Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Short-term Covid-19 forecast for latecomers

Marcelo C. Medeiros a,⁎, Alexandre Street b, Davi Valladão c, Gabriel Vasconcelos d, Eduardo Zilberman e,a

a Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil
b Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Brazil
c Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Brazil
d Bank of Communications - BBM/BOCOM, Brazil
e Gávea Investimentos, Brazil

A R T I C L E   I N F O

Keywords:
Covid-19
LASSO
Forecasting
Pandemics
Infectious diseases

A B S T R A C T

The number of new Covid-19 cases is still high in several countries, despite vaccination efforts. A number of countries are experiencing new and severe waves of infection. Therefore, the availability of reliable forecasts for the number of cases and deaths in the coming days is of fundamental importance. We propose a simple statistical method for short-term real-time forecasting of the number of Covid-19 cases and fatalities in countries that are latecomers—i.e., countries where cases of the disease started to appear some time after others. In particular, we propose a penalized LASSO regression model with an error correction mechanism to construct a model of a latecomer country in terms of other countries that were at a similar stage of the pandemic some days before. By tracking the number of cases in those countries, we use an adaptive rolling-window scheme to forecast the number of cases and deaths in the latecomer. We apply this methodology to 45 countries and we provide detailed results for four of them: Brazil, Chile, Mexico, and Portugal. We show that the methodology performs very well when compared to alternative methods. These forecasts aim to foster better short-run management of the healthcare system and can be applied not only to countries but also to different regions within a country. Finally, the modeling framework derived in the paper can be applied to other infectious diseases.

© 2021 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Accurate forecasts of the number of cases and deaths of infectious diseases, like Covid-19, in the short run (e.g., in the next few days or weeks) are crucial for managing healthcare systems properly. Depending on the expected pressure on the healthcare system, one can make more informed decisions on how to allocate hospital beds and ventilators, whether to setup more field hospitals, whether to train more healthcare workers, and so on.

Furthermore, accurate forecasting models can help trace the evolution of new variants of the disease and the effects of vaccination.

In this paper, we propose a statistical method to forecast in real time the very short-run evolution of the number of Covid-19 cases and deaths in countries that are latecomers. Given that these latecomers were hit by the Covid-19 pandemic after other countries, we can use information from these other countries when they were at a similar stage of the pandemic, a few days or weeks before. Here, a latecomer country refers to one where Covid-19 cases and deaths were identified after others. Note that the labeling of a given country as a latecomer depends on the number of countries we consider. We use the penalized least absolute selection and shrinkage
operator (LASSO) regression model proposed by Tibshirani (1996) to construct an error correction model (ECM) of a latecomer in terms of the other countries. The idea behind the ECM is to adjust the short-run dynamics of the latecomer to departures from the equilibrium (long-run relation) between the latecomer country and its peers. By tracking the number of cases in those countries, we can forecast the evolution of the disease in the latecomer a few days ahead. Forecasts of the number of deaths are constructed as a linear regression on the number of cases. As the pandemic evolves, one can run the model on a daily basis, and in an adaptive rolling-window scheme, to obtain updated forecasts for the next few days. The model is easily estimated and confidence intervals can be computed in a straightforward manner by simulation techniques.

An (adaptive) rolling-window scheme is important in order to acknowledge the dynamic nature of the pandemic and to attenuate the effects of outliers and potential structural breaks (due to, for example, more or less testing after a given period, policy changes, the start of vaccination campaigns, new variants of the virus, and changes in the relations between countries used as explanatory variables and the latecomer). Nonetheless, it is important to emphasize that despite this attenuation, one might expect worse forecasts a few days after a structural break as the model adapts. Hence, and needless to say, the use of the proposed forecasting method should be complemented by evaluations of how the pandemic is evolving.1

We apply the methodology to 45 different countries and present detailed results for four of them: Brazil, Chile, Mexico, and Portugal. The 45 countries were selected based on the following criteria: (1) there must be at least six countries ahead of it by 14 days or more; (2) it has a population of at least 10 million people, and (3) it had at least 30,000 reported cases of Covid-19 by the end of 2020. We show that our method performs very well at forecasting the out-of-sample number of cases and deaths up to the next 14 days. The number of cases correspond to those that are detected by the healthcare system, which is the proper measure to track if the concern is to evaluate the impact on its capacity.

Tracking the evolution of Covid-19 poses several challenges. The proposed method overcomes some of them. First, standard epidemiological models that are used to track the evolution of an epidemic are hard to discipline quantitatively to a new disease. Despite the enormous effort worldwide to understand transmission, recovery, and death rates, many parameters that need to calibrated remain uncertain Atkeson (2020), and the behavioral responses of individuals, as well as containment policies, affect these parameters (Eichenbaum, Rebelo, & Trabandt, 2020). By contrast, the proposed forecasting method has the advantage of being model-free, and makes projections based solely on available data.

Second, even if it were possible to discipline those epidemiological models reliably, they speak to the evolution of the infected population. From the perspective of managing healthcare resources, the relevant figure is the number of infected individuals that end up pressuring the healthcare system. Many individuals with Covid-19 are asymptomatic and do not need access to the healthcare system. Hence, the evolution of the virus among the sub-population that actually needs healthcare—a sub-population with certain characteristics that differ from the rest of the population—might be different from the evolution of the virus among the whole population. The proposed method avoids this problem by directly forecasting the number of infected individuals detected by the healthcare system or specific parts of it that are of interest, e.g., specific regions within a country.

Finally, alternative methods to track the evolution of Covid-19 and forecast the pressure on healthcare resources, such as mass testing, are expensive and unavailable in many countries. The methodology and our codes are immediately and cheaply reproducible to any latecomer that tracks the number of Covid-19 cases (and deaths). Note that the proposed methodology can also be applied to different regions within a country. This is particularly useful in large countries such as Brazil and India, where the disease hits distinct regions with delays.2

The aforementioned challenges are even harder to overcome in poor and developing countries, mostly latecomers, due to the lack of high-quality research, reliable data, and budget limitations. By tracking the very short-run evolution of the number of Covid-19 cases (and deaths) in real time, we hope to provide a methodology that can inform policymakers and the general public. In our view, an adaptive and accurate data-driven forecast is critical to foster better management of the healthcare system, especially in those countries that lack proper resources.

1.1. Main takeaways

The ECM method proposed in this paper provides forecasts of cases and deaths. In general, our method showed lower mean absolute percentage errors (MAPEs) than two benchmark models—namely, a simple quadratic trend regression (trend) that has been shown to be quite precise for short-term forecasts of Covid-19 fatalities (Coroneo, 2020).2 Epidemiologists and researchers from other fields rushed to improve those models and provide simulations on the spread of the disease, some of them taking into account counteracting policy and behavioral responses. A very incomplete list includes Bastos and Cajuheiro (2020), Berger, Herkenhoff, and Moneg (2020), Kucharski et al. (2020), Walker, Whittaker, Watson, et al. (2020), Wu, Leung, and Leung (2020).

2 We post daily updated forecasts for Brazil, along with methodology updates and codes. The domain is https://covid19analytics.com.br/, and one can check there for updates.
lacone, & Monteiro, 2020), and an integrated autoregressive model of order one (AR).

Over the 14 forecasting horizons, the ECM approach produced smaller MAPEs on average than the AR benchmark in 51.10% of the countries studied when the dependent variable was the number of Covid-19 cases, and in 65.50% in the case of deaths. Compared to the trend model, the percentages were 85.24% and 82.06% for cases and deaths, respectively. Compared with the AR model, the gains of the ECM alternative were more noticeable at forecasting horizons of more than two days. On the other hand, the ECM outperformed the trend benchmark at all horizons. Finally, the results improved remarkably when we considered a simple average combination of the ECM and AR forecasts.

Regarding the four countries for which we provide an in-depth analysis, the following results emerge. For Brazil, the ECM delivered MAPEs for Covid-19 cases between 0.44% (one day ahead) and 3.94% (14 days ahead), whereas the trend model MAPEs ranged between 0.76% and 4.86%. The AR benchmark MAPEs were between 0.38% and 6.40%. Although, the performance of the ECM was inferior to the AR model for one-day-ahead forecasts, it considerably outperformed the latter at horizons of two or more days. For Chile, the ECM yielded MAPEs between 0.32% and 3.99%, whereas the trend approach produced MAPEs between 0.48% and 4.09%. The MAPEs from the AR models ranged between 0.32% and 5.22%. The superiority of the proposed ECM in this clear in the case of Chile. For Mexico, our approach delivered MAPEs ranging from 0.23% to 2.41%, whereas the trend alternative yielded MAPEs between 0.41% and 2.78%. The MAPEs from the AR model were between 0.23% and 4.16%. For Portugal, the ECM delivered MAPEs between 0.25% and 5.18%, whereas the trend approach produced MAPEs between 0.74% and 6.83%. However, in this case, the AR performed best, with MAPEs between 0.22% and 4.17%. Nevertheless, when we considered a simple average of the forecasts from the ECM and AR models, the MAPEs reduced to 0.21% and 3.94%. Now, we analyze the results concerning deaths. For Brazil, the ECM produced MAPEs between 0.49% and 4.13%. The trend (AR) model yielded MAPEs between 0.83% (0.35%) and 4.96% (5.39%). For Chile the ECM delivered a minimum MAPE of 0.73% for one-day-ahead forecasts and a maximum of 5.77% at the two-week-ahead horizon. The trend benchmark had MAPEs between 1.44% and 10.60%. The minimum MAPE of the AR model was 0.58. The maximum MAPE was of the order of 10%, due to a major outlier. When this forecast was removed, the maximum MAPE was 8.25%. For Mexico, our model produced MAPEs between 0.61% and 3.17%. The trend method yielded MAPEs between 0.55% and 8.23%. The MAPEs of the AR benchmark were between 0.39% and 6.70%. For Portugal, the minimum (one-day-ahead) and maximum (14-day-ahead) MAPEs of the ECM were 0.30% and 4.14%, respectively. The trend model yielded MAPEs between 0.77% and 7.09%, whereas the AR alternative had MAPEs ranging from 0.14% to 3.59%. For Portugal, the AR benchmark was the model with the lowest MAPEs. However, as before, a simple average combination of the ECM and AR models reduced the MAPEs to 0.17% and 3.53%. Although the forecast combination was not able to beat the AR model for one-day-ahead forecasts, it outperformed the AR benchmark at horizons of more than four days.

Since the outbreak of the Covid-19 pandemic, ample research dealing with short-term forecasts of cases and deaths has been published in a wide collection of academic journals. The models range from different versions of epidemiological compartmental models to pure statistical and machine learning approaches. The models can be as simple as pure trend regression or as complicated as deep learning neural networks. Nevertheless, a number of studies provide strong evidence that it is quite difficult to beat the simplest alternatives. Our approach keeps the simplicity of several statistical models, explores equilibrium relations between a latecomer country and its peers, and shows robustness against many breaks in the dynamics of the series over the years 2020 and 2021. Coronel et al. (2020) compare the predictive accuracy of forecasts for the number of fatalities produced by several forecasting teams and collected by the United States Centers for Disease Control and Prevention (CDC) and a simple benchmark alternative. The set of models includes both statistical (dynamic growth models) and compartmental approaches (different versions of SEIRD models). The authors find that a simple quadratic trend regression outperforms all alternatives at horizons up to one week ahead. For longer horizons, some of the models outperform the simple benchmark. However, the authors show that an ensemble of models outperforms all the other alternatives. Similar quadratic trend models are also considered with some adjustments by Jiang, Zhao, and Sha (2020) and Li and Linton (2021). Due to the previous satisfactory performance of the quadratic trend model, we adopted it as a benchmark specification in our study.

Hendry (2022) consider a flexible trend model and apply it to forecasts of confirmed cases and deaths in a large number of countries. As with ours, their model has no epidemiological component. For confirmed cases, the authors report MAPEs between 0.4% and 2.1% for one-day-ahead forecasts and from 1.7% to 7.6% for four-day-ahead forecasts. In the case of death counts, the MAPEs are higher. Although their results are not directly comparable to ours, as we do not analyze the same countries, our MAPEs are lower than the ones reported by Hendry (2022). See also Petropoulos, Makridakis, and Stylianou (2022) for a similar approach to Hendry (2022).

Nonlinear machine learning methods, such as long short-term memory, deep neural networks, random forests, and support vector machines, have also been used to forecast cases and death counts in the short run. For example, Ribeiro, da Silva, Mariani, and Coelho (2020) estimate a vast amount of statistical and machine learning methods to forecast future cases in Brazil. However, their models are much more complex than the ones considered here, and their MAPEs range from 0.87%–3.51%, 1.02%–5.63%, and 0.95%–6.90% for one-, three-, and six-day-ahead forecasts, respectively. These MAPEs

---

4 We tried many ARIMA specifications and ARIMA(1,1,0) showed the best performance.
Due to the limited amount of data and the large dimension of \( \mathbf{x} \), as compared to the sample size, we use the least absolute and shrinkage operator (LASSO) to recover the parameter vector. The goal of the LASSO is to balance the trade-off between bias and variance. It is a useful tool to select the relevant peers in an environment with very few data points. Therefore, the estimator of the unknown parameter \( \beta_T \) in Eq. (2.2) is defined as

\[
\hat{\beta}_T = \arg \min_\beta \left( \frac{1}{K} \sum_{\tau = T-K+1}^{T} (y_{\tau} - \mathbf{x}_{\tau}^T \beta)^2 + \lambda \sum_{j=1}^{P} |\beta_j| \right),
\]

(2.3)

where \( K \) is the number of days in the estimation window, and \( \lambda > 0 \) is the penalty parameter. Theoretical justification for the use of LASSO to estimate the parameters in this setup with trends can be found in Masini and Medeiros (2021).

Once we estimate Eq. (2.2), we proceed to a second step: estimating the ECM using ordinary least squares (OLS) with the variables selected in the first step with the LASSO. The final prediction \( h \) steps ahead from \( T \) reads as

\[
\hat{y}_{T+h} = \Delta x_{T+h} \hat{\mathbf{\alpha}}_T - \hat{y}_{T-h} \hat{\mathbf{\beta}}_T + (1 + \hat{\gamma}_T) \hat{y}_{T+h-1},
\]

(2.4)

where \( \hat{y}_{T+h-1} \) is the forecast for the previous day. Confidence intervals are obtained through simulation by assuming that the error term \( u_{\tau} \) in (2.1) is normally distributed.

The intuition behind the proposed ECM is to model the dynamics and the reactions to departures from the equilibrium between \( y_{\tau} \) and \( \mathbf{x}_{\tau} \): the disease behaves in a somewhat similar fashion in different countries. Note that we do not claim a causal link, for instance, from the cases in Germany to the cases in Brazil, due to mobility in these two countries. What the model explores is that the evolution of diseases like Covid-19 seems to share similar patterns in different locations. Furthermore, as the first-stage LASSO regression is a model selection tool, if our hypothesis of common dynamics among countries is not valid, the LASSO will not select any country to explain the latecomer behavior and/or the residuals of the first-stage LASSO regression present statistical evidence of non-stationarity.

Our interest relies on the forecasts for the number of cases in levels, not in logs: \( Y_{\tau} := \exp(y_{\tau}) \). Therefore, for the horizon \( T+h \), the forecasts are constructed as follows:

\[
\hat{Y}_{T+h} = \hat{\alpha}_T \cdot \hat{y}_{T+h},
\]

(2.5)

where

\[
\hat{\alpha}_T = \frac{1}{P} \sum_{\tau = T-K+1}^{T} \exp(\hat{u}_{\tau}) := \frac{1}{K} \sum_{\tau = T-K+1}^{T} \exp(y_{\tau} - \hat{y}_{\tau})
\]

is a correction which is essential to attenuate the induced bias when we take the exponential of the forecasted value of \( Y_{T+h} \); see Wooldridge (2019). Note that, in the \( \tau \)-scale, the peers are 'in the future' and we can plug in actual values of \( x_{T+h} \) to construct our forecasts. Further, note that a rolling estimation window of \( K = 28 \) days induces an adaptive forecasting
framework suitable for capturing the dynamic nature of the pandemic and to attenuate the effects of outliers and potential structural breaks. Finally, in order to give more weight to the newest observations, we inflate the data by repeating the last four observations, where the last observation is repeated four times with a linear decay for the observations before. It is worth emphasizing that this model is only a local approximation of a more complex and dynamic process. Therefore, its best use relies on fresh and updated data, and the rolling-window scheme takes care of that. Although the model provides excellent adherence, the proposed forecasting method may be complemented with indexes, such as proxies for social distancing, to guide the evaluation of the future dynamics affecting the number of cases and deaths. However, the inclusion of other regressors such as Google Mobility did not show any improvement to the quality of the forecasts.

3. Guide to practice

The implementation of the proposed forecasting method requires the choice of three tuning parameters: the penalty term in the first-stage LASSO regression ($\lambda$), the length of the estimation window ($K$), and the data inflation mechanism.

The penalty term is selected by the Bayesian information criterion (BIC), as discussed in Medeiros and Mendes (2016). The degrees of freedom of the LASSO are determined by the nonzero estimated coefficients. Cross-validation can also be used to determine the penalty parameters. However, we prefer the BIC in order to avoid any extra computational burden.

The estimation window length ($K$) and the inflation scheme for the most recent observations can be estimated in a rolling-window process. The motivation for the rolling-window estimation and the data inflation is to give weight to more recent observations and attenuate potential structural breaks. Before computing the actual forecasts, one could select these tuning parameters from a rolling window using previous data and choose the values that minimize the out-of-sample error measure (MAPE, for example). However, this procedure gives us the best model for past data, especially because we need a significant number of windows that go back several weeks to obtain stable estimates. Although a procedure like this could lead to some local improvements, it could also lead to situations where the forecast explodes, especially when a very small $K$ is selected with no data inflation. To avoid unnecessary data mining that could lead to unreliable results, we use a fixed value of $K = 28$ and the inflation scheme for the four most recent lags. This sample size is enough to get a satisfactory model, given the number of variables, and it is not so big that it includes many structural breaks. The inflation scheme merely gives more weight to the most recent data, which is likely to be more similar to future data in the short run. Small changes in the inflation strategy do not affect the overall results. However, no inflation yields higher errors.

Another important point is to check whether the first-stage errors are stationary. This can be conducted by common unit-root tests. If the null hypothesis of unit-root is not rejected, the first stage is clearly misspecified and the forecasts will not be reliable. In our study, we ran unit-root tests after every LASSO estimation, and there was no evidence of misspecification.

Finally, the proposed ECM methodology is flexible enough to include other regressors, such as mobility data. However, in our experience the inclusion of such data did not bring any evident improvement in the performance of the model.

4. Data

We used the John Hopkins compiled data\textsuperscript{8} for all countries with Covid-19 cases. We also used the Brazilian Ministry of Health official data.\textsuperscript{9} The data were organized in epidemiological time, i.e., representing the number of days after the 100th case.

\textsuperscript{8} John Hopkins data are available at https://github.com/CSSEGISandData/COVID-19.

\textsuperscript{9} Brazilian Ministry of Health data available at https://covid.saude.gov.br/.
The models were computed on a rolling-window scheme with 28 in-sample observations per window. For each country, model estimation started when the number of confirmed cases of Covid-19 reached 20,000. The last in-sample day for every country was June 30, 2021. As described in Section 3, we set the window length to 28 days because this turned out to be a good trade-off between the quality of the in-sample adjustment and robustness to potential structural breaks.

Fig. 1 illustrates the evolution of Covid-19 cases in several countries. The data are displayed in epidemiological time. That is, the x-axis represents the number of days since the first registered case. It is clear from the figure that some countries are ahead of others in epidemiological time.

5. Results

5.1. Benchmark models

In other to compare the performance of the proposed ECM model, we considered the benchmark model described in Coroneo et al. (2020). The model is a simple quadratic trend regression (trend), defined as

\[ y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \epsilon_t, \]  

(5.1)

where \( \epsilon_t \) is a zero-mean error term. As mentioned above, although this benchmark is simple, it proved to be quite precise for short-term forecasts.

We also considered an integrated first-order autoregressive AR model:

\[ \Delta y_t = \phi_0 + \phi_1 \Delta y_{t-1} + \eta_t, \]  

(5.2)

where \( \eta_t \) is a zero-mean error term.

5.2. Mean absolute percentage errors

We start by reporting summary results for all 45 countries considered. Tables 1 and 2 report summary statistics for the forecasts of cases and deaths, respectively. Specifically, the tables show descriptive statistics for the ratio of the forecasting mean absolute percentage error (MAPE) of either ECM or trend alternatives and the AR benchmark. Ratios below one indicate that the alternative models outperform the benchmark. For cases, the medians of the ratio of the MAPEs from the ECM and the AR models are below one for horizons between three and 12 days ahead. When deaths are considered, the medians are below one for all horizons of more than two days ahead. Furthermore, the ECM is much better than the trend benchmark, which is frequently beaten by the AR model. The medians of the ratios over the 14 forecasting horizons and all countries are 0.99 for cases and 0.78 for deaths, indicating the superiority of the ECM over the AR model.

Now we turn our attention to four specific countries. Tables 3 and 4 present the forecasting results for the full out-of-sample period. The tables show the MAPE for forecasting horizons of 1 to 14 days ahead of the Covid-19 accumulated number of cases (Table 3) and deaths (Table 4) for Brazil, Chile, Mexico, and Portugal. Values in parenthesis are p-values for the Giacomini & White test of superior predictive ability (Giacomini & White, 2006). The null hypothesis of the test is that both forecasts have the same MAPE.

We start by comparing the models with respect to the forecasts for case counts. For Brazil, the ECM has lower MAPEs than the AR benchmark at all horizons of more than one day ahead, and the differences are all significant. It is also clear from the table that the ECM outperforms the trend model. For Chile, the ECM is better than the benchmark at all horizons. The differences are statistically significant at 12 out of 14 horizons. For Mexico, the ECM is superior to the benchmark at all horizons of more than one day ahead, and the differences are significant. Finally, for Portugal, the ECM has a much worse performance and is superior to the AR alternative only at horizons between two and six days ahead, but the differences are only significant for three-day-ahead forecasts. Still, the ECM considerably outperforms the trend model.

Turning to fatalities, the ECM model is clearly superior to the benchmark, and the differences are much more pronounced. For Brazil, the ECM is better than the AR for all horizons of more than one day ahead, and the differences are strongly significant. Our proposal also shows much better results than the trend alternative. For Chile and Mexico, the results are similar. However, we should point out the for Chile there is a strong outlier in the forecasts of the AR models that distorts the results, with ratios close to zero at larger horizons. When we remove this specific outlier, the ratio for one-day-ahead forecasts is 1.323 and ranges between 0.302 and 0.868. This still indicates the superiority of the ECM for most horizons. For Portugal, the AR model delivers the best results at all horizons, followed by the ECM.

In order to complement the analysis, we compute the frequency of days that the ECM has a lower absolute percentage error than the benchmark and the median of the ratio of the daily absolute errors of the ECM and benchmark specification over the forecasting sample. Figs. 2 and 3 report, for each forecasting horizon, the frequency of days over the sample when the daily absolute percentage error of the ECM is smaller than the one from the AR and trend alternatives, respectively. The first column in the figures presents the results for cases, whereas the second column shows the numbers of deaths. To quantify these gains, Figs. 4 and 5 present, for each horizon, the median of the ratios of the daily absolute errors. A number less than one favors the ECM model. As before, the first (second) column concerns cases (deaths). We prefer the use the median instead of the mean to avoid the potential effects of outliers.

For Brazil, the results favor the ECM over the AR alternative for horizons of more than one day ahead. Figs. 3 and 5 both point to the superiority of the benchmark. However, when looking at the results in Tables 3 and 4 above, we reach a different conclusion. Therefore, it is important to uncover the drivers of the best MAPE of the ECM over the entire out-of-sample period. The reason for this finding is that the ECM is far superior to the benchmark during the first 100 days of the sample. This was the period when the number of cases and deaths in Brazil was accelerating the most.
### Table 1
Forecasting Results for Cases: Distribution of Mean Absolute Percentage Error Ratios. The table presents results with respect forecasting models for the number of deaths of Covid-19. The table shows descriptive statistics for the ratio of the forecasting mean absolute percentage error (MAPE) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

#### Descriptive Statistics: Ratio of Forecasting Mean Absolute Percentage Errors

| Days ahead | Cases: ECM x AR | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|------------|----------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Min        | 0.885          | 0.536 | 0.103 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 5% prct    | 0.984          | 0.748 | 0.514 | 0.414 | 0.365 | 0.340 | 0.329 | 0.335 | 0.344 | 0.345 | 0.355 | 0.375 | 0.358 | 0.370 |
| 10% prct   | 1.012          | 0.787 | 0.668 | 0.582 | 0.556 | 0.547 | 0.568 | 0.579 | 0.590 | 0.586 | 0.592 | 0.599 | 0.607 | 0.615 |
| 25% prct   | 1.090          | 0.899 | 0.787 | 0.745 | 0.724 | 0.725 | 0.709 | 0.709 | 0.719 | 0.736 | 0.766 | 0.777 | 0.799 | 0.798 |
| Median     | 1.172          | 1.034 | 0.955 | 0.908 | 0.962 | 0.951 | 0.957 | 0.975 | 0.987 | 0.993 | 0.996 | 1.042 | 1.103 | 1.103 |
| 75% prct   | 1.383          | 1.149 | 1.073 | 1.097 | 1.090 | 1.090 | 1.098 | 1.138 | 1.147 | 1.184 | 1.192 | 1.206 | 1.216 | 1.267 |
| 90% prct   | 1.812          | 1.572 | 1.593 | 1.560 | 1.483 | 1.648 | 1.836 | 1.927 | 1.915 | 1.886 | 1.862 | 1.838 | 1.830 | 1.856 |
| 95% prct   | 2.116          | 2.003 | 1.842 | 1.804 | 1.814 | 1.873 | 1.973 | 1.979 | 2.009 | 2.033 | 2.101 | 2.177 | 2.248 | 2.325 |
| Max        | 2.710          | 2.570 | 2.439 | 2.304 | 2.238 | 2.140 | 2.035 | 2.134 | 2.667 | 3.408 | 4.653 | 6.403 | 9.458 | 12.029 |
| Mean       | 1.301          | 1.123 | 1.019 | 0.973 | 0.963 | 0.970 | 0.991 | 1.010 | 1.027 | 1.053 | 1.094 | 1.155 | 1.244 | 1.330 |
| Std        | 0.368          | 0.389 | 0.416 | 0.425 | 0.431 | 0.436 | 0.451 | 0.461 | 0.500 | 0.569 | 0.708 | 0.929 | 1.343 | 1.708 |

### Table 2
Forecasting Results for Deaths: Distribution Mean Absolute Percentage Error Ratios. The table presents results with respect forecasting models for the number of deaths of Covid-19. The table shows descriptive statistics for the ratio of the forecasting mean absolute percentage error (MAPE) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

#### Descriptive Statistics: Ratio of Forecasting Mean Absolute Percentage Errors

| Days ahead | Cases: Trend x AR | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|------------|-------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Min        | 1.151             | 0.859 | 0.247 | 0.111 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 5% prct    | 1.552             | 0.917 | 0.585 | 0.455 | 0.394 | 0.358 | 0.339 | 0.352 | 0.361 | 0.380 | 0.391 | 0.398 | 0.405 | 0.413 |
| 10% prct   | 1.694             | 1.072 | 0.780 | 0.704 | 0.668 | 0.635 | 0.599 | 0.616 | 0.616 | 0.624 | 0.622 | 0.626 | 0.632 | 0.639 |
| 25% prct   | 1.975             | 1.299 | 1.011 | 0.902 | 0.879 | 0.917 | 0.934 | 0.921 | 0.908 | 0.906 | 0.891 | 0.853 | 0.811 | 0.783 |
| Median     | 2.685             | 1.726 | 1.468 | 1.322 | 1.292 | 1.256 | 1.253 | 1.230 | 1.195 | 1.167 | 1.151 | 1.159 | 1.151 | 1.157 |
| 75% prct   | 3.452             | 2.373 | 1.910 | 1.691 | 1.598 | 1.600 | 1.607 | 1.562 | 1.531 | 1.508 | 1.524 | 1.538 | 1.531 | 1.526 |
| 90% prct   | 5.728             | 3.157 | 2.623 | 2.349 | 2.292 | 2.128 | 2.062 | 2.024 | 1.956 | 1.904 | 1.856 | 1.815 | 1.786 | 1.777 |
| 95% prct   | 9.620             | 4.615 | 3.644 | 3.083 | 2.722 | 2.510 | 2.338 | 2.183 | 2.059 | 1.975 | 1.950 | 1.963 | 1.990 | 2.004 |
| Max        | 11.652            | 7.210 | 5.460 | 4.737 | 4.170 | 3.815 | 3.572 | 3.386 | 3.240 | 3.130 | 3.070 | 3.055 | 3.074 | 3.104 |
| Mean       | 3.195             | 2.029 | 1.618 | 1.431 | 1.340 | 1.308 | 1.297 | 1.270 | 1.237 | 1.211 | 1.197 | 1.193 | 1.196 | 1.203 |
| Std        | 1.912             | 1.195 | 0.956 | 0.833 | 0.740 | 0.681 | 0.642 | 0.608 | 0.582 | 0.560 | 0.547 | 0.542 | 0.542 | 0.545 |

### Tables with Additional Information

- **Deaths: ECM x AR**
  - [Days ahead](#)
  - [Cases](#)
- **Deaths: Trend x AR**
  - [Days ahead](#)
  - [Cases](#)
Table 3
Forecasting Results for Cases: Mean Absolute Percentage Error Ratios. The table shows the ratios of the forecasting mean absolute percentage error (MAPE) of either the ECM or the trend model over the AR benchmark. Numbers below one indicate that the ECM or the trend model outperforms the AR. The table presents results for forecasting horizons of 1 to 14 days ahead of the Covid-19 accumulated number of cases. The models were computed on a rolling window scheme with 28 in-sample observations per window. For each country, the model estimation started when the number of cases of Covid-19 reached 20,000. The last out-of-sample day for every country is July 1, 2021. Values in parenthesis are p-values for the Giacomini & White test (Giacomini & White, 2006).

| Cases: Forecasting Mean Absolute Percentage Errors |
|---------------------------------------------------|
| Days ahead | Brazil | Chile | Mexico | Portugal |
|------------|--------|-------|--------|----------|
| Days ahead |        |       |        |          |
| Brazil     |        |       |        |          |
| Trend      | 1.983  | 1.476 | 1.800  | 3.353    |
| ECM        | 1.157  | 0.994 | 1.003  | 1.147    |
| Days ahead |        |       |        |          |
| Chile      |        |       |        |          |
| Trend      | 1.126  | 0.869 | 1.107  | 2.079    |
| ECM        | 0.899  | 0.758 | 0.831  | 0.965    |
| Days ahead |        |       |        |          |
| Mexico     |        |       |        |          |
| Trend      | 0.813  | 0.587 | 0.850  | 1.677    |
| ECM        | 0.813  | 0.543 | 0.745  | 1.536    |
| Days ahead |        |       |        |          |
| Portugal   |        |       |        |          |
| Trend      | 0.832  | 0.540 | 1.540  | 1.506    |
| ECM        | 0.847  | 0.598 | 1.605  | 1.664    |

Fig. 2. Frequency of days when ECM is better than the AR model.

The figure illustrates for different countries and horizons, the frequency of days when the absolute error of the ECM is smaller than the one from the AR specification.
Table 4
Forecasting Results for Deaths: Mean Absolute Percentage Error Ratios. The table shows the ratios of the forecasting mean absolute percentage error (MAPE) of either the ECM or the trend model over the AR benchmark. Numbers below one indicate that the ECM or the trend model outperforms the AR. The table presents results for forecasting horizons of 1 to 14 days ahead of the Covid-19 accumulated number of deaths. The models were computed on a rolling window scheme with 28 in-sample observations per window. For each country, the model estimation started when the number of cases of Covid-19 reached 20,000. The last out-of-sample day for every country is July 1, 2021. Values in parenthesis are p-values for the Giacomini & White test. (Giacomini & White, 2006).

Deaths: Forecasting Mean Absolute Percentage Errors

| Days ahead | Brazil       |               |               |               |               |               |               |               |               |               |               |               |               |               |               |
|------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|            | Trend        | 2.369         | 1.259         | 1.021         | 0.969         | 0.973         | 1.007         | 1.054         | 1.043         | 1.006         | 0.972         | 0.956         | 0.946         | 0.933         | 0.920         |
|            |              | (0.000)       | (0.141)       | (0.463)       | (0.444)       | (0.453)       | (0.489)       | (0.420)       | (0.437)       | (0.490)       | (0.457)       | (0.433)       | (0.417)       | (0.398)       | (0.375)       |
|            | ECM          | 1.400         | 0.881         | 0.745         | 0.695         | 0.686         | 0.696         | 0.725         | 0.738         | 0.738         | 0.732         | 0.738         | 0.743         | 0.751         | 0.766         |
|            |              | (0.000)       | (0.138)       | (0.027)       | (0.010)       | (0.002)       | (0.000)       | (0.000)       | (0.001)       | (0.016)       | (0.024)       | (0.032)       | (0.036)       | (0.039)       | (0.059)       |

| Days ahead | Chile        |               |               |               |               |               |               |               |               |               |               |               |               |               |               |
|------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|            | Trend        | 2.471         | 1.511         | 1.207         | 0.966         | 0.745         | 0.522         | 0.296         | 0.131         | 0.047         | 0.013         | 0.003         | 0.000         | 0.000         | 0.000         |
|            |              | (0.000)       | (0.031)       | (0.268)       | (0.465)       | (0.294)       | (0.210)       | (0.175)       | (0.162)       | (0.158)       | (0.156)       | (0.156)       | (0.156)       | (0.155)       | (0.155)       |
|            | ECM          | 1.256         | 0.818         | 0.668         | 0.542         | 0.418         | 0.295         | 0.169         | 0.076         | 0.027         | 0.007         | 0.002         | 0.000         | 0.000         | 0.000         |
|            |              | (0.079)       | (0.208)       | (0.137)       | (0.103)       | (0.086)       | (0.103)       | (0.130)       | (0.146)       | (0.153)       | (0.155)       | (0.156)       | (0.156)       | (0.155)       | (0.155)       |

| Days ahead | Mexico       |               |               |               |               |               |               |               |               |               |               |               |               |               |               |
|------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|            | Trend        | 1.799         | 0.987         | 0.796         | 0.719         | 0.677         | 0.656         | 0.649         | 0.634         | 0.610         | 0.591         | 0.576         | 0.565         | 0.557         | 0.552         |
|            |              | (0.000)       | (0.453)       | (0.024)       | (0.006)       | (0.002)       | (0.000)       | (0.000)       | (0.000)       | (0.000)       | (0.001)       | (0.001)       | (0.001)       | (0.001)       | (0.001)       |
|            | ECM          | 1.552         | 0.983         | 0.756         | 0.627         | 0.528         | 0.481         | 0.492         | 0.511         | 0.513         | 0.494         | 0.476         | 0.455         | 0.450         | 0.454         |
|            |              | (0.000)       | (0.418)       | (0.000)       | (0.000)       | (0.000)       | (0.000)       | (0.000)       | (0.000)       | (0.000)       | (0.001)       | (0.001)       | (0.001)       | (0.001)       | (0.001)       |

| Days ahead | Portugal      |               |               |               |               |               |               |               |               |               |               |               |               |               |               |
|------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|            | Trend        | 5.425         | 3.881         | 3.341         | 2.983         | 2.736         | 2.570         | 2.445         | 2.340         | 2.258         | 2.182         | 2.113         | 2.061         | 2.016         | 1.974         |
|            |              | (0.000)       | (0.000)       | (0.000)       | (0.001)       | (0.001)       | (0.002)       | (0.004)       | (0.006)       | (0.009)       | (0.012)       | (0.015)       | (0.018)       | (0.021)       | (0.024)       |
|            | ECM          | 2.114         | 1.640         | 1.526         | 1.419         | 1.347         | 1.292         | 1.267         | 1.237         | 1.217         | 1.197         | 1.183         | 1.172         | 1.161         | 1.153         |
|            |              | (0.000)       | (0.000)       | (0.000)       | (0.001)       | (0.001)       | (0.004)       | (0.005)       | (0.057)       | (0.092)       | (0.122)       | (0.151)       | (0.171)       | (0.191)       | (0.212)       |

Fig. 3. Frequency of days when ECM is better than the trend model.
The figure illustrates for different countries and horizons, the frequency of days when the absolute error of the ECM is smaller than the one from the trend specification.
For Chile, the one-day-ahead forecasts of ECM are the best ones in 72.17% (61.75%) of the days when cases (deaths) are considered. These numbers drop to 46.52% (53.91%) when the forecasting horizon is set to 14 days. For almost all the horizons, the proportion of days when the ECM is better than the benchmark is larger than 50%. From the analysis of the results in Fig. 4, it is clear that the ECM is better than the benchmark at almost all horizons when cases are considered. When fatalities are analyzed, the ECM is always superior.

For Mexico, the results are very supportive of the ECM specification. The ECM is superior to the benchmark in more than 50% of the days in almost every case considered in the analysis. The median ratios of the absolute
percentage errors are always below one when Covid-19 cases are considered. For deaths, the ratios are below one at horizons of more than four days ahead.

Finally, for Portugal, the superiority of the ECM draws attention. For all horizons considered, the ECM is better than the benchmark in terms of the number of days with lower errors, as well as in terms of the relative magnitude of the absolute percentage errors.

5.3. Diagnostic tests

We report two diagnostic tests. First, Fig. 6 illustrates the empirical distribution of the estimated coefficient of a first-order autoregressive (AR(1)) model estimated with the residuals from the first-stage LASSO regression. The distribution is over all the rolling windows and each of the four countries analyzed. It is clear from the figure that, apart from a single case, all the estimates are below one in absolute values. Unit-root tests also strongly reject the null of unit roots in all but one case. This specific negative case is related to a huge outlier in the data that distorts the estimation of the AR coefficient and, consequently, the unit-root test. Based on this analysis, we are confident that our methodology is adequate for the present data.

The second diagnostic is related to the data inflation heuristic. Table 6 presents the MAPEs of the ECM with data inflation divided by the MAPEs of the ECM without data inflation. Numbers lower than one favor the inflation heuristic. For Brazil, Chile, and Portugal, it is clear that data inflation is superior to no inflation at all. For Mexico, we see improvements when the forecasts for cases are considered but not deaths. Changing the number of inflated observations seemed to have no significant effect. These additional results are available upon request.

Finally, it is important to understand which variables are being selected by the LASSO during the first stage of the methodology. Table 5 shows the frequency of selection of each variable over the rolling windows. Mexico is the latecomer country where each variable in the pool is selected at least once. Portugal seems to be the country with the most parsimonious model. Note that the frequency of selection of each variable differs from country to country.

5.4. Robustness

In this section, we report a number of robustness checks. We start by reporting results concerning a number of alternative loss functions. In addition, we provide results showing the potential effects of vaccination on the performance of the ECM model.

5.4.1. Alternative loss functions

In addition to the MAPE, we present forecasting results for the following loss functions: (i) the median absolute deviation from the median (MAD), (ii) the mean absolute error (MAE), and (iii) the mean squared error. The results for the MAD, which is a loss function that is robust to outliers, are shown in Tables 7, 8, and 14. The tables provide descriptive statistics of the MAD ratios with respect to the AR benchmark. With respect to cases, the ECM outperforms the AR alternative in more than 50% of the countries at horizons between three and 12 days. Furthermore, the ECM is much better than the trend model. With respect to deaths, the performance of the ECM improves. In Table 14, we report the results for the combination of ECM and AR models. It is clear from the table that the combination of the models strongly improves the results for both cases and deaths.

The results concerning the MAE are reported in Tables 9, 10, and 15. The results in Table 9 for the ECM are not very encouraging. However, the combination of the ECM and the AR model is clearly superior to the AR benchmark. For deaths, the results for the ECM are much better.

Finally, the results for the MSE are shown in Tables 11, 12, and 16. The results for the ECM are not particularly good, especially due to some outliers. However, the combination of the ECM and the AR model seems to be very robust no matter which loss function is chosen.

5.4.2. Forecast errors and vaccination

One question of interest is whether the evolution of vaccination affects the performance of the ECM. To answer this question, we ran the following regression for each country:

$$\log \left( \frac{\epsilon_{ECM, i,t+h}}{\epsilon_{AR, i,t+h}} \right) = \varrho_i + \rho_i V_{i,t+h} + \epsilon_{i,t+h}, \quad i = 1, \ldots, 45,$$

(5.3)
### Table 5
Proportion of times each variable is selected by the LASSO. The table shows the frequency of times each variable is selected by the LASSO in the first-stage regression.

| Target       | $t$ | $t^2$ | France | Iran | Italy | Japan | South Korea | Singapore | Germany | Spain | United Kingdom | US |
|--------------|-----|-------|--------|------|-------|-------|-------------|-----------|---------|-------|----------------|----|
| Brazil       | 0.52| 0.07  | 0.50   | 0.64 | 0.70  | 0.37  | 0.53        | 0.57      | –       | –     | –              | –  |
| Chile        | 0.63| 0.16  | 0.55   | 0.55 | 0.38  | 0.45  | 0.32        | 0.69      | 0.25    | –     | –              | –  |
| Mexico       | 0.41| 0.09  | 0.08   | 0.66 | 0.17  | 0.31  | 0.32        | 0.25      | 0.39    | 0.29  | 0.33           | 0.67|
| Portugal     | 0.54| 0.42  | –      | 0.69 | 0.69  | 0.40  | 0.52        | –         | –       | –     | –              | –  |

### Table 6
Effects of data inflation. Forecasting MAPEs of the ECM with data inflation divided by the forecasting MAPEs of the ECM without data inflation. Numbers lower than one favor the inflation heuristic.

| Horizon | Country | Cases | Deaths | Cases | Deaths | Cases | Deaths | Cases | Deaths | Cases | Deaths | Cases | Deaths |
|---------|---------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|
| 1       | Brazil  | 0.868 | 1.040  | 0.828 | 0.942  | 0.915 | 1.009  | 0.807 | 0.880  | 0.886 | 0.952  | 0.883 | 0.922  |
| 2       | Chile   | 0.857 | 0.840  | 0.818 | 0.922  | 0.948 | 1.051  | 0.883 | 0.875  | 0.886 | 0.952  | 0.883 | 0.922  |
| 3       | Mexico  | 0.870 | 0.854  | 0.812 | 0.925  | 0.959 | 1.055  | 0.892 | 0.873  | 0.892 | 0.955  | 0.892 | 0.972  |
| 4       | Portugal| 0.866 | 0.820  | 0.852 | 0.927  | 0.992 | 1.064  | 0.932 | 0.862  | 0.932 | 0.955  | 0.932 | 0.862  |

### Table 7
Forecasting Results for Cases: Distribution of Median Absolute Deviation from the Median Ratios. The table presents results with respect forecasting models for the number of cases of Covid-19. The table shows descriptive statistics for the ratio of the forecasting median absolute deviation from the median (MAD) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

**Descriptive Statistics: Ratio of Forecasting Median Absolute Deviation from the Median Ratios**

| Days ahead | Cases: ECM x AR | Cases: Trend x AR |
|------------|-----------------|------------------|
| 1          | Min 0.910       | 1.230            |
|            | 5% prct 0.972   | 1.624            |
|            | 10% prct 1.027  | 1.884            |
|            | Median 1.193    | 2.091            |
|            | 75% prct 1.441  | 2.509            |
|            | 90% prct 1.906  | 2.959            |
|            | 95% prct 2.158  | 3.267            |
|            | Max 2.656       | 4.237            |
|            | Mean 3.236      | 3.237            |
|            | Std 0.373       | 0.406            |

**Days ahead**

| Cases: ECM x AR |
|----------------|
| Min 0.904      |
| 5% prct 0.950  |
| 10% prct 1.003 |
| Median 1.304   |
| 75% prct 3.262 |
| 90% prct 7.304 |
| 95% prct 12.273|
| Max 2.236      |
| Mean 2.059     |
| Std 0.190      |
Table 8
Forecasting Results for Deaths: Distribution Median Absolute Deviation from the Median Ratios. The table presents results with respect forecasting models for the number of deaths by Covid-19. The table shows descriptive statistics for the ratio of the forecasting median absolute deviation from the median (MAD) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Deaths: ECM x AR |
|-----------|------------------|
|           | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
| Min       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 10% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 25% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Median     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 75% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 90% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 95% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Max        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Mean       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Std        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

Table 9
Forecasting Results for Cases: Distribution of Mean Absolute Error Ratios. The table presents results with respect forecasting models for the number of cases of Covid-19. The table shows descriptive statistics for the ratio of the forecasting mean absolute error (MAE) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Cases: ECM x AR |
|-----------|-----------------|
|           | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
| Min       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 10% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 25% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Median     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 75% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 90% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 95% prct   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Max        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Mean       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Std        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
Table 10
Forecasting Results for Deaths: Distribution Mean Absolute Error Ratios. The table presents results with respect forecasting models for the number of deaths by Covid-19. The table shows descriptive statistics for the ratio of the forecasting mean absolute ratio (MAE) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Deaths: ECM x AR | Days ahead | Deaths: Trend x AR |
|------------|------------------|------------|-------------------|
|            |                  |            |                   |
| Min        |                  |            |                   |
| 5% prct    |                  |            |                   |
| 10% prct   |                  |            |                   |
| 25% prct   |                  |            |                   |
| Median     |                  |            |                   |
| 75% prct   |                  |            |                   |
| 90% prct   |                  |            |                   |
| 95% prct   |                  |            |                   |
| Max        |                  |            |                   |
| Mean       |                  |            |                   |
| Std        |                  |            |                   |

Table 11
Forecasting Results for Cases: Distribution Mean Squared Error Ratios. The table presents results with respect forecasting models for the number of cases of Covid-19. The table shows descriptive statistics for the ratio of the forecasting mean squared error (MSE) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Cases: ECM x AR | Days ahead | Cases: Trend x AR |
|------------|-----------------|------------|------------------|
|            |                 |            |                  |
| Min        |                 |            |                  |
| 5% prct    |                 |            |                  |
| 10% prct   |                 |            |                  |
| 25% prct   |                 |            |                  |
| Median     |                 |            |                  |
| 75% prct   |                 |            |                  |
| 90% prct   |                 |            |                  |
| 95% prct   |                 |            |                  |
| Max        |                 |            |                  |
| Mean       |                 |            |                  |
| Std        |                 |            |                  |
Table 12
Forecasting Results for Deaths: Distribution Mean Squared Error Ratios. The table presents results with respect forecasting models for the number of deaths by Covid-19. The table shows descriptive statistics for the ratio of the forecasting mean squared error (MSE) of either ECM or Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Deaths: ECM x AR | Deaths: Trend x AR |
|------------|------------------|--------------------|
|            |                  |                    |
|            | Days ahead       |                    |
|            | Min              | Mean              |
|            | Std              | Max               |
|            | 95%prct          | 90%prct           |
|            | 75%prct          | Median            |
|            | 25%prct          | 10%prct           |
|            | 5%prct           | 0%prct            |
|            |                  | Min               |
|            | Days ahead       |                    |
|            | 1                | 2                 |
|            | 3                | 4                 |
|            | 5                | 6                 |
|            | 7                | 8                 |
|            | 9                | 10                |
|            | 11               | 12                |
|            | 13               | 14                |

Table 13
Forecasting Combination Results for Cases and Deaths: Distribution of Mean Absolute Percentage Error Ratios. The table presents results with respect forecasting models for the number of cases and deaths. The table shows descriptive statistics for the ratio of the forecasting mean absolute percentage error (MAPE) of either the combination ECM and AR or ECM and Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Cases: ECM and AR x AR | Deaths: ECM and AR x AR |
|------------|------------------------|-------------------------|
|            | Days ahead             |                        |
|            | Min                    | Mean                    |
|            | Std                    | Max                     |
|            | 95%prct                | 90%prct                 |
|            | 75%prct                | Median                  |
|            | 25%prct                | 10%prct                 |
|            | 5%prct                 | 0%prct                  |
|            |                        | Min                     |
|            | Days ahead             |                        |
|            | 1                     | 2                      |
|            | 3                     | 4                      |
|            | 5                     | 6                      |
|            | 7                     | 8                      |
|            | 9                     | 10                     |
|            | 11                    | 12                     |
|            | 13                    | 14                     |

Descriptive Statistics: Ratio of Forecasting Mean Squared Error Ratios

| Days ahead | Deaths: ECM x AR | Deaths: Trend x AR |
|------------|------------------|--------------------|
|            | Days ahead       |                    |
|            | Min              | Mean              |
|            | Std              | Max               |
|            | 95%prct          | 90%prct           |
|            | 75%prct          | Median            |
|            | 25%prct          | 10%prct           |
|            | 5%prct           | 0%prct            |
|            |                  | Min               |
|            | Days ahead       |                    |
|            | 1                | 2                 |
|            | 3                | 4                 |
|            | 5                | 6                 |
|            | 7                | 8                 |
|            | 9                | 10                |
|            | 11               | 12                |
|            | 13               | 14                |

Descriptive Statistics: Ratio of Forecasting Mean Absolute Percentage Errors

| Days ahead | Cases: ECM and AR x AR | Deaths: ECM and AR x AR |
|------------|------------------------|-------------------------|
|            | Days ahead             |                        |
|            | Min                    | Mean                    |
|            | Std                    | Max                     |
|            | 95%prct                | 90%prct                 |
|            | 75%prct                | Median                  |
|            | 25%prct                | 10%prct                 |
|            | 5%prct                 | 0%prct                  |
|            |                        | Min                     |
|            | Days ahead             |                        |
|            | 1                     | 2                      |
|            | 3                     | 4                      |
|            | 5                     | 6                      |
|            | 7                     | 8                      |
|            | 9                     | 10                     |
|            | 11                    | 12                     |
|            | 13                    | 14                     |
Table 14
Forecasting Combination Results for Cases and Deaths: Distribution of Median Absolute Deviation from the Median Ratios. The table presents results with respect forecasting models for the number of cases and deaths. The table shows descriptive statistics for the ratio of the forecasting median absolute deviation from the median (MAE) of either the combination ECM and AR or ECM and Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Cases: ECM and AR x AR | Deaths: ECM and AR x AR |
|------------|------------------------|------------------------|
|            | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
| Min        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5% prct     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 10% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 25% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Median      |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 75% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 95% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Max         |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Mean        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Std         |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

Table 15
Forecasting Combination Results for Cases and Deaths: Distribution of Mean Absolute Error Ratios. The table presents results with respect forecasting models for the number of cases and deaths. The table shows descriptive statistics for the ratio of the forecasting mean absolute error (MAE) of either the combination ECM and AR or ECM and Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Cases: ECM and AR x AR | Deaths: ECM and AR x AR |
|------------|------------------------|------------------------|
|            | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
| Min        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5% prct     |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 10% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 25% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Median      |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 75% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 95% prct    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Max         |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Mean        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Std         |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
Table 16
Forecasting Combination Results for Cases and Deaths: Distribution of Mean Squared Error Ratios. The table presents results with respect forecasting models for the number of cases and deaths. The table shows descriptive statistics for the ratio of the forecasting mean squared error (MSE) of either the combination ECM and AR or ECM and Trend models and the AR benchmark. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

### Descriptive Statistics: Ratio of Forecasting Mean Squared Error

#### Cases: ECM and AR x AR

| Days ahead | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   |
|------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Min        | 0.374| 0.280| 0.256| 0.250| 0.250| 0.250| 0.250| 0.250| 0.250| 0.249| 0.247| 0.243| 0.241| 0.240|
| 5% prct    | 0.585| 0.397| 0.331| 0.289| 0.278| 0.263| 0.260| 0.255| 0.254| 0.250| 0.250| 0.250| 0.250| 0.250|
| 10% prct   | 0.646| 0.508| 0.453| 0.397| 0.368| 0.315| 0.338| 0.293| 0.323| 0.311| 0.332| 0.310| 0.348| 0.301|
| 25% prct   | 0.702| 0.556| 0.500| 0.472| 0.446| 0.441| 0.450| 0.466| 0.460| 0.461| 0.466| 0.465| 0.488| 0.498|
| Median     | 0.854| 0.685| 0.630| 0.599| 0.563| 0.585| 0.664| 0.683| 0.686| 0.696| 0.703| 0.701| 0.746| 0.802|
| Max        | 1.693| 1.933| 1.762| 2.908| 2.450| 2.284| 1.953| 1.885| 1.926| 2.077| 2.248| 2.637| 3.401| 4.638|
| Mean       | 0.958| 0.841| 0.876| 0.894| 0.855| 0.805| 0.840| 0.862| 0.997| 1.229| 1.810| 2.838| 5.536| 9.048|
| Std        | 0.407| 0.546| 1.104| 1.054| 0.835| 0.604| 0.695| 0.797| 1.612| 3.091| 6.838| 13.516| 30.806| 53.518|

#### Deaths: ECM and AR x AR

| Days ahead | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   |
|------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Min        | 0.247| 0.250| 0.250| 0.247| 0.251| 0.250| 0.250| 0.250| 0.249| 0.245| 0.244| 0.244| 0.246| 0.247|
| 5% prct    | 0.250| 0.250| 0.270| 0.251| 0.255| 0.253| 0.252| 0.252| 0.250| 0.250| 0.250| 0.250| 0.250| 0.250|
| 10% prct   | 0.473| 0.344| 0.321| 0.276| 0.260| 0.258| 0.257| 0.257| 0.253| 0.252| 0.252| 0.250| 0.250| 0.250|
| 25% prct   | 0.679| 0.505| 0.437| 0.424| 0.383| 0.359| 0.302| 0.326| 0.325| 0.300| 0.302| 0.329| 0.345| 0.342|
| Median     | 0.871| 0.685| 0.599| 0.546| 0.509| 0.485| 0.470| 0.461| 0.435| 0.444| 0.441| 0.452| 0.426| 0.431|
| Max        | 1.244| 1.087| 0.937| 0.856| 0.788| 0.775| 0.783| 0.745| 0.749| 0.759| 0.785| 0.824| 0.859| 0.914|
| Mean       | 7.681| 4.169| 2.354| 2.050| 1.726| 1.723| 2.434| 2.802| 2.791| 3.240| 3.908| 4.951| 6.799| 9.317|
| Std        | 45.043| 22.583| 10.325| 8.846| 7.053| 7.260| 12.112| 14.543| 14.314| 17.337| 21.569| 28.231| 40.160| 56.002|

**Fig. 7.** t-statistic.

The figure illustrates the t-statistic of the regression of $\log \left( \frac{\text{ECM}_{i,t+h}}{\text{AR}_{i,t+h}} \right)$ on the proportion of the population vaccinated in each country.

where $\text{ECM}_{i,t+h}$ and $\text{AR}_{i,t+h}$ are the out-of-sample forecasting errors of the ECM and AR models, respectively, at the $h$-step-ahead horizon, and $V_{i,t+h}$ is the proportion of the population that is fully vaccinated.

In Fig. 7, we report the t-statistic for the null hypothesis $H_0: \rho_i = 0$. The red lines indicate the $+/− 1.96$ threshold. Standard errors for the coefficients are robust to heteroskedasticity and autocorrelation. As we
can see from the plots, for most countries, the null hypothesis is not rejected, such that there is no evidence that vaccination is correlated with the performance of the models. For the case where the null is rejected, the effects can be either positive or negative. The adaptiveness of the locally regularized linear models capture the slowly-varying dynamics in the relationship between dependent and independent variables. As such, we conclude that cross-country differences in vaccination do not notably affect the performance of the ECM.

6. Conclusion

We proposed a statistical model for very short-run forecasts of Covid-19 cases and deaths in countries and regions that are latecomers. We believe this is a useful tool to inform healthcare management. Nonetheless, structural breaks might worsen forecasts a few days after such breaks. So the use of this tool should be complemented by other external information, such as proxies for social distancing, to guide subjective or objective assessments of potential dynamic changes in the pandemic’s evolution. We hope to enhance the model by improving the methodology and incorporating more information. And we aim to keep the forecasts, methodology, and codes updated on a daily basis at https://covid19analytics.com.br/.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Acknowledgments

The authors wish to thank the Associate Editor and two anonymous referees for very helpful comments, and CAPES, CNPq, and FAPERJ for partial financial support. The authors are extremely grateful to all the Covid19 Analytics (www.covid19analytics.com.br) team for their work and for many enlightening discussions. The authors are also in debt to Marcelo Fernandes for providing important comments on the paper.

Appendix. Additional results

In this appendix we report a number of additional results.

A.1. Rolling MAPEs

In order to analyze how the errors unfold over the evolution of the pandemic, we plot rolling MAPEs over 14 days in Figs. 8–11. We report only results for selected horizons. It is clear from the figures that both models improve over time. In some cases the reductions are larger than 50%. For Chile, the gains of the ECM over the benchmark are more evident during the months of July and August when we look at cases. On the other hand, for deaths, the ECM is better than the benchmark for all windows. In case of Brazil, the superiority of the ECM forecasts of cases is more evident at the beginning of the sample, when the benchmark performs very poorly. A similar pattern is visible in the case of deaths. In the case of Mexico and for forecasts of case counts, the ECM is systematically superior to the benchmark over the sample and when we consider one-day-ahead forecasts. For the other horizons, the gains are more concentrated at the beginning and end of the sample. Equivalent conclusions emerge when we look at forecasts of deaths. In the case of Portugal, the benefits of using the ECM instead of the benchmark are clearer in the first half of the sample.

![Fig. 10. Rolling Mean absolute percentage error - Mexico.](image1)
The figure illustrates, for different horizons, the Mean Absolute Percentage Error (MAPE) computed over rolling windows with 14 observations.

![Fig. 11. Rolling Mean absolute percentage error - Portugal.](image2)
The figure illustrates, for different horizons, the Mean Absolute Percentage Error (MAPE) computed over rolling windows with 14 observations.
Table 17
Forecasting Results: Distribution of Exception Rates (90% Prediction Interval). The table presents results with respect to the prediction intervals produced by the ECM for both cases and deaths. The table shows descriptive statistics for the frequency that the absolute out-of-sample errors of the ECM model exceed the 90% prediction interval. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Cases: ECM | Deaths: ECM |
|------------|------------|-------------|
|            | Min        | 0.016       | 0.016       |
|            | 5% prct    | 0.021       | 0.023       |
|            | 10% prct   | 0.026       | 0.027       |
|            | 25% prct   | 0.054       | 0.045       |
|            | Median     | 0.084       | 0.093       |
|            | 75% prct   | 0.095       | 0.099       |
|            | Max        | 0.115       | 0.124       |
|            | Mean       | 0.076       | 0.071       |
|            | Std        | 0.029       | 0.027       |

Table 18
Forecasting Results: Distribution of Exception Rates (95% Prediction Interval). The table presents results with respect to the prediction intervals produced by the ECM for both cases and deaths. The table shows descriptive statistics for the frequency that the absolute out-of-sample errors of the ECM model exceed the 95% prediction interval. The results are shown for forecasting horizons of 1 to 14 days ahead. The models were computed on a rolling window scheme with 28 in-sample observations per window.

| Days ahead | Cases: ECM | Deaths: ECM |
|------------|------------|-------------|
|            | Min        | 0.026       | 0.016       |
|            | 5% prct    | 0.029       | 0.023       |
|            | 10% prct   | 0.033       | 0.028       |
|            | 25% prct   | 0.049       | 0.050       |
|            | Median     | 0.059       | 0.061       |
|            | 75% prct   | 0.072       | 0.073       |
|            | Max        | 0.134       | 0.092       |
|            | Mean       | 0.059       | 0.060       |
|            | Std        | 0.015       | 0.016       |
A.2. Forecast combination

We consider a simple average combination of the forecasts from the ECM and AR models. The results are shown in Table 13. The table reports descriptive statistics for the ratio of the MAPEs of the combined models and the autoregressive benchmark. The upper panel presents the results for cases, and the lower panel considers forecasts of the number of deaths. Compared with the results in Tables 1 and 2, the combination of models reduces the MAPE ratios by more than 20% for cases and by more than 10% for deaths. The results for other losses are reported in Tables 14–16.

A.3. Prediction intervals

We report results concerning prediction intervals from the ECM in Tables 17–19. The tables report descriptive statistics for the frequency that the out-of-sample forecasting errors of the ECM violate the 90%, 95%, and 99% prediction interval, respectively. The prediction intervals are computed under the assumption that the errors are normally distributed. As we can see, the ECM produces very precise intervals for 95%. The intervals are slightly conservative for 90%. In the 99% case, the ECM underestimates the prediction bands. Overall, we believe that the results are satisfactory, but that they can be improved if the normality assumption is dropped and bootstrapping is used to compute the prediction intervals.

References

Atkeson, A. (2020). How Deadly is COVID-19? Understanding the difficulties with estimation of its fatality rate: Working paper 26965, National Bureau of Economic Research.

Bastos, S., & Cajueiro, D. (2020). Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil: Technical report, arXiv.org.

Berger, D., Herkenhoff, K., & Monegy, S. (2020). An SEIR infectious disease model with testing and conditional quarantine: Working paper 597, Federal Reserve Bank of Minneapolis.

Carroll, C., Bhattacharjee, S., Chen, Y., Dubey, P., Fan, J., Gajardo, A., et al. (2020). Time dynamics of COVID 19. Scientific Reports, 10.

Chimmula, V., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Coas. Solitons & Fractals, 135.

Coroneo, L., Iacone, F., & Monteiro, P. (2020). Testing the predictive accuracy of Covid-19 forecasts: Working paper 20/10, University of York.

da Silva, R., Ribeiro, M., Mariani, V., & Coelho, L. (2020). Forecasting Brazilian and American Covid-19 cases based on artificial intelligence coupled with climatic exogenous variables. Coas. Solitons & Fractals, 139.

Eichenbaum, M., Rebelo, S., & Trabandt, M. (2020). The macroeconomics of epidemics: Working paper 26882, National Bureau of Economic Research.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. Econometrica, 74, 1545–1578.

Hendry, J. D. J. C. D. (2022). Short-term forecasting of the coronavirus pandemic. International Journal of Forecasting, 38, 453–466.

Jiang, F., Zhao, Z., & Sha, X. (2020). Time series analysis of COVID-19 infection curve: A change-point perspective. Journal of Econometrics, (in press).

Kucharski, A., Russell, T., Diamond, C., Liu, Y., Edmunds, J., Funk, S., et al. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. The Lancet Infectious Diseases.
Li, S., & Linton, O. (2021). When will the Covid-19 pandemic peak? *Journal of Econometrics*, 220, 130–157.

Masini, R., & Medeiros, M. (2021). Counterfactual analysis with artificial controls: Inference, high dimensions and nonstationarity. *Journal of the American Statistical Association*, (in press).

Medeiros, M., & Mendes, E. (2016). $\ell_1$-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic innovations. *Journal of Econometrics*, 191, 255–271.

Petropoulos, F., Makridakis, S., & Stylianou, N. (2022). COVID-19: Forecasting confirmed cases and deaths with a simple time series model. *International Journal of Forecasting*, 38, 439–452.

Ribeiro, M., da Silva, R., Mariani, V., & Coelho, L. (2020). Short-term forecasting Covid-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, 135.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 267–288.

Walker, P., Whittaker, C., Watson, O., et al. (2020). The global impact of COVID-19 and strategies for mitigation and suppression: Technical report, Imperial College London.

Wooldridge, J. (2019). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.

Wu, J., Leung, K., & Leung, G. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCOV outbreak originating in Wuhan, China: A modelling study. *The Lancet*, 395(10225), 689–697.

Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-series data: A comparative study. *Chaos, Solitons & Fractals*, 140.