI policies, the $RN_t^{nor,swe}$ estimate lies between the original estimate and the "do nothing", or no containment, estimate. However, relatively small changes in the reproduction number can have large effects on the evolution of the pandemic. As seen in the SIR simulations in the lower right graph in Figure 9, if Norway had followed Swedish containment policies the total number of infected persons would have been over 100 thousand persons by the end of the spring in 2020. Accordingly, the value

[13] These estimates should be interpreted with caution since RN_t^c is a generated regressor. However, all of the quantities used in this analysis are highly uncertain and should be interpreted accordingly.

of saved lives in Norway from following the Norwegian strategy relative to the Swedish policy in these simulations is roughly 40 billion NOK while the more stringent policy likely lowered GDP by less than 12 billion per quarter, yielding a positive net gain.[14] Or, in other words, the output cost per life saved in Norway from following the Norwegian strategy relative to the Swedish one is roughly 10 million NOK.

Does this imply that Norway handled the pandemic better than Sweden? If we take the counterfactual simulations above literally, the answer might be yes for the time span we analyze here. However, if one takes a more international perspective, the difference in outcomes within Scandinavia are likely exaggerated. E.g., a large range of western countries had much worse mortality and economic outcomes (see Section 6). Likewise, as discussed below, a fairly large degree of similarity between Norway and Sweden, rather than difference, is what's striking when considering a longer time period.

The left graph in Figure C.6, in Appendix C, reports an estimate of the overall cumulative excess mortality in Sweden and Norway for the time period 2019 to 2022. As seen in the graph, Sweden and Norway went into the pandemic with negative cumulative excess mortality. However, this effect was much larger for Sweden than in Norway. Thus, an analysis of mortality that normalizes these statistics to the same starting value in March 2020 would easily exaggerate the differences, a point first made by Juul et al. (2022). As also seen in the graph, while Norway had a relatively flat curve throughout 2020, it started to climb in mid 2021. Around the same time, the Swedish numbers start to fall, and by mid 2022 they actually lower than in the Norwegian case.

The right graph in Figure C.6 illustrates the aggregate economic developments in these two countries. Both Sweden and Norway had fairly parallel growth trends up to 2019. When the pandemic hit, GDP in both countries contracted significantly, but by late 2021, the level of GDP was at the same level as one would predict using data prior to 2020. I.e., the large decline following the initial phases of the pandemic was outweighed by an even more significant rebound. More than highlighting a large difference, the graph illustrates a relatively large degree of similarity.[15]

[14] The difference in the number of infected is $120000 - 10000 = 110000$, and we continue to assume an IFR of one percent and a VLS of 35 million. As discussed in Section 5.1, implementing containment policies likely lowered GDP by roughly 12 billion NOK per quarter, creating an upper bound for how much GDP could fall following the policy change.

[15] The observant reader will notice that Norway's economy contracted less than the Swedish one in the spring of 2020. In Appendix B.3 we describe how controlling for international developments gives estimates suggesting that Sweden had roughly a one percentage point higher growth rate than Norway, which is fairly consistent with the effect sizes of containment policies documented earlier.

6 Discussion and robustness

Modeling the joint dynamics between health and macroeconomic developments during the COVID-19 pandemic is challenging, and relatively small differences in assumptions and implementation can give rise to differences in results. To the best of our knowledge, only two other empirical studies taking a similar dynamic general equilibrium perspective as we do exist; Camehl and Rieth (2023) and Arias et al. (2023). Below we discuss relevant similarities and differences between our and their studies to highlight important considerations for this type of modeling.

Camehl and Rieth (2023) pull information from 44 countries, accounting for (an impressive) 81 percent of worldwide infections and deaths due to COVID-19 and for 72 percent of global GDP. They then estimate a Bayesian Panel VAR and identify three shocks; COVID-19 incidence shocks, containment policy shocks, and economic mobility shocks, using the sign-restrictions algorithm proposed by Arias et al. (2018).

The containment policy shock in Camehl and Rieth (2023) has the same interpretation as in our model, and does to a large extent generate similar impulse response dynamics as we find. However, whereas they find that this shock accounts for between 20 and 30 percent of the variability in cases and deaths, our estimates are more in the range 40 to 50 percent. Related to this, the most important shock in their study is the COVID-19 incidence shock, which explains roughly three times as much of the variation in cases and deaths in their work than our comparable infection shock does.

One plausible reason for this difference is that Camehl and Rieth (2023) estimate a global model whereas we estimate a model for one single (small) country and include a daily business cycle index. This allows us to separately identify mobility and news shocks, where the latter captures international economic and health developments that likely will affect the domestic economy, but yet (contemporaneously) has not. As we document, the forward-looking economic variables in our model respond to this shocks, and so does policy. I.e., economic agents and policy-makers observe an infection wave coming and respond accordingly. In turn, this likely decreases the importance of infection shocks in our model relative to in the work by Camehl and Rieth (2023). Put differently, it is difficult for the world to import COVID-19, but for a single country the pandemic most likely is imported and followed by news that can be acted upon before arrival.

The study by Arias et al. (2023) shares our single economy focus by studying the case of Belgium. They first use Sequential Monte Carlo methods to fit a time-varying compartment model and then use the estimated output from this model, i.e., a time-varying reproduction number, new cases, and mortality, together with external variables measuring NPI policies, economic sentiment, and mobility, in a SVAR model. An advantage of this approach is that statistics such as new cases (and thereby the estimate of the time-varying reproduction number) becomes consistently estimated together with, e.g., mortality, within the compartment model machinery, and thereby more robust to measurement errors (due to, e.g., under-reporting).[16]

In their SVAR implementation they use the estimation and identification procedures developed in Arias et al. (2018) and Arias et al. (2019) to identify one shock, namely unexpected stringency policy innovations. The interpretation of this shock is similar to that in our work, but is identified under the assumption that the policy equation of the system responds contemporaneously and positively to the reproduction number, mobility, new cases, deaths, and the economic sentiment index. In contrast, we assume that policy only responds with a lag to any other developments in the model.

In line with our results they find that the stringency shock has a short-lived effect on the stringency index itself, but gives rise to a persistent decline in mortality and also leads to a drop in economic sentiment and mobility. However, where we find a substantial delay of close to three weeks, the cumulative number of deaths is significantly negative already after 10 days in Arias et al. (2023). Moreover, the drop in economic sentiment and mobility in their study is only short-lived. After roughly two weeks, both variables have turned from negative to positive. In our work the related variables display a much more persistent negative response to containment shocks, lasting for at least 40 days.

As a result, Arias et al. (2023) estimate that a containment shock which leads to roughly 1000 fewer deaths only reduces GDP by 330 million EURO or less with 84 percent probability, i.e., an output cost per life saved equaling 330000 EURO, or roughly 3 million NOK. From this they conclude that additional government-mandated mobility curtailments would have reduced deaths at a very small cost in terms of GDP. Although their (lower tail) GDP response is similar to what we find for Norway, their mortality implications are vastly different from our SVAR analysis.

The reason for these differences is not that we identify more shocks. As discussed in Section 4.1, the data supports four relative to fewer common shocks, but as shown in Figure C.8, in Appendix C, the impulse response results for the key outcome variables are very similar if we instead estimate a model with only the first two or three shocks. The figures also show that the differences in identification assumptions regarding the

[16] Although under-reporting is an issue also with Norwegian data, particularly in the beginning of the pandemic, the severity of this issue is likely smaller than in many other countries. Figure C.7, in Appendix C, provides evidence suggesting that the share of reported cases in Norway during in March and April 2020 was likely close to 40 percent. In contrast, Arias et al. (2023) estimate that the share of reported cases in Belgium during March and April was around 20 percent, which is consistent with the estimates provided in Figure C.7. In their study they also analyze the effects of reproduction shocks and whether its effects on the virus's spread, deaths, and economic activity depend on the level of government stringency. This analysis has no counterpart in our study, so we do not discuss it further.

NPI policy do not explain the discrepancies. I.e., re-estimating our model and allowing for containment policies to respond contemporaneously to the other shocks in the model does not lead to qualitatively different conclusions.

A more likely explanation is that country differences matter a lot. For the sample we look at, Norway had one of the lowest COVID-19 mortality rates in the world. In contrast, Belgium was one of the European countries hardest hit, with over 16000 dead (1600 dead per million people) by the end of 2020. For Norway and Sweden the comparable numbers were 400 and 9000 (80 and 900 per million people), respectively. Moreover, the relative sizes of these statistics were roughly the same already by the early spring of 2020.[17] From this perspective, the initial conditions and implied steady-state likely play a large role in terms of determining the effect sizes. In particular, a stringent lock-down policy will drive down economic activity irrespective of how many people are infectious. However, the mortality effect of the same policy will depend heavily on the number of infectious when the policy is put in place. In the SVAR, the impulse responses capture departures from steady-state, and since the average number of infected and mortality was significantly higher in Belgium than in Norway, the effect of containment shocks on mortality is also stronger there.

We finally note that the NPI policy variables used in both Camehl and Rieth (2023) and Arias et al. (2023) are sourced from OxCGRT. We have not evaluated to what extent these series, for countries other than Norway and Sweden, contain measurement errors. Thus, it might very well be that some of the differences in results discussed above also stem from this type of issue. Another difference is that both Camehl and Rieth (2023) and Arias et al. (2023) use a sample covering the first two waves of the pandemic, and end estimation late 2020. To obtain more observations for estimating the high-dimensional SVAR we include also the third wave of the pandemic in our sample. However, Figure C.8, in Appendix C, shows that shortening the sample to roughly one year of observations does not qualitatively change our conclusions.[18]

7 Conclusion

We use a daily business cycle and containment policy index to model the high-frequency dynamic interaction between COVID-19 NPI policies, health, and business cycle outcomes

[17] As such, the containment shock size considered in Arias et al. (2023) accounts for roughly six percent of observed deaths in Belgium, and our containment shock for roughly one percent of observed deaths in Norway.

[18] For this experiment we end estimation in mid February 2021, which is the earliest endpoint we can use given the number of parameters in the model. At this point in time the third COVID-19 wave had already started in Norway.

within a SVAR framework. Focusing on Norway and Sweden our analysis documents potentially large measurement errors in commonly used containment policy measures, which we correct for using a manual audit, and significant endogeneity between important variables, such as infection rates, mobility, and containment policies.

Applying sign restrictions, and assuming reduced rank for the stochastic elements of the model, our results suggest that both containment policy shocks and unexpected variation in precautionary behavior lower the pandemic burden, and that both types of shocks have significant adverse economic effects. Moreover, we document that a large share of the variation in containment policies is driven by anticipatory precautionary (news) innovations.

To highlight the difference between unexpected and systematic NPI strategies, and the nexus between the Norwegian and Swedish experience in particular, we run a number of counterfactual simulations using a simple compartment model and the SVAR output. According to our simulations, the gains from departing from the widely used "Do nothing" benchmark strategy are likely inflated, while the effect of systematic relative to unexpected NPI policies are stronger for health outcomes than economic outcomes. As a result, we estimate that the output cost per life saved in Norway from following the Norwegian strategy relative to the Swedish one is roughly 10 million NOK.

A natural extension of our analysis would be to merge the general equilibrium relationships we study in the SVAR with a more structural epidemiological model. This would likely require the usage of more sophisticated Sequential Monte Carlo methods, but offer the benefit of one coherent framework for analyzing the role of both systematic and unexpected policy changes.

References

Aksoy, C. G., J. M. Barrero, N. Bloom, S. J. Davis, M. Dolls, and P. Zarate (2022). Working from Home Around the World. Working Paper 30446, National Bureau of Economic Research.

Alvarez, F. E., D. Argente, and F. Lippi (2020). A simple planning problem for Covid-19 lockdown. Technical report, National Bureau of Economic Research.

Andersen, T., S. Holden, S. Honkapohja, W. Eichhorst, J. Brunner, and G. Zoega (2022). COVID-19 Effects on the Economy in the Nordics. *Nordic Economic Policy Review*.

Andersson, S. and N. Aylott (2020). Sweden and coronavirus: Unexceptional exceptionalism. *Social Sciences 9*(12), 232.

Arias, J. E., D. Caldara, and J. F. Rubio-Ramirez (2019). The systematic component of monetary policy in SVARs: An agnostic identification procedure. *Journal of Monetary Economics 101*, 1–13.

Arias, J. E., J. Fernández-Villaverde, J. R. Ramírez, and M. Shin (2023). The causal effects of lockdown policies on health and macroeconomic outcomes. *American Economic Journal: Macroeconomics 15*(3), 287–319.

Arias, J. E., J. F. Rubio-Ramirez, and D. F. Waggoner (2018). Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications. *Econometrica 86*(2), 685–720.

Arroyo-Marioli, F., F. Bullano, S. Kucinskas, and C. Rondon-Moreno (2021, 01). Tracking R of COVID-19: A new real-time estimation using the Kalman filter. *PLOS ONE 16*(1), 1–16.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Beaudry, P. and F. Portier (2014). News-Driven Business Cycles: Insights and Challenges. *Journal of Economic Literature 52*(4), 993–1074.

Berger, D., K. Herkenhoff, C. Huang, and S. Mongey (2020). Testing and reopening in an SEIR model. *Review of Economic Dynamics*.

Bjørnland, H. C. and L. A. Thorsrud (2016). Boom or gloom? Examining the Dutch disease in two-speed economies. *The Economic Journal 126*(598), 2219–2256.

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni (2013). News, Noise, and Fluctuations: An Empirical Exploration. *American Economic Review 103*(7), 3045–70.

Brodeur, A., D. Gray, A. Islam, and S. Bhuiyan (2021). A literature review of the economics of COVID-19. *Journal of Economic Surveys 35*(4), 1007–1044.

Camehl, A. and M. Rieth (2023). Disentangling Covid-19, economic mobility, and containment policy shocks. *American Economic Journal: Macroeconomics 15*(4), 217–248.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Deb, P., D. Furceri, J. Ostry, and N. Tawk (2020). The Economic Effects of COVID-19 Containment Measures. Technical Report 20/158.

Deb, P., D. Furceri, J. D. Ostry, N. Tawk, and N. Yang (2021). The Effects of Fiscal Measures During COVID-19. Technical Report 21/262.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2021). The macroeconomics of epidemics. *The Review of Financial Studies 34*(11), 5149–5187.

Esaiasson, P., J. Sohlberg, M. Ghersetti, and B. Johansson (2021). How the coronavirus crisis affects citizen trust in institutions and in unknown others: Evidence from 'the Swedish experiment'. *European Journal of Political Research 60*(3), 748–760.

Ferguson, N. M., D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, et al. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand.

Forni, M., L. Gambetti, M. Lippi, and L. Sala (2017). Noisy news in business cycles. *American Economic Journal: Macroeconomics 9*(4), 122–52.

Gamtkitsulashvili, T. and A. Plekhanov (2023). Mobility and economic activity around the world during the Covid-19 crisis. *Applied Economics Letters 30*(5), 608–614.

Hale, T., N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour 5*(4), 529–538.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics 100*(5), 831–843.

Herby, J., L. Jonung, and S. Hanke (2022, January). A Literature Review and Meta-Analysis of the Effects of Lockdowns on COVID-19 Mortality. Studies in Applied Economics 200, The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise.

Holdenutvalget (2020). Samfunnsøkonomiske vurderinger av smitteverntiltak (COVID-19). Technical report, Helsedirektoratet.

Johansson, M. A., T. M. Quandelacy, S. Kada, P. V. Prasad, M. Steele, J. T. Brooks, R. B. Slayton, M. Biggerstaff, and J. C. Butler (2021). SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA network open 4*(1), e2035057–e2035057.

Jones, C. J., T. Philippon, and V. Venkateswaran (2020). Optimal mitigation policies in a pandemic: Social distancing and working from home. Technical Report 26984.

Juranek, S. and F. Zoutman (2020). The effect of social distancing measures on intensive care occupancy: Evidence on COVID-19 in Scandinavia. *NHH Dept. of Business and Management Science Discussion Paper* (2020/2).

Juul, F., H. Jodal, I. Barua, E. Refsum, Ø. Olsvik, L. Helsingen, M. Løberg, M. Bretthauer, M. Kalager, and L. Emilsson (2022). Mortality in Norway and Sweden during the COVID-19 pandemic. *Scandinavian Journal of Public Health 50*(1), 38–45.

Korobilis, D. (2022). A new algorithm for structural restrictions in Bayesian vector autoregressions. *European Economic Review 148*, 104241.

Larsen, V. H. and L. A. Thorsrud (2019). Business cycle narratives. Technical report.

Levin, A. T., W. P. Hanage, N. Owusu-Boaitey, K. B. Cochran, S. P. Walsh, and G. Meyerowitz-Katz (2020). Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *European Journal of Epidemiology 35*(12), 1123–1138.

Linton, N. M., T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine 9*(2), 538.

McKibbin, W. and R. Fernando (2021). The global macroeconomic impacts of COVID-19: Seven scenarios. *Asian Economic Papers 20*(2), 1–30.

Mishra, S., J. A. Scott, D. J. Laydon, S. Flaxman, A. Gandy, T. A. Mellan, H. J. T. Unwin, M. Vollmer, H. Coupland, O. Ratmann, M. Monod, H. H. Zhu, A. Cori, K. A. M. Gaythorpe, L. K. Whittles, C. Whittaker, C. A. Donnelly, N. M. Ferguson, and S. Bhatt (2021). Comparing the responses of the UK, Sweden and Denmark to COVID-19 using counterfactual modelling. *Scientific Reports 11*(16342).

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Padhan, R. and K. Prabheesh (2021). The economics of COVID-19 pandemic: A survey. *Economic Analysis and Policy 70*, 220–237.

Rubio-Ramirez, J. F., D. F. Waggoner, and T. Zha (2009). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies* (4), 1–34.

Sheridan, A., A. L. Andersen, E. T. Hansen, and N. Johannesen (2020). Social distancing laws cause only small losses of economic activity during the COVID-19 pandemic in Scandinavia. *Proceedings of the National Academy of Sciences 117*(34), 20468–20473.

Sims, C. A. and T. Zha (2006). Does Monetary Policy Generate Recessions? *Macroeconomic Dynamics 10*(02), 231–272.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

Thorsrud, L. A. (2018). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, 1–17.

World Health Organization (2020). WHO Report of the WHO-China Joint Mission on Coronavirus Disease 28 February, 2020.

Yarmol-Matusiak, E. A., L. E. Cipriano, and S. Stranges (2021). A comparison of COVID-19 epidemiological indicators in Sweden, Norway, Denmark, and Finland. *Scandinavian Journal of Public Health 49*(1), 69–78.

Appendices for online publication

Appendix A Revised NPI indexes

In this section we first shortly present the most important features of the Oxford Covid-19 Government Response Tracker (OxCGRT) code-book. We then describe how we have used this code-book, together with additional data sources to construct revised NPI measures for Norway and Sweden. Finally, we present the manual audit which we used to verify the proposed revisions.

A.1 The original OxCGRT codebook

In this section, we provide a detailed overview of the original Oxford COVID-19 Government Response Tracker (OxCGRT) index. The OxCGRT quantifies non-pharmaceutical policy interventions (NPIs) across 180 countries, initiated from January 2020 onward. It utilizes ordinal or continuous scales to record government actions in 24 distinct policy areas, including but not limited to, school closures, border restrictions, and limitations on gatherings. Of particular interest to this paper is the "Containment and Health Index," one of several composite indexes generated by OxCGRT. This index amalgamates both lock-down measures and health-related policies such as testing protocols and contact tracing, covering a subset of 14 policy indicators.

Indicators in the Containment and Health Index are constructed using two primary inputs: an ordinal value and a flag variable for geographical specificity. The ordinal values, denoted as $v_{j,t}$, range from 0 to 5 and are assigned according to the OxCGRT code-book. A higher value signifies a more restrictive policy. For instance, the indicator $C1$—school closings—would be assigned a value of 2 if schools have been required to close, but only for some levels, such as high schools.

Since policies can vary within a country, each indicator captures the strictest policy in place within a given jurisdiction. To illustrate, if only high schools are closed nationally but all schools are closed in Oslo, the latter, stricter measure is the one recorded. A flag variable f_j is used to indicate whether the strictest policy is local or national. Each sub-index score ($I_{j,t}$) for a given indicator j on a given day t is calculated by the function below:

$$I_{j,t} = 100 \times \frac{v_{j,t} - \omega(F_j - f_{j,t})}{N_j} \tag{7}$$

F_j is a binary variable which records presence or absence of a flag for a given indicator. For example, certain indicators, such as international travel control (C8), are relevant to a country's entire jurisdiction and hence do not have a geographic flag, rendering $F_j = 0$.

Meanwhile, N_j represents the maximum attainable value for the respective indicator. The weight ω is set to 0.5 in the original formulation of OxCGRT to adjust for the possibility that the strictest measure may not be universally applied within the jurisdiction. This normalizes the different ordinal scales to produce a sub-index score between 0 and 100 where each full point on the ordinal scale is equally spaced.

There are a few features of the OxCGRT that need to be highlighted for a comprehensive understanding: 1) The indicators are updated to reflect the day a policy is implemented rather than when it is announced. 2) Indicators reflect only the strictest restriction put in place, independently of whether these restrictions are local or national. 3) The indicators capture policy restrictiveness, not compliance levels.

Specifically, if a localized policy is the strictest in the jurisdiction, it will dominate the country-wide stringency index. This is particularly evident when the geographical distance between the area with the strictest policy and the rest of the country is considerable. For instance, consider a scenario where $v_{j,t} = 1$ in Oslo and $v_{j,t} = 0$ elsewhere in the country. The indicator C_1 would then yield $I_{c_1,t} = \frac{0.5}{3} \approx 16.66$. The discrepancy between 0 and 16.66 is relatively minor. However, if $v_{j,t} = 3$ in Oslo while remaining 0 elsewhere, the resulting $I_{j,t} = \frac{2.5}{3} \approx 83.33$. In this case, the divergence becomes substantial, potentially exaggerating the level of stringency as experienced by the larger population.

A.2 The revised codebook

We rely on the framework provided by OxCGRT as a baseline measure of the aggregate level of NPIs in Norway and Sweden. We then introduce three alterations to the original composite index, with a particular emphasis on key features 1) and 2) described in Section A.1.

The first adjustment concerns the removal of geographic flags that were originally part of the index. These flags serve to indicate whether a policy measure is in force nationally or specific to a particular geographic region, with the index reflecting the more stringent of the two. The flag variable is problematic for two reasons. First, it may misrepresent the overall perceived stringency level, especially when a localized but highly stringent policy, compared to the overall restriction level, significantly influences the index for an extended period, as highlighted in the example is Section A.1. Moreover, it introduces a risk of measurement errors, as local restrictions are more difficult to track than the national ones.

This removal of geographic flags necessitates two adjustments to the overall containment and health index. First, the formula, described in Equation (7), need modification by setting ω to 0, as opposed to its previous value of 0.5. Second, for instances where a restriction is both local and more stringent compared to national measures, the task arises to ascertain the value of the policy measure for the remainder of the country. To accurately represent these values, we consult government press releases and media archives in Norway and Sweden. Specifically, the timeline of press releases from the Norwegian Ministries related to COVID-19 serves as a benchmark for determining stringency measures in Norway[19].

Secondly, our examination of the various indicators recorded from Jan 2020 - July 2021 revealed numerous inconsistencies in the interpretation of the OxCGRT code-book. Consequently, we undertook a comprehensive review of all indicators to ensure they align with the level of stringency indicated by national restrictions as documented in the media and official government press releases. This review uncovered multiple inconsistencies between how we interpret the OxCGRT code-book and how it was initially implemented. For example, consider the C1 indicator, which reflects the degree of school closures. According to the code-book, a value of '1' suggests that the government recommends either closing all schools or conducting significantly altered operations. In the period from May 11, 2020, to June 15, 2020, C1 was coded as '1' for Norway. This does not accurately reflect the situation, as universities were not fully open, and classes were available only to specific groups. A more accurate coding, as per the code-book, would assign a value of '2' to the C1 indicator for this period.

The source of many such discrepancies can often be traced back to the ambiguous wording of policy measures. This is particularly evident in the case of Swedish policies. To illustrate, on March 16, 2020, Swedish authorities issued a statement advising, "Employers who have the opportunity to allow employees to work from home may consider recommending this." The statement leaves room for interpretation. The statement could either be interpreted as a formal recommendation, warranting a score of '1,' or as a simple suggestion, corresponding to a score of '0' in the OxCGRT code-book.

Third, the index is a measure of policy responses at the date of implementation. This poses a concern as measures announced prior to their implementation are often anticipated, leading to shifts in variables such as mobility and business cycle indicators ahead of the official implementation. While the time lag between policy announcement and implementation is typically short—often just a few hours to a few days—we opt to refine the index to capture this anticipation effect. Specifically, we adjust the index to reflect changes as of the date of the policy announcement, rather than the date of its formal implementation.

[19]https://www.regjeringen.no/no/tema/Koronasituasjonen/tidslinje-koronaviruset/id2692402/

A.3 A manual audit

The audit of the OxCGRT index, discussed in Section A.2, was conducted by one of the authors and two independent reviewers. These reviewers were Norwegian master's students in economics from BI Norwegian Business School and were compensated for their efforts.

The audit took place between June and July of 2023 and relied on data series and accompanying notes from the OxCGRT GitHub repository[20], extracted August 2021, covering policies in both Sweden and Norway. The reviewers were individually tasked with examining all 14 indicators from 1. January 2020 - 31. August 2021, with explicit instructions to account for gaps in the data set resulting from the removal of geographic flags and to address these gaps as needed.

For the purpose of the audit, reviewers were given access to a benchmark data set, comprising the extracted time series and notes from volunteer contributors. These notes outlined changes in policy indicators and their respective values which were primarily supported by media reports. Furthermore, reviewers were provided with the OxCGRT's general code-book as well as a detailed supplementary guide tailored to circumstantial use. Reference to the timeline of press releases from the Norwegian Ministries related to COVID-19 was also provided, as discussed in Section A.2.

Appendix B Data and auxiliary calculations

B.1 A public spending indicator

The G variable is constructed using newspaper archives and the timeline of press releases from the Norwegian Ministries related to COVID-19 as discussed in Section A.1. We use this timeline of press releases to construct a data set with all fiscal interventions differentiated by type of intervention, magnitude, announcement date and whether the intervention is a proposal, an adoption or an implementation. We direct our attention solely to proposals. The type of intervention varies from proposals of state loan guarantee for bank loans to large compensations schemes to help companies cover fixed costs to increases in the unemployment benefit. The monetary size of the policy proposal ranges from five million Norwegian kroner up to 100 billion kroner. The press releases are supplemented by and cross-checked with Yale's COVID-19 Financial Response Tracker (CFRT). In cases where the two data sources differ, we favor the press releases from the Norwegian Ministriese. Both data sources lack information about the economic magnitude of certain interventions. Common for these policy proposals are that the size of the policy depends

[20] https://github.com/OxCGRT/covid-policy-dataset/tree/main

on the degree to which people will be affected by it. For example, the cost of increasing the unemployment benefits will depend on the number of unemployed and the duration of their unemployment status. Precise measures of magnitude is therefore difficult to produce. In these instances, we use cost estimates from the Ministry of Finance, found in draft resolutions and bills issued by the government and cross-check these estimates with newspaper articles.

B.2 Calculating excess mortality

To calculate the excess mortality statistics in Figure C.6 we obtain monthly data on all case mortality in Norway and Sweden from the two countries' health authorities for the sample January 2012 to December 2022. We then compute the expected number of deaths, and the standard deviations, for each month in the period 2012 to 2018, and use these statistics to generate predictions for the period 2019 to 2022. The uncertainty associated with these predictions are simply drawn from the normal distribution using the associated standard deviations. In the figure, the (percentage) difference between the actual and predicted cumulative mortality statistics are reported.

Although much more sophisticated methods can be used to produce this type of statistic, the excess mortality statistics obtained here are well in line with what other studies find. The crucial difference is that we take into account the mortality displacement due to the low all-cause mortality in the year prior to the pandemic, i.e., 2019, which was particularly pronounced for Sweden (Juul et al., 2022).

B.3 Controlling for international business cycles

The numbers presented in Figure C.6 suggest that Norway's economy contracted less than the Swedish one in the spring of 2020. As Norwegian containment policies were more stringent than in Sweden, and we find that containment policies drive down growth, this might seem like a puzzle. Although difficult to measure precisely, there is little evidence suggesting that differences in governmental support packages can explain the growth differences (Andersen et al., 2022). A more plausible explanation is that Sweden's economy is more dependent on international business cycles than Norway's economy.

To control for international business cycle developments in Figure C.6 we first obtain real GDP growth data for the ten most important trading partners for Norway and Sweden, respectively. From these data we then extract one common growth factor using Principal Component Analysis on the sample 1996:Q1-2021:Q4. The common factors explains roughly 90 percent of the variability in each of the two datasets. Finally, we regress (log) GDP growth in Norway or Sweden on the associated common trading partner growth factor and use the resulting residual as a measure of the purely domestic growth contribution. In the figure, the cumulative sum of the domestic growth contributions, starting in 2019:Q1, is reported.

B.4 The response of the Norwegian authorities to the pandemic

To interpret the findings in Section 4.3 and put them into context, it is illuminating to provide some more details about how the Norwegian authorities managed the crisis. In particular, in the early spring of 2020, an expert group of economists and health scientists proposed and analyzed four different strategies. In turn, these strategies were continuously used as reference points in the subsequent policy debate. Table B.1 lists the strategy options together with the expected mortality and economic costs associated with each policy option. The four strategies considered were: (i) "do nothing", i.e., the government would not impose restrictions to prevent the number of infected from increasing, (ii) "break" the trend in new infection rates, by imposing sufficient restrictions (iii) "suppress" the number of infected, by imposing strong restrictions following each wave of infections and (iv) "suppress and hold down" by imposing the strongest set of restrictions and keep them in place so that the number of new infection rates remained suppressed.

Of the four policy options considered, it is commonly agreed upon that it was the "Suppress" strategy that was followed, at least during the first year of the pandemic.[21] In light of this, a positive containment shock in our model can be interpreted as making an unexpected move to the right in Table B.1, starting from a "Suppress" steady-state. As a one-time one standard deviation containment shock in our model saves three lives and reduces growth by three billion NOK within a quarter, a series of such shocks, or one large shock in the range of 7-10 standard deviations, is needed to move the mortality and economic costs statistics all the way from the "Suppress" scenario to the "Suppress and hold down" policy.[22] Since this policy regime is systematically more stringent than the "Suppress" regime, the required shock sequence and size implications are far from

[21] Data outcomes are also relatively consistent with this "Suppress" interpretation. We know now that roughly 400 persons were confirmed dead by COVID-19 by the end of 2020, and roughly 800 persons by the end of our sample, in July 2021. Clearly, the "Do nothing" and "Brake" scenarios are nowhere close to match this, while the two scenarios to the right in the table are much more reasonable. Comparing the economic scenarios to actual outcomes is much more problematic because governmental support packages can have helped reduce the actual losses. I.e., governmental support packages can compensate economic losses but not bring back lost life.

[22] The mortality and economic costs statistics reported in Table B.1 apply for 2020 as a whole. To make these numbers comparable to those in Figure 7, we assume for simplicity that they can be divided equally into each quarter of the year. E.g., the change in mortality is $(200 - 320)/4 = -30$, and the containment shock in our model needs to be of size $-30/-3 = 10$ standard deviations.

Table B.1. Expert group scenarios. Obtained from the report Holdenutvalget (2020), produced in early April 2020 for the year 2020. See $https://no.wikipedia.org/wiki/Holden-utvalget_(covid-19)$ for this and later reports. Numbers in parenthesis reflect the difference between subsequent column scenarios. The mortality estimates were obtained from relatively standard compartment model simulations without economic and behavioral interactions, while the economic scenarios where produced using economic models entertained by the Norwegian authorities also prior to the pandemic. In the report three different economic models where used to construct the economic cost projections. The average estimate is reported here. The two last rows are our own calculations based on assuming an IFR of one percent and a VSL of 35 million.

	Do nothing	Policy options		
		Brake	Suppress	Suppress and hold down
Mortality (persons)	47600	12400 (-35200)	320 (-12080)	200 (-120)
Economic cost (billion)		-80 (-80)	-120 (-40)	-200 (-80)
Total cost (billion)		-514	-131	-207
Cost change (billion)		-1152	-383	+75

implausible. The important point here is that the effect of shocks in our model are totally unrealistic for the scenarios to the left in the table, but reasonable for the scenarios to the right in the table, which also reflect the systematic NPI policies actually followed.

The simulations provided by the Norwegian expert group provide a cross-check for our own estimates, but also highlight how to interpret our analysis in general and the cost-benefit analysis in particular: Our results reflect the effects of unexpected fluctuations around a given steady-state conditional on the systematic policy response adopted, i.e., the policy strategy. For this reason, we can claim that there is evidence in the data that a containment shock has as a negative cost-benefit effect, but cannot claim that the adopted NPI strategy is better or worse relative to, e.g., doing nothing. However, getting the alternative scenarios fairly well calibrated is important. As an example, is it really plausible that a benchmark "Do nothing" scenario would make the whole population infected and cause 50000 deaths within short amount of time, as implied by Table B.1?[23] Or, is this mortality number too large and thereby inflating the potential gains from adopting more stringent NPI policies? This question is addressed in Section 5.

[23] The "Do nothing" scenario reported in Table B.1 reflects a simulation of a compartment model without economic and behavioral interactions. If the IFR is around one percent, the scenario simply reflects that the whole population would be infected within short amount of time and one percent of these die. For Noway this would have amounted to roughly 50000 persons. Even in the countries in the world that was hardest hit by the pandemic, mortality rates of this size has not been recorded for the sample we look at. Thus, irrespective of the chosen NPI strategy, the "Do nothing" reference point illustrated in Table B.1 seems highly questionable. One important reason for this is that the reproduction number, which is an important parameter in compartment model simulations, often is held fixed, or not allowed to respond to economic and behavioral interactions, during the evolution of the pandemic.

Appendix C Additional results

Table C.1. GDP growth and approximations. The table reports the coefficients from regressing GDP growth (Δy_t or $\Delta_8 y_t$) on the FNI, (log) electricity consumption, and (log) No2 emissions. Consumption and No2 emissions are transformed to year-on-year changes to remove seasonality in the raw data. Sample lengths vary due to data availability. All regressions end in 2019:Q4. Standard errors are computed following Newey and West (1987).

	Δy_t				$\Delta_8 y_t$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FNI	0.529***			0.702*	2.668***			3.066***
	(0.105)			(0.423)	(0.401)			(0.717)
El. Consumption		2.290		-0.452		13.787*		-2.184
		(1.615)		(1.520)		(8.156)		(7.337)
No2 emissions			0.156	0.958			0.150	3.641*
			(1.389)	(1.559)			(2.765)	(1.901)
No. obs	120	96	48	48	120	96	48	48
R^2	0.13	0.01	0.00	0.08	0.47	0.05	0.00	0.42

Table C.2. DIC. The table reports the Deviance Information Criterion for four different model specifications. The shock abbreviations are infection (INF), containment (CON), new (N), and mobility (M).

	Number of shocks			
	2	3	4	5
	INF & CON	INF & CON & N	INF & CON & N & M	INF & CON & N & M & ...
DIC	4288	2814	-24171	-47261

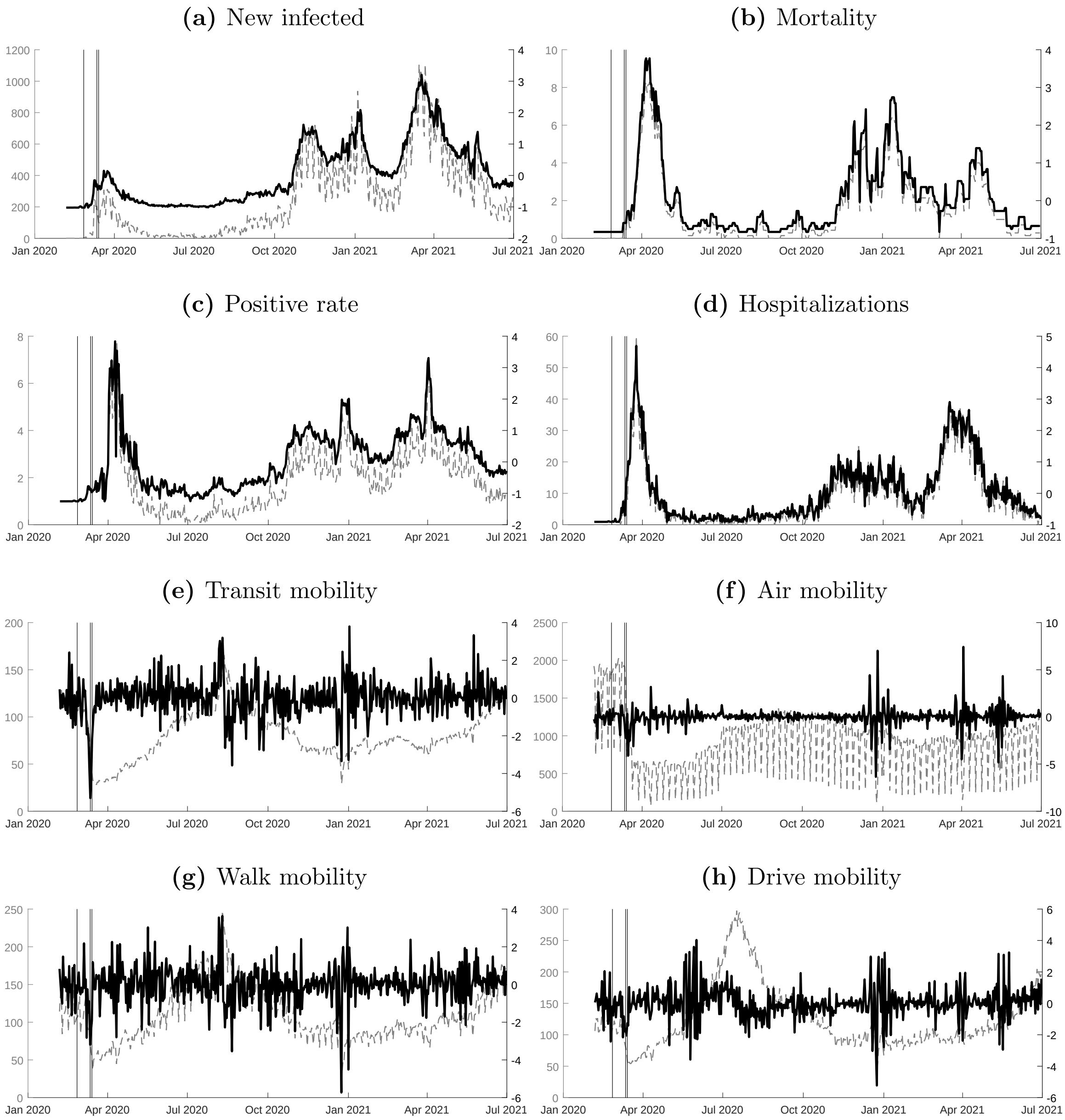

Figure C.1. Raw and transformed data. The raw data is reported on the left axis and with broken gray lines. The transformed data is normalized and reported on the right axis in black. See the text for details about transformation routines.

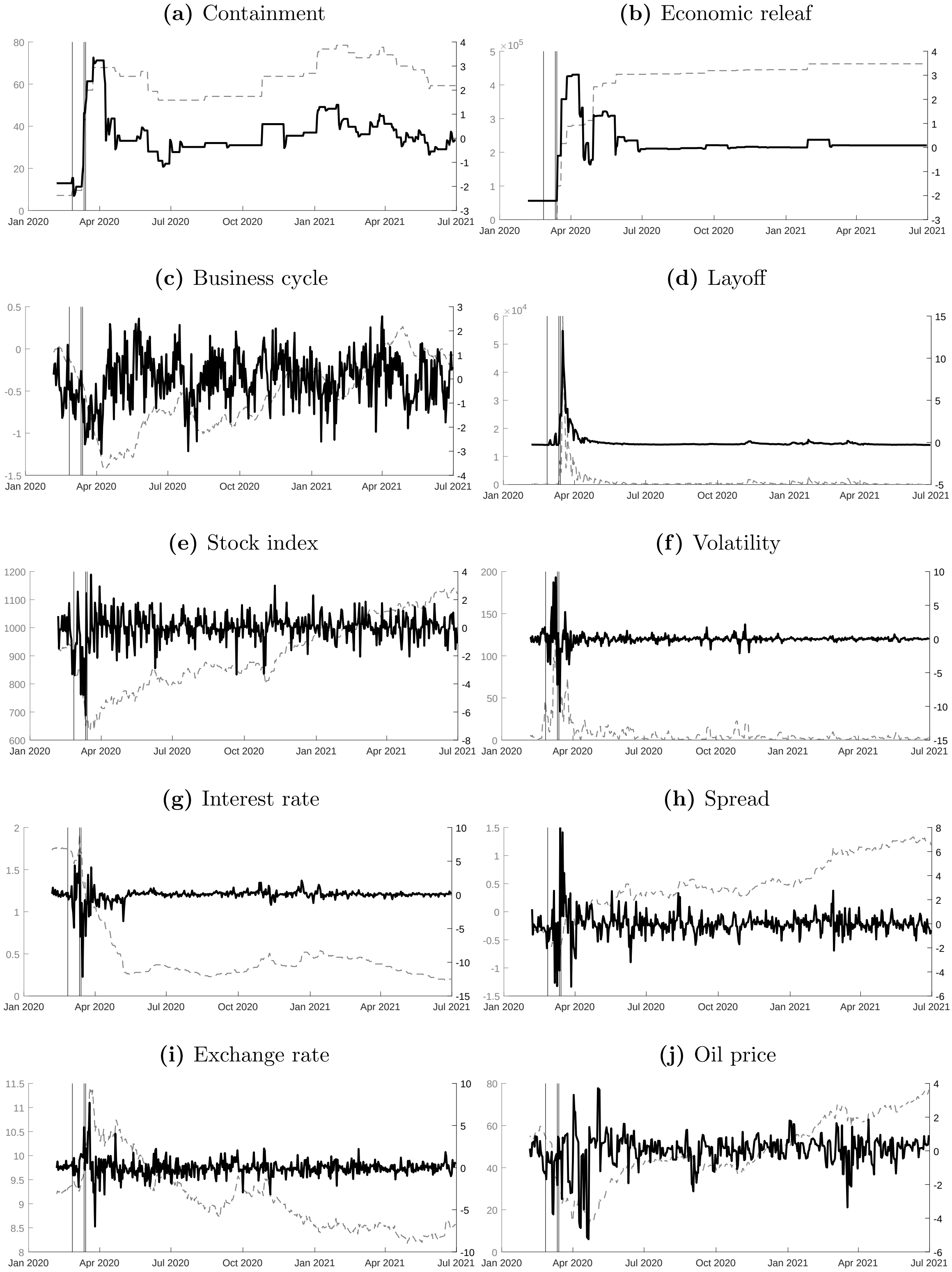

Figure C.2. Raw and transformed data. The raw data is reported on the left axis and with broken gray lines. The transformed data is normalized and reported on the right axis in black. See the text for details about transformation routines.

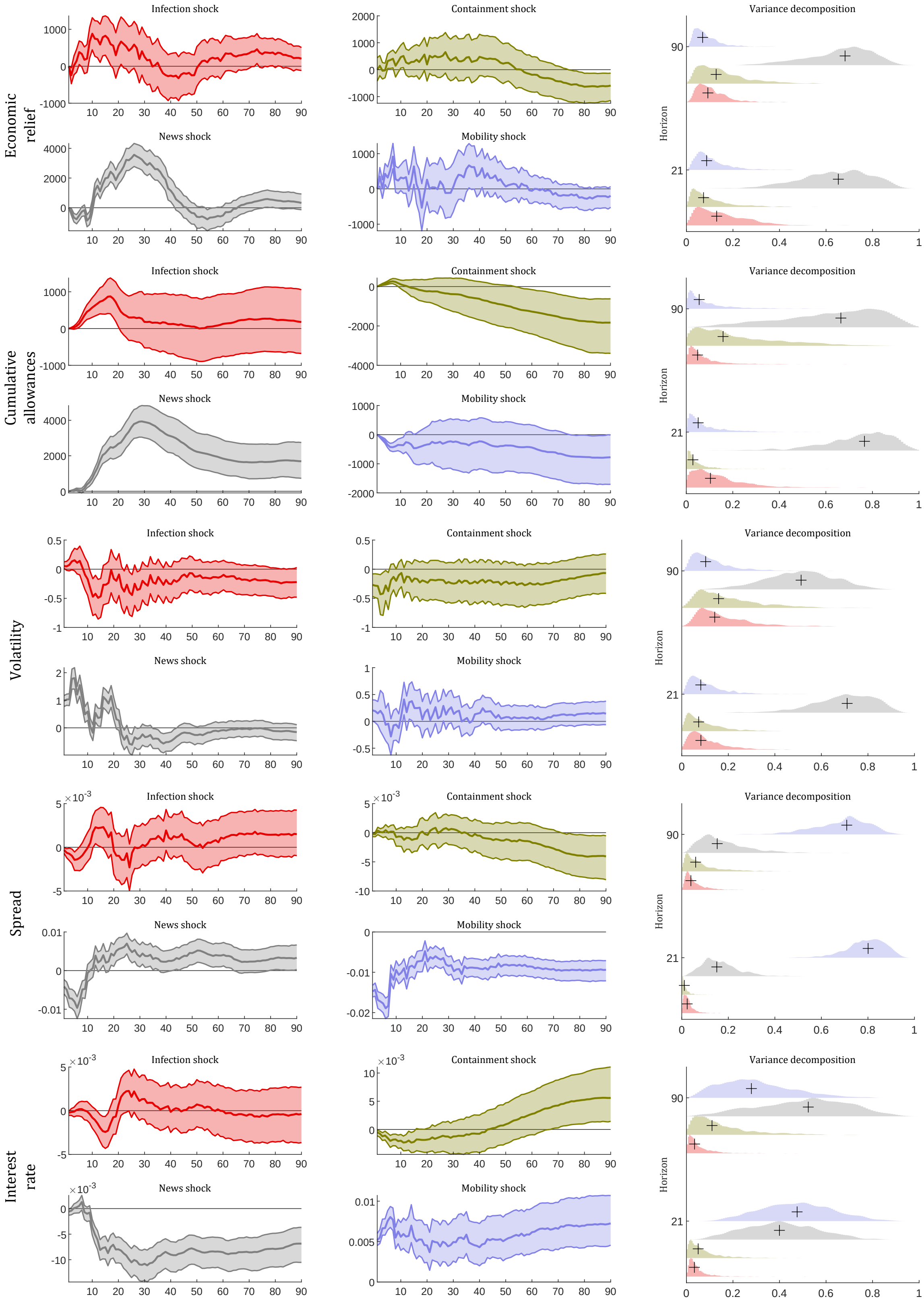

Figure C.3. Impulse responses and variance decompositions. Median response to a one standard deviation shock. Shaded areas correspond to 68 percent posterior probability bands. The variance decompositions report the full distribution, together with the median (+), for two different horizons.

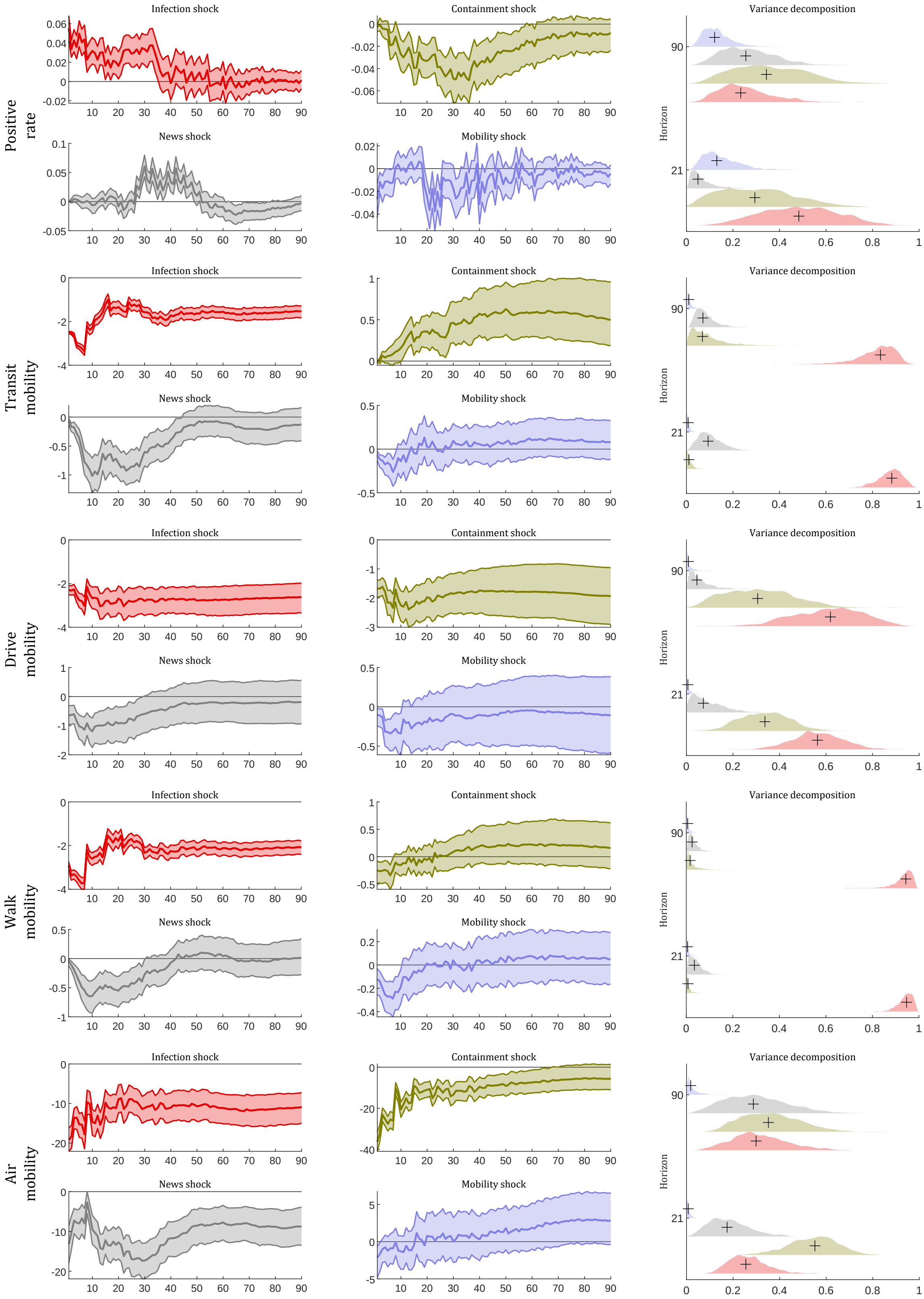

Figure C.4. Impulse responses and variance decompositions. Median response to a one standard deviation shock. Shaded areas correspond to 68 percent posterior probability bands. The variance decompositions report the full distribution, together with the median (+), for two different horizons.

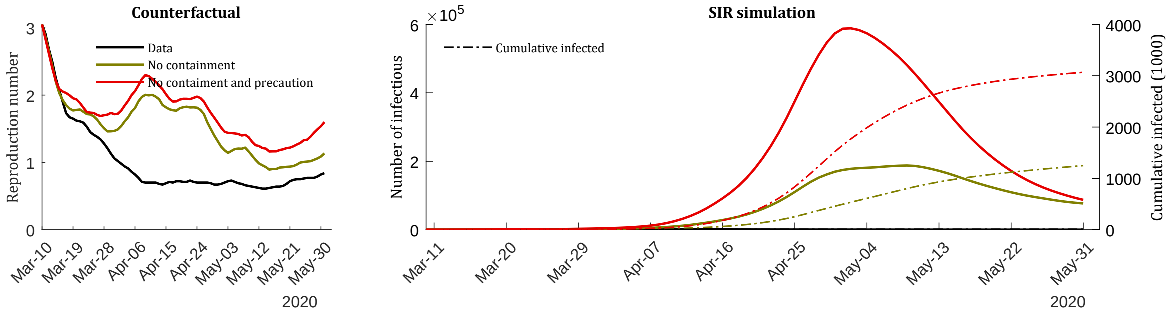

Figure C.5. SIR simulations. The graph to the left reports the reproduction number in Norway (from Arroyo-Marioli et al. (2021)) together with two counterfactual estimates; Without any systematic containment policies and without systematic containment policies and precautionary behavior. The graph to the right reports the resulting pandemic curves and cumulative number of infected when a standard SIR model is simulated using these time-varying reproduction numbers. See the text for further explanations.

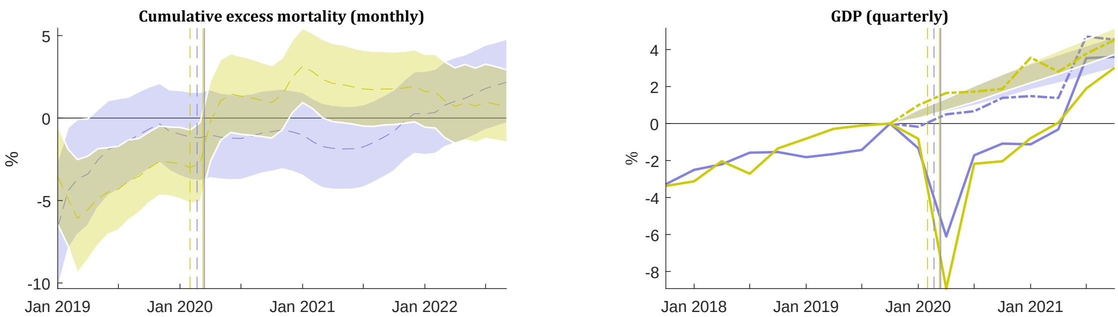

Figure C.6. Excess mortality and GDP growth. The left graph reports the monthly excess mortality in Norway and Sweden computed as described in Appendix B.2. The graph to the right reports the quarterly (log) level of GDP in Norway and Sweden, normalized to zero in 2019:Q4. Solid lines report actual outcomes, while broken lines report outcomes adjusted for a common international factor (as described in Appendix B.3). Shaded areas represent simple AR(2) projections of future (log) GDP growth, conditional on 2019:Q4 information.

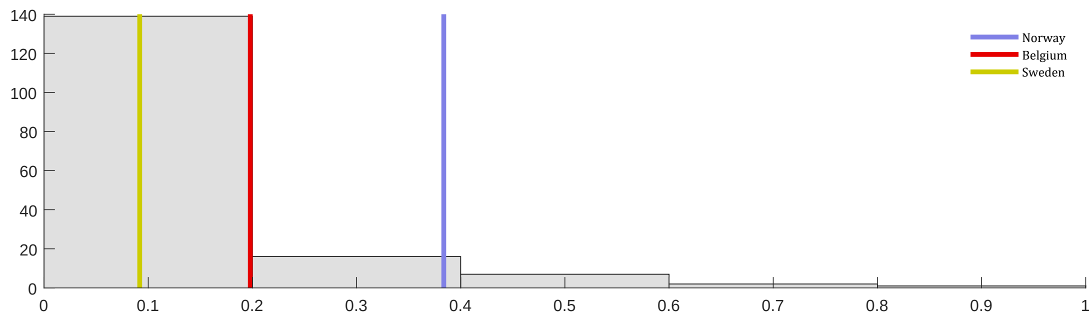

Figure C.7. Share of reported cases. Estimates for the true number of cases are mean estimates sources from the Institute for Health Metrics and Evaluation (IHME). IHME estimates the true number of COVID-19 infections for 165 countries and territories. Data on recorded cases are sourced from the "WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020. Available online: https://covid19.who.int/".

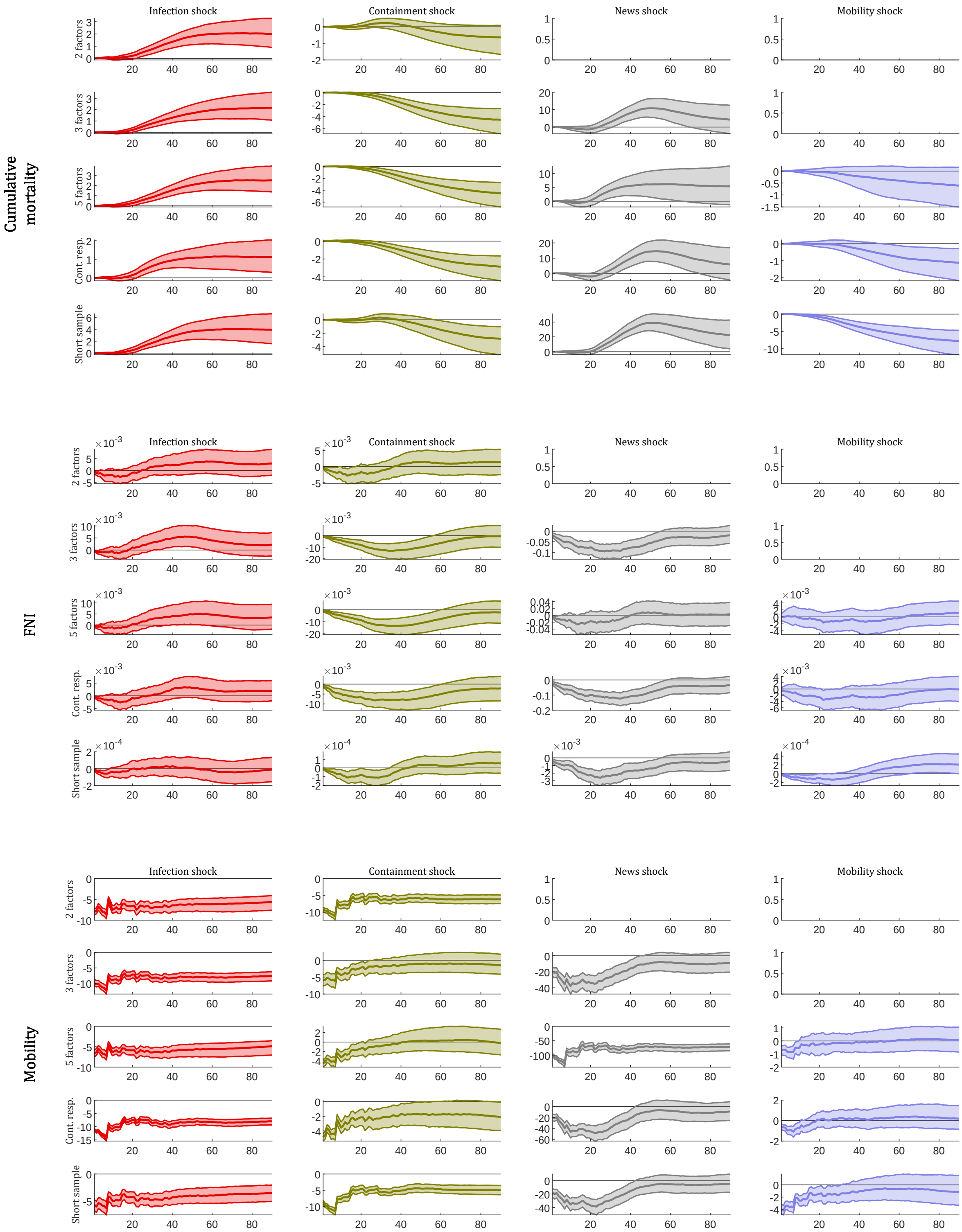

Figure C.8. Impulse responses and variance decompositions. Median response to a one standard deviation shock. Shaded areas correspond to 68 percent posterior probability bands. The variance decompositions report the full distribution, together with the median (+), for two different horizons.