UNIVERSITY of York

Discussion Papers in Economics

No. 21/06

Predicting the COVID-19 epidemic: is a regional approach preferable?

Laura Coroneo, Fabrizio Iacone, Giancarlo Manzi, and Silvia Salini

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Predicting the COVID-19 epidemic: Is a regional approach preferable?*

Laura Coroneo[1], Fabrizio Iacone[†2,1], Giancarlo Manzi[2], and Silvia Salini[2]

[1]University of York

[2]Università degli Studi di Milano

19th August 2021

Abstract

We use a SIRD model to predict the dynamics of the COVID-19 epidemic in the Italian regions at 1 to 4 weeks ahead. Out of sample forecasting results indicate that national forecasts obtained by aggregating regional forecasts are more accurate than predictions from a national model. These results suggest that national health authorities should take into account the level of heterogeneity across regions when predicting the spread of a national epidemic.

JEL classification codes: C12; C53; I18.

Keywords: Forecasting, Aggregation, Forecast evaluation, Epidemic.

*We thank Giuseppe Gerardi DEMM Data Management (University of Milan) for its availability and support

†Corresponding author: Department of Economics and Quantitative Methods, Universitá degli Studi di Milano, Via Conservatorio 7, 20122 Milano, Italy. Email: fabrizio.iacone@unimi.it.

1 Introduction

Accurate forecasts for the diffusion of a pandemic are crucial to inform public health interventions, and their effectiveness. During a pandemic, public health authorities are called to take decisions such as diverting, postponing or cancelling elective surgeries, imposing social distancing, school closures or lockdowns. These measures can contain and reduce the spread of the infection, but they create care backlogs, and are costly for the economy and for the mental health of the population. It is thus important that the timing and intensity of these actions are based on accurate forecasts, to balance effectively their benefits and costs.

Mechanistic compartment models, such as the SIR (Susceptible-Infectious-Recovered) model introduced by Kermack and McKendrick (1927), are the standard tools in infectious disease epidemiology to understand the dynamics of epidemics and predict their evolution. In this type of models, people within a population move between different compartments as a pathogen spreads from person to person. For example, in the standard SIR model people move between being "Susceptible", "Infected", and "Recovered". This implies that these models can produce realistic forecasts, even with a long term horizon, and for this reason SIR-type models have been routinely used to predict the COVID-19 pandemic.

Predictions from SIR-type models are obtained by first fitting the model on the country/region of interest, and then projections are generated from the estimated model. This is what in the forecasting literature is referred to as "direct" approach. However, one could alternatively use an "indirect" approach, where a SIR-type model is fitted to each of the components/sub-units and then a forecasts is obtained by aggregation of the forecasts of each component/sub-units. Heuristically, the direct approach may be seen as more parsimonious, but it has the drawback of imposing the same parametric model for all the components. Thus, there is a trade-off in terms of variance and bias, and it is an empirical question which of the two approaches is preferable for each case.

In this paper, we compare the predictive ability of a national model for the spread of the

COVID-19 epidemic in Italy to projections obtained by aggregating forecast from models fitted at a regional level. Italy was hit severely by the pandemic, and the spread of the epidemic was heterogeneous across regions. The response of the Italian health authorities was modulated across the regions, matching this diversity. The structure of the Italian health system, that contemplates a strong governance devolution to the regional authorities, may have facilitated this heterogeneity, occasionally resulting in very different regional policies adopted in face of the pandemic. These considerations suggest the possible presence of a degree of diversity across regions that cannot be captured by a model at the national level, implying that an indirect approach to forecasting the evolution of the pandemic might be more accurate than the standard direct one.

We forecast the evolution of the COVID-19 epidemic in Italy over the period October 18th, 2020 to January 31st, 2021 by using SIRD (Susceptible-Infectious-Recovered-Deceased) compartment models at national and regional levels. We then compare the predictive accuracy of the predictions of these two approaches using the Diebold and Mariano (1995) test of equal predictive accuracy with fixed-smoothing asymptotics, as recently proposed by Coroneo and Iacone (2020) to overcome the small sample size distortions of the standard test.

Our results indicate that the indirect approach of generating a national forecast by aggregating regional forecasts delivers more accurate predictions for the spread of the COVID-19 epidemic in Italy than directly forecasting the national series. We conjecture that this is primarily due to the heterogeneity of the dynamics and interventions across regions, that is amplified by the exponential nature of the SIRD models and also by the fact that movement across regions was restricted, reducing the correlation of regional shocks. Overall, this evidence caution public health authorities to consider the level of heterogeneity across regions when producing projection of the national spread of the epidemic. In the presence of substantial heterogeneity across regions, public health decisions should not be based on standard projection of the national spread of the epidemic obtained from models fitted at national level. A better

approach in this case is to fit a model for each region and then obtain the national prediction by aggregating the regional ones.

These results are in line with the economics literature, where the prevalent finding is that the indirect approach provides more accurate predictions (see Rose 1977, Tiao and Guttman 1980, Kohn 1982, Lütkepohl 1984, Marcellino, Stock and Watson 2003); however some evidence to the contrary also exists (see Hubrich 2005, Benalal, Diaz del Hoyo, Landau, Roma and Skudelny 2004). In epidemiology, very few analyses using regional aggregation have been presented in the literature. One example is in Ben-Nun, Riley, Turtle, Bacon and Riley (2018) who calculated the US influenza incidence forecast as the weighted sum of some regional profiles, with the weights given by the relative populations of US regions. Therefore, to the best of our knowledge, this is the first analysis formally comparing the predictive accuracy of national SIR-type models to projections obtained by aggregating predictions from regional SIR-type models.

The rest of the paper is organised as follows: in Section 2 we review the literature on forecasting using the indirect approach; in Section 3 we present the data. We discuss the model and its estimation in Section 4 and in Section 5, respectively. The forecast evaluation is introduced in Section 6 and it is discussed in Section 7. Conclusions are in Section 8.

2 Literature review on the indirect approach to forecasting aggregated time series

The indirect approach of generating forecasts by aggregating its components has a long tradition in econometrics. Rose (1977), Tiao and Guttman (1980) and Kohn (1982), among others, discussed the advantage of aggregating the forecasts of individual components. These early contributions are mainly theoretical, and consider the case in which the Data Generating Process (DGP) is known. Lütkepohl (1984), however, showed that the superiority of the

indirect approach is no longer warranted when the DGP is not known and the model has to be selected and the parameters estimated. Thus, in practice a decision regarding this question can only be taken on a case by case basis after a comparison of the performances of the two approaches with real data.

In an empirical study in which the aggregated consumption and investment series are considered against their components, Lütkepohl (1984) found that the indirect approach still delivered the best forecasts. This finding that aggregating individual forecast gives better performance seems to be prevalent in the economic literature, but evidence to the contrary also exists. Hubrich (2005) found that forecasts generated by means of the indirect approach did not improve the year-on-year forecast of inflation 12 months ahead; Benalal et al. (2004) also found that forecasting the aggregate HCPI for the euro-area had a better root Mean Square Error (RMSE) performance.

In practice, the two main arguments in favour of the indirect approach are that in this way it is possible to make a specific model for each variable, which may induce a forecast bias, and that idiosyncratic shocks that cause forecast errors may cancel out as the forecasts are aggregated. On the other hand, when model and parameter uncertainties are taken into account, direct forecasting may be advantageous because it may provide a more parsimonious model. It is interesting that both Hubrich (2005) and Benalal et al. (2004) have a rather short sample, as their finding may suggest that in that case the advantage of using a parsimonious model is stronger.

In other studies, Bermingham and D'Agostino (2014) again considered forecasting inflation by looking at its components. One interesting aspect of this work is that one may conjecture that if disaggregating into components is beneficial for forecasting, then one could try and refine the disaggregation and consider as many components as possible. Considering disaggregation up to 164 items for the US and 32 for the euro-area, Bermingham and D'Agostino (2014) find that the indirect approach does indeed improve forecasts, although it is not clear if the

improvement is statistically significant.

Indirect forecasting over such wide range of components poses a problem in terms of dimensionality. Lütkepohl (1984) considered only two or three components, and modelled these with a VAR. However, when a large quantity of components is used, the number of parameters that must be estimated in a VAR becomes very large. Hubrich (2005), Bermingham and D'Agostino (2014) and Marcellino et al. (2003) all found that a VAR was overall outperformed by scalar autoregressions, thus providing support for the more parsimonious modelling approach. Indeed, Bermingham and D'Agostino (2014) and Marcellino et al. (2003) found that the aggregation of the scalar autoregressions performed well, even when compared against forecasting using a principal component approach.

In much of the empirical literature, the indirect approach was considered aggregating series in the economic sense. For example, in Lütkepohl (1984) the personal consumption expenditure series was decomposed as the sum of the expenditure for durable goods, services and nondurable goods; and Hubrich (2005) disaggregated inflation in five components: unprocessed food, processed food, industrial goods, energy and services prices. On the other hand, Marcellino et al. (2003) considered geographical aggregation instead, and forecasted HCPI inflation and other relevant macro variables for the Euro-area aggregating forecasts for the individual countries, finding that the indirect approach to forecasting yielded better results. In view of the geographical nature of the aggregation, our exercise seems closer to the one in Marcellino et al. (2003), but the basic units of interest in our models are individuals rather than prices of baskets of goods. This is novel in this literature, and it adds a level of uncertainty, because individuals may move across regional boundaries.

3 Data

We use weekly data on Covid-19 new cases, deceased and recovered people available at the national level and for each region in Italy from the Italian Presidency of the Council of

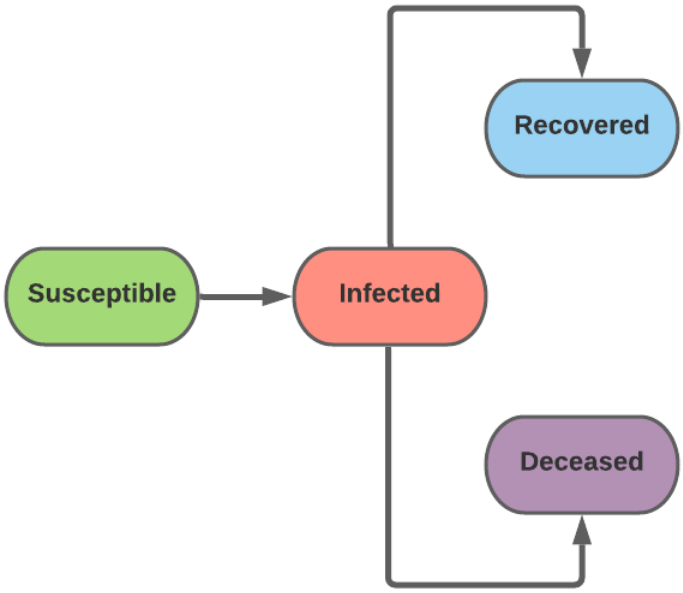
Ministers (PCM) - Department of Civil Protection Agency (CPA). Italy is constituted by 21 EU Nomenclature of Territorial Units for Statistics (NUTS) level 2 regions, as shown in Figure 1, that are very heterogeneous for size and population.

We focus our attention in particular on the so-called pandemic's "second wave", i.e. the period from October 2020 to January 2021. In particular, our sample for the estimation starts on August 31st, 2021, and we evaluate the forecasts on the period October 18th, 2020, until January 31st, 2021. This period seems particularly interesting and crucial, as it comes after the experiences learnt in the first wave in spring 2020, and before the vaccination campaign began in earnest, as the first vaccine jab in Italy was delivered on December 27th, 2020, but during the first month of the campaign only very few vaccinations, and solely for health personnel, have been administered.

During the second wave, the health authorities had the opportunity to use what was learnt during the first wave, and implement policies that could minimise damage to the country while waiting for the development and the rollover of the vaccine. The Italian government managed this period with a combination of national and regional measures with an alert system based on four colours characterising the level of alert for each region: white (very low restrictions), yellow (mild restrictions), orange (strong restrictions) and red (very high level of restrictions, comparable to that of a real lockdown). Sometimes restricted areas within regions had different alert colours with respect to the region colour. This system resulted in a widespread reduction of mobility between regions, if not an almost complete ban on inter-regional movements on the basis of the regional situations, with the intention to prevent the spread of the virus. This level of heterogeneity of interventions across regions, coupled with the mobility restrictions, provide a strong motivation for constructing national forecast aggregating forecasts from regional models rather than directly from a national one. In addition, we take into account the level of restriction using a stringency index computed since the beginning of the pandemic by the Covid-19 Response Tracker

Figure 1: Regions in Italy

Figure 2: SIRD model schematic



Research Group at the University of Oxford, United Kingdom (https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker).

4 Model

We model the dynamics of the COVID-19 epidemic in Italy and the 21 regions by means of a SIRD compartmental model. This is a generalisation of the classic SIR model where the population is divided into four different groups: susceptible individuals (S) who are healthy and can contract the disease; infected individuals (I); recovered individuals (R) who are immune to the disease; and deceased individuals (D). All the population starts in a single compartment, the susceptible compartment, and, with the evolution of the pandemic, can move to the compartment of the infected and then to the one of recovered people. See Figure 2 for a visual idea of the flows in a SIRD model.

Building on what recently proposed by Ferrari, Gerardi, Manzi, Micheletti, Nicolussi, Biganzoli and Salini (2021), the SIRD model we use to forecast the development of the pandemic is an

adjusted time-dependent SIRD model. Denoting S, I, R and D as the number of susceptible, infected, deceased and recovered people, respectively, the differential equations governing the SIRD model are the following:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\frac{\beta(t)S(t)I(t)}{n} \\
\frac{dI(t)}{dt} &= \left(\frac{\beta(t)S(t)}{n} - \gamma_R(t) - \gamma_D(t)\right) I(t) \\
\frac{dR(t)}{dt} &= \gamma_R(t)I(t) \\
\frac{dD(t)}{dt} &= \gamma_D(t)I(t)
\end{aligned}$$

subject to the constraint $S(t) + I(t) + R(t) + D(t) = n$ (where n is the total population). We don't take into account new births and deaths (not related to Covid-19) in the period considered, so n is considered constant. The parameters of interest, that govern the dynamics, are $\beta(t)$, $\gamma_R(t)$ and $\gamma_D(t)$: these are the time varying transmission rate, the recovery rate, and the mortality rate, respectively.

This model is deterministic in the sense that the dynamics ruling the evolution of the variables of interest $S(t)$, $I(t)$, $R(t)$, $D(t)$ are not random, nor they are subject to random fluctuations. Once the current values are observed, and given the parameters $\beta(t)$, $\gamma_R(t)$ and $\gamma_D(t)$, the trajectories of the variables of interest are certain. In practice, as $S(t)$, $I(t)$, $R(t)$, $D(t)$ are observable, the only uncertainty in this model comes from the three rates parameters, that are not constant and are not observed at time t.

Given that our data is observed at weekly frequency, we transform the previous system of

ODEs into a system of discrete time difference equations,

$$
\begin{aligned}
S_{t+1} - S_t &= -\frac{\beta_t S_t I_t}{n} && (1)\\
I_{t+1} - I_t &= \left(\frac{\beta_t S_t}{n} - \gamma_{R,t} - \gamma_{D,t}\right) I_t && (2)\\
R_{t+1} - R_t &= \gamma_{R,t} I_t && (3)\\
D_{t+1} - D_t &= \gamma_{D,t} I_t && (4)
\end{aligned}
$$

for S_t, I_t, R_t, D_t observed at week t. Given the pandemic rates β_t, $\gamma_{R,t}$ and $\gamma_{D,t}$, equations (1)-(4) can be used to forecast future values of the variables of interest. We thus now turn to discussing the estimation of the parameters. We start by rewriting equations (1)-(4) as

$$
\begin{aligned}
\beta_t &= \frac{n(S_{t+1} - S_t)}{S_t I_t} && (5)\\
\gamma_{R,t} &= \frac{R_{t+1} - R_t}{I_t} && (6)\\
\gamma_{D,t} &= \frac{D_{t+1} - D_t}{I_t} && (7)
\end{aligned}
$$

where equations (5)-(7), given observations up to time t, allow to compute the series of parameters $(\beta_1, ..., \beta_{t-1})'$, $(\gamma_{R,1}, ..., \gamma_{R,t-1})'$, $(\gamma_{D,1}, ..., \gamma_{D,t-1})'$; on the other hand, β_t, $\gamma_{R,t}$, $\gamma_{D,t}$ are not observable at time t, so these must be estimated. We assume a simple AR model for each parameter, and estimate the coefficients of the autoregression by minimising a ridge regression-type loss function:

$$
\hat{c}_{j,t} = argmin_{c_{j,t}} \left(\sum_{s=t-w}^{t-1} (\theta_s - \hat{\theta}_s)^2 - \lambda_t \sum_{j=0}^{J} c_{j,t}^2 \right) \quad (8)
$$

$$
\hat{\theta}_s = c_{0,t} + \sum_{j=1}^{J} c_{j,t} \theta_{s-j} \quad (9)
$$

where θ_t denotes each of the evolving parameters of the model, i.e. $\beta_t, \gamma_{R,t}, \gamma_{D,t}$, and $c_{0,t}$ and $c_{j,t}$ are the usual intercept and autoregression coefficients parameters. The penalty

term λ_t is applied to the sum of squares of the regression coefficients, so this is therefore a ridge regression regularization, based on a ℓ_2 norm. J and w and are the number of lags in the autoregression and the length of the rolling window in the estimation, respectively. We selected these using the Akaike Information criterion, fitting the AR model on data corresponding to observations from February 2020 up to September 2020. We considered pairs $w = 2J$ and selected $w = 10$ and $J = 5$ as they provided the best performance for most models.

We first obtain the estimates $\hat{\theta}_t$ from the AR part of the model, then regularize them using the shrunken $c_{j,t}$ coefficients from the ridge regression, see Hoerl and Kennard (1976). Ridge regularization alters the bias variance trade-off of the estimates, reducing the variance at the expense of increasing the bias. This reduces the overfitting and improves the reliability of the predictions. It has been widely used in applied times series econometrics, both in simulation studies comparing its performance versus other regularization methods, as for example in Inoue and Kilian (2008), and in applications, for example to improve the forecasts of urban traffic times, see Haworth, Shawe-Taylor, Cheng and Wang (2014). In our work, different penalty values have been used for λ_t in the ridge regressions to estimate the transmission, recovery and death rates. The values of λ_t for each parameter has been obtained using a K-fold cross-validation and picking the value of λ_t minimizing the mean square error.

Forecasts h periods ahead are computed iteratively using equations (2)-(4). For example, for the number of infected $\hat{I}_{t+h|t} = E(I_{t+h}|S_t, I_t, R_t, D_t)$ and, abbreviating the notation to $\hat{I}_{t+h}$, this is

$$\hat{I}_{t+h} = \hat{I}_{t+h-1} + \left(\frac{\hat{\beta}_{t+h-1}\hat{S}_{t+h-1}}{n} - \hat{\gamma}_{R,t+h-1} - \hat{\gamma}_{D,t+h-1} \right) \hat{I}_{t+h-1}$$

where $\hat{\beta}_{t+h-1}$, $\hat{\gamma}_{R,t+h-1}$, $\hat{\gamma}_{D,t+h-1}$ are obtained by iterating the $AR(J)$ projections, so for example for two periods ahead we compute $\hat{\theta}_{t+1} = \hat{c}_{0,t} + \hat{c}_{1,t}\hat{\theta}_t + \hat{c}_{2,t}\theta_{t-1}...$ where $\hat{c}_{0,t}, \hat{c}_{1,t}, ...$ are the estimates from (8)-(9).

To take into account the effect of the social distancing policies adopted in Italy, we multiply

the predictions of the transmission rate $\hat{\beta}_{t+h-1}$, for $h = 1, \ldots, 4$, by a factor $1 - s$, where s is the national average of the Italian Covid-19 stringency index, computed according to Hale (2021). The stringency index is a composite indicator formed by equally weighting 9 sub-indicators related to the measures adopted to contrast the pandemic, i.e. school closures, workplace closures, cancel public events, restrictions on gatherings, close public transport, public information campaigns, stay at home, restrictions on internal movement, international travel controls, testing policy, contact tracing, face covering, and vaccination policy. Each sub-indicator is formed by three to five levels of measure intensity. Data about the stringency index are available on the *Our World in Data* repository. For the forecasts in this exercise, we use the national sample mean over the period March 2020 to October 11, 2020, which the period before the starting date of the evaluation period for our out of sample exercise. The stringency index is also convenient to accommodate a well known shortcoming of the SIRD models, namely the tendency of these models to overpredict when there is a fast increase of the infections, and to underpredict when there is a sudden decrease, see Boguá, Pastor-Satorras and Vespignani (2003).

The Forecasting Algorithm

At each point in time t, variables S_1, ..., S_t, I_1, ..., I_t, R_1, ..., R_t, D_1, ..., D_t are observed, and we use the following algorithm to predict the evolution of these variables h steps ahead:

(i) Compute series of parameters $(\beta_1, ..., \beta_{t-1})'$, $(\gamma_{R,1}, ..., \gamma_{R,t-1})'$, $(\gamma_{D,1}, ..., \gamma_{D,t-1})'$ using equations (5)-(7).

(ii) Estimate parameters $\hat{c}_{0,t-1}$, ..., $\hat{c}_{J,t-1}$ for β_{t-1}, $\gamma_{R,t-1}$, $\gamma_{D,t-1}$, using (8).

(iii) Predict parameters $\hat{\beta}_t$, $\hat{\gamma}_{R,t}$, $\hat{\gamma}_{D,t}$ using (9), and scale parameter $\hat{\beta}_t$ using the stringency index.

(iv) Compute one-step ahead forecasts $\hat{I}_{t+1}$, $\hat{R}_{t+1}$, $\hat{D}_{t+1}$, using equations (2)-(4) and the

predicted parameters $\hat{\beta}_t$, $\hat{\gamma}_{R,t}$, $\hat{\gamma}_{D,t}$:

$$\begin{aligned}
\hat{I}_{t+1} &= I_t + \left(\frac{\hat{\beta}_t S_t}{n} - \hat{\gamma}_{R,t} - \hat{\gamma}_{D,t} \right) I_t \\
\hat{R}_{t+1} &= R_t + \hat{\gamma}_{R,t} I_t \\
\hat{D}_{t+1} &= D_t + \hat{\gamma}_{D,t} I_t
\end{aligned}$$

(v) Compute h-step ahead forecasts $\hat{I}_{t+h}$, $\hat{R}_{t+h}$, $\hat{D}_{t+h}$ in two steps:

1. Predict $\hat{\beta}_{t+h-1}$, $\hat{\gamma}_{R,t+h-1}$, $\hat{\gamma}_{D,t+h-1}$, using (9). For example, for $h = 2$,

$$\hat{\theta}_{t+1} = \hat{c}_{0,t+1} + \hat{c}_{1,t+1}\hat{\theta}_t + \sum_{j=2}^{J} \hat{c}_{j,t+1}\theta_{t-j}$$

and scale parameter $\hat{\beta}_{t+h-1}$ using the stringency index.

2. Compute the forecasts using (2)-(4). For example,

$$\hat{I}_{t+2} = \hat{I}_{t+1} + \left(\frac{\hat{\beta}_{t+1}\hat{S}_{t+1}}{n} - \hat{\gamma}_{R,t+1} - \hat{\gamma}_{D,t+1} \right) \hat{I}_{t+1}$$

5 Evolution of the Reproduction Numbers

Using a simplified definition, the basic reproduction number at time t is

$$R_{0,t} = \frac{\beta_t}{\gamma_{R,t} + \gamma_{D,t}},$$

with parameters β_t, $\gamma_{R,t}$ and $\gamma_{D,t}$ as defined in Equations (5) - (7) (for a more general definition considering multiple compartmental states, see Diekmann, Heesterbeek and Metz (1990)). The basic reproduction number is the number of people infected by one individual in a population in which all the individuals are susceptible. This number is of great importance

as it summarises the future dynamics of the pandemic, for example $R_{0,t} > 1$ means that the infection is spreading. A larger value of $R_{0,t}$ is also informative about the speed at which the pandemic is spreading, and it is therefore a primary object of interest for health authorities.

Given that in our simple model past values of β_t, $\gamma_{R,t}$ and $\gamma_{D,t}$ can be computed directly from the observations of S_t, I_t, R_t and D_t, and their past values, this is not a direct object of investigation in the forecast evaluation exercise. We nonetheless compute this number as a heuristic measure of the heterogeneity that is present at regional level, in contrast with a one-size-fits-all feature of the national model.

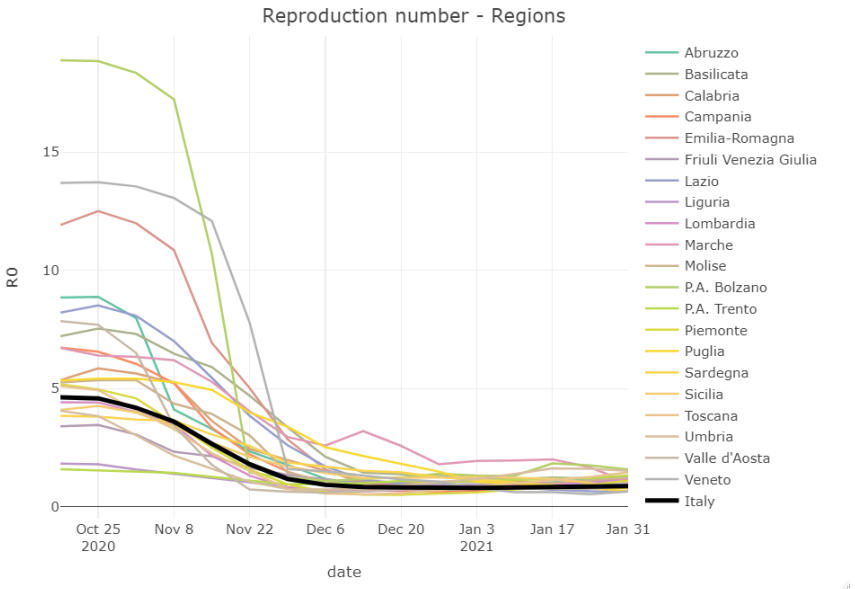
Figure 3 displays the values of the $R_{0,t}$ for the Italian regions compared with the value of the $R_{0,t}$ for Italy in the considered evaluation period. These plots document that the real behavior of the pandemic is mainly characterized by local infection surges and hotspots, and the national $R_{0,t}$ does not capture all the regional variability, especially in the first part of the sample. The dispersion of the values of $R_{0,t}$ drops in the second part of the sample, but a certain variation in the regional $R_{0,t}$ remains even for that period. As a result of this heterogeneity across regions, the indirect approach may prove more effective at forecasting the epidemic. This will be tested formally in Section 7.

6 Forecast Evaluation

We assess the out of sample forecasting performance using the Diebold and Mariano test of equal predictive accuracy (see Diebold and Mariano 1995, Giacomini and White 2006). To obtain reliable size properties, we apply fixed-smoothing asymptotics, as recently proposed for this test by Coroneo and Iacone (2020).

We denote by $e^{(j)}_{t|t-h}$ the forecast error at time t for forecast j formulated h periods ahead, where $j = 1$ refers to the indirect forecast obtained aggregating regional forecasts into a national one, and $j = 2$ to the direct forecast from a national model. Given a user-chosen

Figure 3: Reproduction numbers for Italy and its regions, by week



loss function $L(.)$, for example a squared loss, we define $d_{t,h} = L\left(e^{(1)}_{t|t-h}\right) - L\left(e^{(2)}_{t|t-h}\right)$ and $\overline{d_h} = \frac{1}{T}\sum_{t=1}^{T} d_{t,h}$. Under regularity conditions, such as that $e^{(j)}_{t|t-h}$ is mixing with sufficient rate, $L(.)$ continuous, and $E(d_{t,h}) = d$ for all t, then a standard CLT gives $\sqrt{T}(\overline{d_h} - d) \to_d N(0, \sigma^2)$, where σ^2 is the long run variance of $d_{t,h}$. Under the hypothesis of equal predictive ability, then $d = 0$. To test this hypothesis we need to standardise $\sqrt{T}(\overline{d_h} - d)$, as σ is unknown, however, we need an estimate, say $\widehat{\sigma}$.

A commonly used approach is to use a weighted covariance estimate with Bartlett kernel, $\widehat{\sigma}^2 = \widehat{\gamma}_0 + 2\sum_{l=1}^{M}(1 - l/M)\widehat{\gamma}_l$, where $\widehat{\gamma}_l$ is the lag-l sample autocovariance of the loss differential $d_{t,h}$. As long as this estimate is consistent, the limit distribution of $\sqrt{T}\frac{(\overline{d_h} - d)}{\widehat{\sigma}}$ remains a standard normal. The standard DM test is particularly unreliable when only a few out-of-sample observations are available, as the test is heavily size distorted, especially for multi-step forecasts, Clark and McCracken (2013). However, Sun (2014) showed that the approximation in the limit distribution may be improved upon by taking the ratio $b = M/T$ fixed. In this case, the limit distribution of $\sqrt{T}\frac{(\overline{d_h} - d)}{\widehat{\sigma}}$ is not normal, but it may be tabulated, see Kiefer and Vogelsang (2005). Thus, using fixed-b asymptotics we can test the null hypothesis of

equal predictive ability as $H_0 : E(d_{t,h}) = 0$ using the statistic $|\sqrt{T}\frac{\overline{d_h}}{\widehat{\sigma}}|$ and checking if its realisation exceeds the fixed-b critical value tabulated in Kiefer and Vogelsang (2005).

Coroneo and Iacone (2020) analyze the size and power properties of the test of equal predictive accuracy using fixed-b asymptotics in an environment with asymptotically non-vanishing estimation uncertainty, as in Giacomini and White (2006). Results indicate that the test with fixed-b asymptotics delivers correctly sized predictive accuracy tests for correlated loss differentials even in small samples, and that the power of these tests mimics the size-adjusted power. Considering size control and power loss in a Monte Carlo study, they recommend the bandwidth $M = \lfloor T^{1/2} \rfloor$ for the weighted autocovariance estimate of the long-run variance using the Bartlett kernel (where $\lfloor \cdot \rfloor$ denotes the integer part of a number).

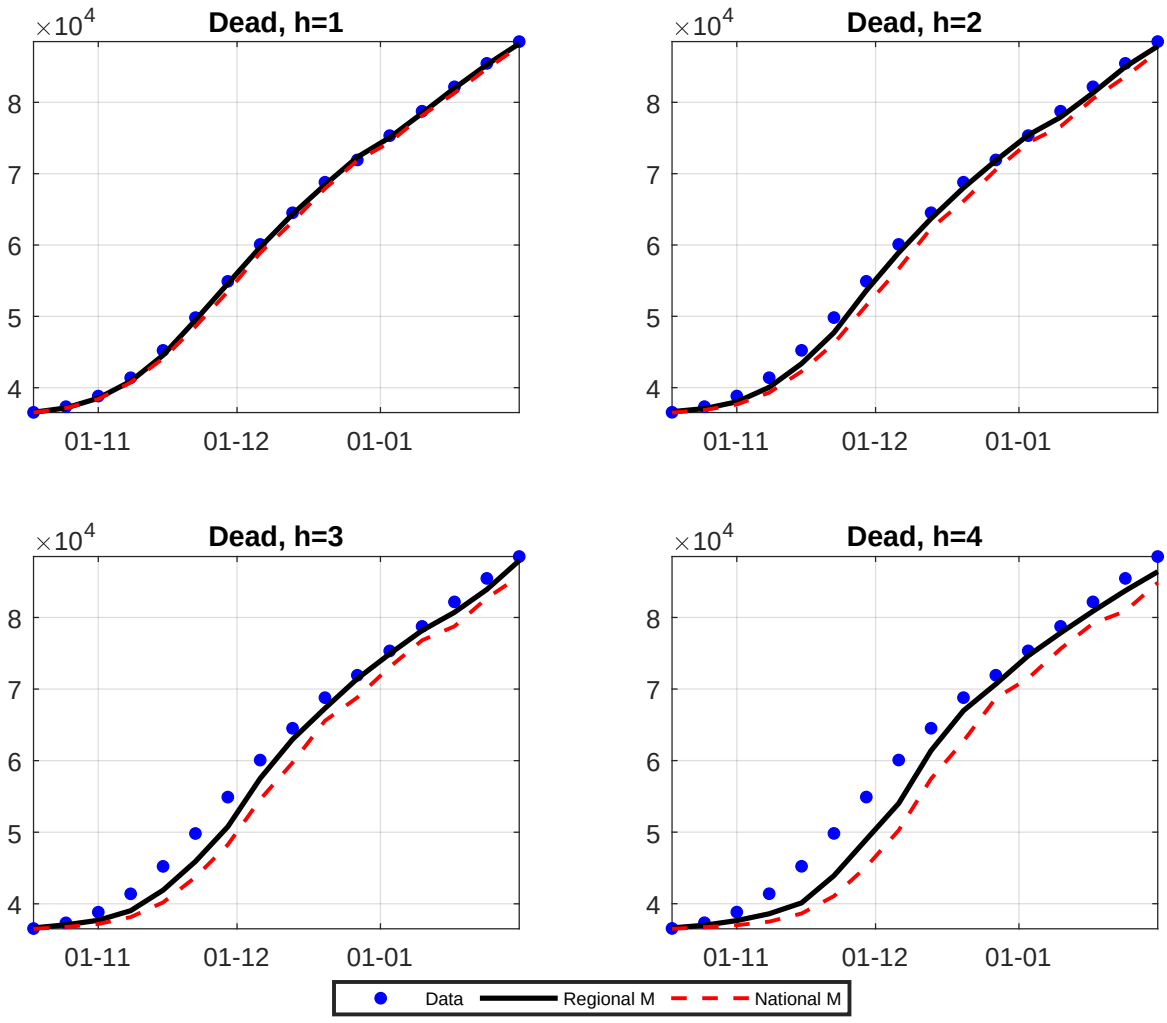
7 Forecasting Results

We compare out of sample predictions at 1 to 4 weeks ahead for the number of infections, the cumulative number of recovered and the cumulative number of deaths obtained directly from the national SIRD model and forecasts obtained indirectly aggregating regional level predictions obtained by fitting regional SIRD models. Our evaluation period is 18 October 2020 to 31 January 2021, for a total of 16 weekly observation.

In Figures 4-6 we report the national forecasts at 1 to 4 weeks ahead obtained using the indirect and the direct approach, along with the realised values. Summary statistics of the forecast errors are in Table 1. On balance, both approaches underpredict the number of deaths and recovered, resulting in a positive average forecast error. For the number of infected, instead, we can see a marked difference between the direct and the indirect forecasts, with the direct forecast undepredicting most of the times, and the indirect forecasts overpredicting the number of infected in the second half of the evaluation period.

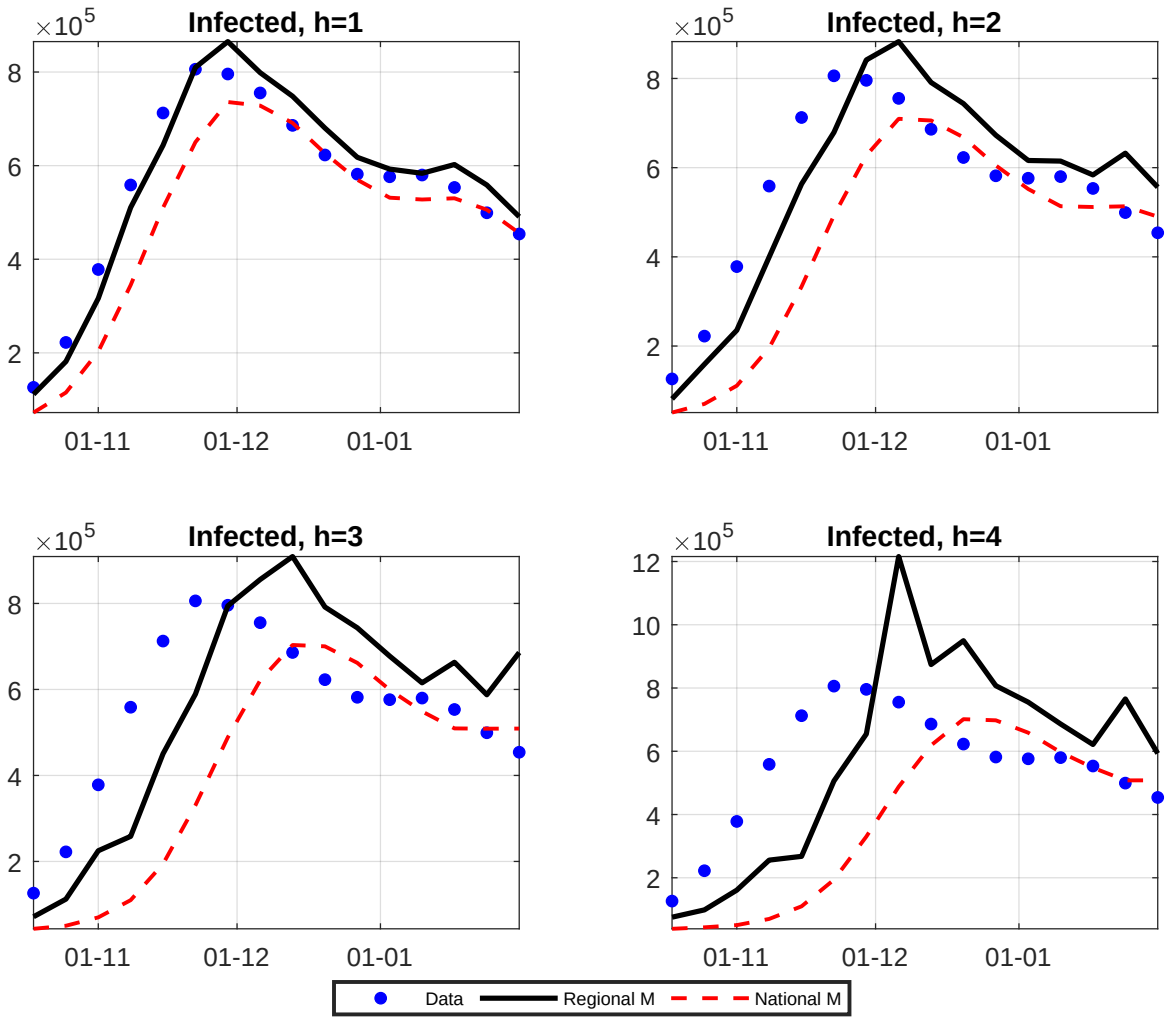
Interestingly, it seems that the main weakness of both approaches is that they tend to predict

Figure 4: Cumulative deaths in Italy, observed vs. forecasts



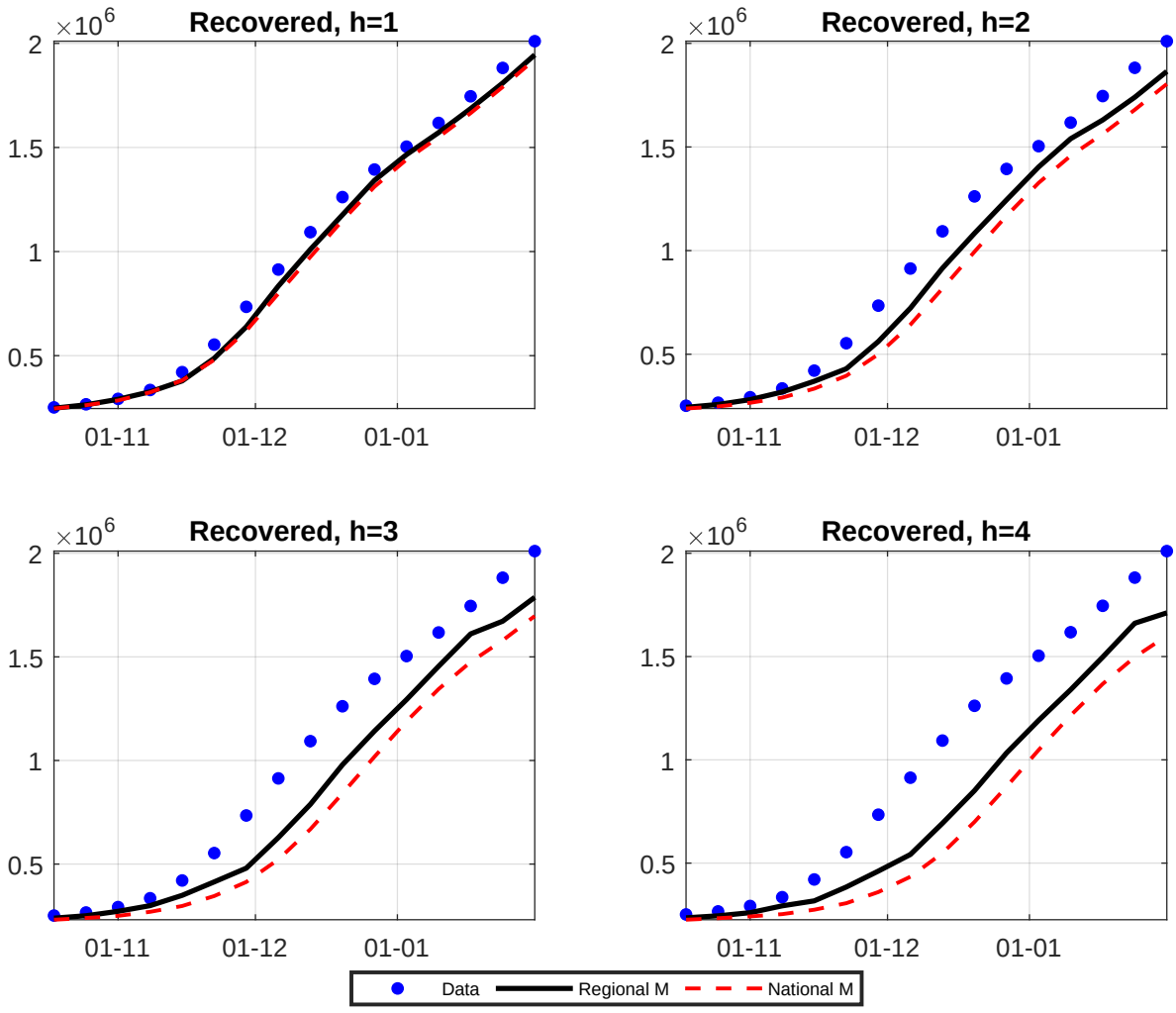
Note: Realised values for fatalities and forecasts at 1-week ahead (top left), 2-week ahead (top right), 3-week ahead (bottom left) and 4-week ahead (bottom right). Blue dots denote realised values, the black line indicates the national forecast obtained aggregating all the regional forecasts (indirect approach), and the dashed red line indicates the forecast from a national level model (direct approach).

Figure 5: Cumulative infected in Italy, observed vs. forecasts



Note: Realised values for infections and forecasts at 1-week ahead (top left), 2-week ahead (top right), 3-week ahead (bottom left) and 4-week ahead (bottom right). Blue dots denote realised values, the black line indicates the national forecast obtained aggregating all the regional forecasts (indirect approach), and the dashed red line indicates the forecast from a national level model (direct approach).

Figure 6: Recovered in Italy, observed vs. forecasts



Note: Realised values for recovered and forecasts at 1-week ahead (top left), 2-week ahead (top right), 3-week ahead (bottom left) and 4-week ahead (bottom right). Blue dots denote realised values, the black line indicates the national forecast obtained aggregating all the regional forecasts (indirect approach), and the dashed red line indicates the forecast from a national level model (direct approach).

Table 1: Summary statistics Forecast Errors

Dead

Model	h	Mean	Median	Std	rMSE	AC1	AC2	AC3	AC4
Reg	1	256	285	224	340	0.129	0.104	-0.092	-0.134
Reg	2	825	813	613	1028	0.701	0.233	-0.096	-0.354
Reg	3	1605	1477	1273	2049	0.790	0.378	-0.065	-0.463
Reg	4	2511	1770	2038	3234	0.836	0.456	-0.052	-0.535
Nat	1	779	815	417	884	0.562	0.336	-0.087	-0.530
Nat	2	1968	1982	1020	2217	0.729	0.408	-0.095	-0.505
Nat	3	3317	3166	1829	3788	0.840	0.467	-0.077	-0.606
Nat	4	4742	3865	2912	5565	0.845	0.515	-0.043	-0.579

Infected

Model	h	Mean	Median	Std	rMSE	AC1	AC2	AC3	AC4
Reg	1	-12715	-26225	45360	47108	0.754	0.454	0.220	-0.105
Reg	2	-9048	-37303	103924	104317	0.825	0.540	0.257	-0.027
Reg	3	-7633	-61880	166407	166582	0.840	0.610	0.250	0.027
Reg	4	-23695	-87223	249850	250971	0.689	0.489	0.140	-0.042
Nat	1	69395	48523	74530	101835	0.852	0.518	0.167	-0.085
Nat	2	109872	56092	141933	179490	0.869	0.541	0.189	-0.075
Nat	3	141235	63452	199984	244828	0.881	0.577	0.229	-0.058
Nat	4	171675	77422	246897	300716	0.894	0.618	0.277	-0.031

Recovered

Model	h	Mean	Median	Std	MSE	AC1	AC2	AC3	AC4
Reg	1	50091	55596	30610	58704	0.840	0.538	0.130	-0.421
Reg	2	104396	119322	64704	122821	0.888	0.600	0.228	-0.223
Reg	3	163414	186504	101333	192282	0.903	0.673	0.352	-0.066
Reg	4	222462	260647	136344	260920	0.932	0.768	0.497	0.119
Nat	1	67206	77049	40273	78349	0.899	0.646	0.275	-0.218
Nat	2	159383	179829	90094	183084	0.930	0.723	0.402	-0.035
Nat	3	243272	287781	139124	280244	0.944	0.777	0.507	0.132
Nat	4	318779	381352	186036	369092	0.956	0.820	0.596	0.282

This table reports summary statistics for the Forecast errors for the predictions of Infected, Recovered and Dead. Results from the aggregation at national level of regional models are summarised in the rows with Model set as Reg; results from the National model are summarised in the rows with Model set as Nat. The forecast horizon is denoted by h. rMSE is the root of the mean squared error, AC(1), AC(2), AC(3), AC(4) report the sample autocorrelations at the specified lag.

the number of infected with some lag. Given that infections were rising sharply in the first part of the period under scrutiny, this resulted in underestimating the number of recovered and deaths in subsequent weeks. The indirect approach does a better job at keeping the pace of the fast change of infections, but at the price of overshooting the peak of November 2021, with the error becoming particularly severe as the forecast horizon is increased, due to the exponential nature of SIRD models. This might be explained by the fact that the rise of infections in October 2021 prompted a quick response from the Health Authorities, so the observed overshoot might well indicate that without those measures the infections would have spread even further.

The better ability of the indirect forecasts to keep the pace of the fast change of infections in the first part of the evaluation sample is due to the fact that the infections picked up in autumn with different speed across the regions as shown in Figure 3. Such heterogeneous dynamics are better approximated by a regions-specific model, rather than with the "one-size-fits-for-all" approach of the national model.

The indirect approach also performed better in terms of dispersion of the forecast errors (except for the prediction of infected cases four periods ahead), as it is expected when the errors of the regional models composing the indirect forecast are idiosyncratic; finally, we also observe that the forecast errors from the indirect approach are less persistent (and therefore less predictable), especially for forecasts one or two periods ahead. Overall, all these empirical regularities points towards the indirect approach to forecasting as being better.

The outcome of the tests of equal predictive ability of the direct and indirect predictions are in Table 2. The test statistic is computed using a quadratic loss function and as described in Section 6, therefore a negative value means that the indirect forecasts constructed from regional SIRD models are more accurate that forecasts produced directly using a national SIRD model. Broadly speaking, the test statistic is negative and significant for the prediction of the number of recovered and dead at all forecasting horizons, signalling that the forecasts

Table 2: Predictive ability - Italy

Horizon	Dead	Infected	Recovered
1	-3.055**	-1.369	-2.789**
2	-3.125**	-1.235	-3.086**
3	-3.010**	-1.090	-3.058**
4	-2.644**	-0.689	-2.995**

This table reports the Diebold and Mariano (1995) test for equal predictive ability using fixed-b asymptotics as in Coroneo and Iacone (2020). A positive value for the test statistic means that the national model is more accurate. The forecast horizons h are 1, 2, 3 and 4 weeks ahead. ** and * indicate respectively, two-sided significance at the 5% and 10% level using fixed-b asymptotics.

from the indirect approach are significantly more accurate. For the number of infections, the test statistic is still negative for all forecasting horizons, but not significant.

Overall, these results suggest a superior predictive ability for the indirect approach, primarily due to the heterogeneity of the dynamics across the Italian regions that were amplified by the regional specific interventions and the restrictions to the movement of people between regions.

8 Conclusion

We use a SIRD model to predict the the spread of the second wave of the COVID-19 epidemic in all the 21 Italian regions at 1 to 4 weeks ahead during the second wave of the epidemic. Out of sample forecasting results indicate that the indirect approach of generating a forecast by aggregating regional forecasts into a national one delivers more accurate predictions for the national spread of the epidemic than directly forecasting the national series.

These results suggest that national health authorities should take into account the level of

heterogeneity across regions when predicting the spread of a national epidemic. This because the standard approach of predicting the national spread of an epidemic from a national model might not produce the most accurate predictions in the presence of important heterogeneity across regions. A better approach in this case is to fit a model for each region and then obtain the national prediction by aggregating the regional ones.

References

Ben-Nun, Michal, Pete Riley, James Turtle, David Bacon, and Steven Riley (2018) 'Forecasting national and regional influenza- like illness for the usa.' *PLOS Computational Biology* 15(5), e1007013

Benalal, Nicholai, Juan Luis Diaz del Hoyo, Bettina Landau, Moreno Roma, and Frauke Skudelny (2004) 'To aggregate or not to aggregate? euro area inflation forecasting.' *Euro Area Inflation Forecasting (July 2004)*

Bermingham, Colin, and Antonello D'Agostino (2014) 'Understanding and forecasting aggregate and disaggregate price dynamics.' *Empirical Economics* 46, 765 – 788

Boguá, Marián, Romualdo Pastor-Satorras, and Alessandro Vespignani (2003) 'Epidemic spreading in complex networks with degree correlations.' In *Statistical Mechanics of Complex Networks,* ed. Romualdo Pastor-Satorras, Miguel Rubi, and Albert Diaz-Guilera (Springer) pp. 127–147

Clark, Todd, and Michael McCracken (2013) 'Advances in forecast evaluation.' *Handbook of economic forecasting* 2, 1107–1201

Coroneo, Laura, and Fabrizio Iacone (2020) 'Comparing predictive accuracy in small samples using fixed-smoothing asymptotics.' *Journal of Applied Econometrics* 35(4), 391–409

Diebold, Francis X, and Roberto S Mariano (1995) 'Comparing predictive accuracy.' *Journal of Business & Economic Statistics* pp. 253–263

Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz (1990) 'On the definition and the computation of the basic reproduction ratio ro in models for infectious diseases in heterogeneous populations.' *Journal of Mathematical Biology* 28, 365–382

Ferrari, Luisa, Giuseppe Gerardi, Giancarlo Manzi, Alessandra Micheletti, Federica Nicolussi, Elia Biganzoli, and Silvia Salini (2021) 'Modeling provincial covid-19 epidemic data

using an adjusted time-dependent sird model.' *International Journal of Environmental Research and Public Health* 18(12), 6563

Giacomini, Raffaella, and Halbert White (2006) 'Tests of conditional predictive ability.' *Econometrica* 74(6), 1545–1578

Hale, Thomas (2021) 'A global panel database of pandemic policies (oxford covid-19 government response tracker).' *Nature Human Behaviour* 5, 529–538

Haworth, James, John Shawe-Taylor, Tao Cheng, and Jiaqiu Wang (2014) 'Local online kernel ridge regression for forecasting of urban travel times.' *Transportation Research Part C* 46, 151–178

Hoerl, Arthur E., and Robert W. Kennard (1976) 'Ridge regression iterative estimation of the biasing parameter.' *Communications in Statistics - Theory and Methods* 5(1), 77–88

Hubrich, Kirstin (2005) 'Forecasting euro area inflation: Does aggregating forecasts by hicp component improve forecast accuracy?' *International Journal of Forecasting* 21(1), 119–136

Inoue, Atsushi, and Lutz Kilian (2008) 'How useful is bagging in forecasting economic time series? a case study of u.s. consumer price inflation.' *Journal of the American Statistical Association* 103(42), 511–522

Kermack, William Ogilvy, and Anderson G McKendrick (1927) 'A contribution to the mathematical theory of epidemics.' *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115(772), 700–721

Kiefer, Nicholas M, and Timothy J Vogelsang (2005) 'A new asymptotic theory for heteroskedasticity-autocorrelation robust tests.' *Econometric Theory* 21(6), 1130–1164

Kohn, Robert (1982) 'When is an aggregate of a time series efficiently forecast by its past?' *Journal of Econometrics* 18(3), 337–349

Lütkepohl, Helmut (1984) 'Forecasting contemporaneously aggregated vector arma processes.' *Journal of Business & Economic Statistics* 2(3), 201–214

Marcellino, Massimiliano, James H Stock, and Mark W Watson (2003) 'Macroeconomic forecasting in the euro area: Country specific versus area-wide information.' *European Economic Review* 47(1), 1–18

Rose, David E. (1977) 'Forecasting aggregates of independent arima processes.' *Journal of Econometrics* 5(3), 323–345

Sun, Yixiao (2014) 'Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference.' *Journal of Econometrics* 178, 659–677

Tiao, G.C., and Irwin Guttman (1980) 'Forecasting contemporal aggregates of multiple time series.' *Journal of Econometrics* 12(2), 219–230