HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 20/16

Real Time Forecasting of Covid-19 Intensive Care Units demand

Paolo Berta; Pietro Giorgio Lovaglio;

Paolo Paruolo and Stefano Verzillo

August 2020

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Real Time Forecasting of Covid-19 Intensive Care Units demand[☆]

Paolo Berta[a,1], Pietro Giorgio Lovaglio[a,2], Paolo Paruolo[b,3], Stefano Verzillo[b,4]

[a] *University of Milan-Bicocca, Department of Statistics and Quantitative Methods, and CRISP*
[b] *European Commission, Joint Research Centre (JRC)*

Abstract

Response management to the SARS-CoV-2 outbreak requires to answer several forecasting tasks. For hospital managers, a major one is to anticipate the likely needs of beds in intensive care in a given catchment area one or two weeks ahead, starting as early as possible in the evolution of the epidemic. This paper proposes to use a bivariate Error Correction model to forecast the needs of beds in intensive care, jointly with the number of patients hospitalised with Covid-19 symptoms. Error Correction models are found to provide reliable forecasts that are tailored to the local characteristics both of epidemic dynamics and of hospital practice for various regions in Europe in Italy, France and Scotland, both at the onset and at later stages of the spread of the disease. The forecast performance is encouraging for all analysed regions, suggesting that the present approach may be useful also beyond the analysed cases.

Keywords: SARS-CoV-2, Covid-19, Intensive Care Units, Forecasting, Vector error correction model, VAR, Cointegration
JEL: C53, C32

[☆]Information and views set out in this paper are those of the authors and do not necessarily reflect the ones of the institutions of affiliation. The authors thank Camillo Rossi, MD (Healthcare Director at Spedali Civili di Brescia, Italy), for stimulating this research by requesting reliable short-run forecasts of Intensive Care Unit needs at the very beginning of the COVID-19 outbreak. The authors acknowledge useful comments on a previous version of the paper from Ian Vollbracht, Brian Doherty, and Helen Johnson from the European Commission and the ECDC respectively. The first two authors acknowledge financial support from the Region of Lombardy, project 2014IT16RFOP012 'Misura a sostegno dello sviluppo di collaborazioni per l'identificazione di terapie e sistemi di diagnostica, protezione e analisi per contrastare l'emergenza Coronavirus e altre emergenze virali del futuro'.
Declarations of interest: none.

[1]Email: paolo.berta@unimib.it ORCID: 0000-0003-0984-4288.
[2]Email: piergiorgio.lovaglio@unimib.it ORCID: 0000-0002-2340-0547.
[3]Email: paolo.paruolo@ec.europa.eu ORCID: 0000-0002-3982-4889, **Corresponding author.**
Address: European Commission, Joint Research Centre (JRC), Via Enrico Fermi 2749, TP 723, 21027 Ispra (VA), Italy
[4]Email: stefano.verzillo@ec.europa.eu ORCID: 0000-0002-1895-8554.

1. Introduction

The human-to-human transmission of SARS-CoV-2 has spread around the world, prompting the World Health Organization (WHO) to declare it a worldwide pandemic on March 12, 2020. Since the start of the pandemic, different regions have experienced rising clusters of cases and community transmission, inducing health authorities to introduce social-distance measures and various forms of lock-down. For instance, the Italian region of Lombardy experienced a rapid increase in Covid-19 cases in February 2020. Spanish regions, including Cataluña, experienced a similar evolution, with a delay of one or a few weeks.

Many forecasting problems are generated by the pandemic; the specific one analysed in the present paper is described in Subsection 1.1. Subsection 1.2 describes the corresponding data availability and its geographical scale. The contribution of the paper is presented in Subsection 1.3, which is next linked to the literature in Subsection 1.4.

The rest of the paper is organised as follows: Section 2 describes the type of data used to forecast Intensive Care Units (ICU) demand. Section 3 introduces the relevant class of models; Section 4 discusses model estimation and testing while Section 5 reports full-sample estimates. Section 6 discusses point and interval forecasts, while Section 7 contains the real-time application to weekly forecast of Covid-19 demand for ICU. Section 8 reports conclusions. Appendix A reports proofs.

1.1. Motivation

A major concern in the pandemic is the mounting pressure on the health care systems (Rodriguez-Llanes et al., 2020) due to the steep demand of sub-intensive units and ICU during the outbreak. The ECDC (2020) indicated 'bed occupancy in ICU' as the reference indicator of seriousness and spread of Covid-19, both at community or regional level and at hospital level.

In this context, there is a concrete need for public health managers to forecast ICU demand in real-time, with an horizon of one or two weeks, to plan or adjust health care resources. These actions may involve staff reallocation or re-deployment, the retrofitting or the creation of new wards for ICU or sub-intensive facilities with ventilators (Grasselli et al., 2020).

This forecast problem is addressed in the present paper and it is relevant at community or regional level, i.e. at the level where health management takes place. This corresponds to the geographical level of NUTS-2 regions in Europe. To showcase the proposed approach, data for some European regions were selected.

Italy was the first country to experience the Covid-19 outbreak, and its four most affected regions were included in the analysis. Other European countries experienced similar outbreaks, with a delay. Among these, French health departments provided consistent data in open access across time, and the Department of Paris has been chosen as a representative case. Similarly, data on Scotland has been provided consistently up to the last weeks of July 2020, and these were also selected for analysis.

1.2. Data

The pandemic has prompted public agencies to provide real-time data on the Covid-19 spread with open access. The data consist typically of daily time series of numbers of people swabbed, tested, infected, hospitalized or recovered, as well as Covid-19-related number of deaths. The data is usually provided at the sub-national level at which health management takes place.

Some of the reported indicators are prone to misreporting or selection bias. For instance, the number of deaths may be under-reported due to deaths that were untested for Covid-19, see e.g. Buonanno et al. (2020). The number of newly infected patients is of course affected by heterogeneity in testing policies within and among countries, see e.g. Richardson and Spiegelhalter (2020).

The number of patients (*IC*) admitted to Covid-19 ICU appears to be less subject to these sources of bias, as it is taken from admission records that apply well-defined hospital procedures. Possible anomalies may still be present, caused by lags in reporting, or mis-attribution of events to specific dates, such as during weekends and bank holidays.[5]

This Covid-19 demand for ICU reflects both the spread of Covid-19 as well as current hospital health practice, and policies put in place by Institutions; these may change during the evolution of the epidemic. Recent analyses, in fact, find that interventions such as lock-downs, social distancing, business closures, quarantine, face masks etc. not only directly affect the spread (growth rates) of the disease, but also indirectly affect its spread by changing people's behavior. Evidence on this has been provided in Hsiang et al. (2020), Zhang et al. (2020), Abaluck et al. (2020) and Chernozhukov et al. (2020).

The population at risk for the ICU bed occupancy is mostly made of the hospitalized patients with Covid-19 symptoms, labelled *HwS*. Data on *HwS* comes from admission records that reflect well-defined hospital procedures, and hence appears less prone to the sources of bias cited above, similarly to *IC*. *HwS* is also reported with the same frequency and spatial aggregation as *IC*. *IC* and *HwS* correspond to the Recognised and Threatened in the classical Compartmental/SIR-like methodology, see Adda (2016), Giordano et al. (2020) and Bollon et al. (2020).

This paper considers data on *IC* and *HwS* at sub-national level in various European countries as case studies. The most affected regions in Italy were Lombardy, Piemonte, Emilia Romagna, and Veneto; these Italian regions are analysed in the paper. Data on all Italian regions are published since February 24, 2020, by the Italian Civil Protection[6].

The paper validates the proposed approach also on Scottish data and French data from the health Department of Paris.[7] All data were last updated on August 7, 2020.

1.3. Contribution of the paper

This paper discusses the bivariate daily forecast for *IC* and *HwS* at regional level. It proposes (i) a model-based approach, based on Vector Error Correction models (VECM) i.e. cointegrated Vector Autoregressive models (VAR) (Davidson et al., 1978, Johansen, 1996), which take advantage of the joint dynamics of the time series. Moreover, (ii) it adapts the VECM current technology to accommodate the observed reality of high initial growth rates at the onset of the epidemic, that tend to persist over time. Additionally, (iii) the flexible deterministic components are considered in the VECM to allow for changes both in the diffusion of the epidemic and in hospital practice. Finally (iv) the paper proposes an adaptive approach that can be used to forecast *IC* with as little as 2 or 3 weeks of data. The proposed forecasting approach appears to work well on all data set analysed.

[5] Several examples of these kind of issues were for instance reported by the Italian media.

[6] https://github.com/pcm-dpc/COVID-19

[7] https://www.data.gouv.fr and https://statistics.gov.scot/.

Specifically, the VECM class of models accommodates the presence of a common trend between the two series, which corresponds to one cointegrating relation in the sense of Engle and Granger (1987). The cointegrating relation can be interpreted as determining an equilibrium level of *IC* as a function of the associated population at risk *HwS*. More specifically, this relation takes the form

$$IC_t^\star = \gamma_t HwS_t^\omega, \tag{1}$$

where * indicates the equilibrium level. Here the 'pass-through' coefficient γ_t may vary over time as a result to changes in the epidemic and in hospital practice. The parameter ω satisfies the equality $\omega = \partial \log IC_t^\star / \partial \log HwS_t$, and it is hence dubbed here the 'elasticity' parameter in (1), even if cointegrating coefficients do not necessarily correspond to elasticity parameters in the economic sense, see Johansen (2005).

Both the pass-through coefficient γ_t and the elasticity parameter ω reflect (a) the diffusion of the epidemic as well as (b) the hospital admission protocols of patients into ICU, and they are likely to differ between different regions and periods.

The presence of an equilibrium relation of the type (1) implies that *HwS* is a natural candidate predictor for *IC*, in the sense that (1) implies Granger causality at least of one of the two variables *IC* and *HwS*, see e.g. Granger (2004). The model-based approach allows one to test the hypothesis that this is the case empirically.

In addition to adjustment to the equilibrium relation of the type (1), the prediction of *IC* may be influenced by the lagged values of the growth rates of *IC* and *HwS*. This dynamic adjustment may incorporate some dynamic components with long half-time, which may be responsible for the slow decay in growth rates experienced at the onset of the outbreak.

The flexible deterministic components in the present model can accommodate changes in the evolution of the epidemic and medical practice. This paper also shows how the implied point and interval forecasts that can used to predict turning points (peaks and troughs). All these features are potentially heterogeneous across regions; a model fitting approach by region allows one to adapt these tools to the different situations.

The above characteristics of the present approach innovate with respect to other approaches in the literature, which, as discussed in the following subsection, focus on univariate models and techniques.

1.4. Related literature

Forecasts of demand of *IC* in the literature are based on several approaches. One of them is based on managerial strategies, such as scheduling and allocative rules, see Sadki et al. (2013) for a review. A different one applies probabilistic rules, such as discrete event simulation and queuing models, based on simulated scenarios (Ridge et al., 1998, Zhu et al., 2012) with Poisson models (Pearson et al., 2012, Milne and Whitty, 1995); they predict *IC* as a function of population size plus seasonal and organizational factors. All these approaches are applied to data at facility level, such as an emergency department or ICU.

When patients microdata are available in a given facility, a different approach uses classification and regression trees to predict *IC* for as a function of patients' case-mix, admission patterns, lengths of stay, ICU available capacity (Costa et al., 2003, Harper and Shahani, 2002) and refusal rate (Harper and Shahani, 2002). Other authors propose state-transition models such as Markov

modelling and Monte Carlo simulations (Kreke et al., 2004) that predict when cohorts switch from ordinary hospitalizations to ICU treatment.

In a time series framework, daily ICU demand has been modelled using univariate approaches; these include classical exponential and damped-trend smoothing, SARIMA models (Aboagye-Sarfo et al., 2015, Angelo et al., 2017, Kadri et al., 2014, Zhu et al., 2015) and artificial neural network models (Buzaev et al., 2016, Jilani et al., 2019).

Rarely, however, demand for ICU or for treatment in Emergency Departments has been modelled with a multivariate time series approach, with a few notable exceptions. Jones et al. (2009) model jointly arrivals, inpatients, computed tomography orders and other variables by means of VAR models, albeit without allowing for cointegration. Aboagye-Sarfo et al. (2015) compare vector-ARMA (VARMA) forecast models for predicting emergency department (ED) demand in Western Australia with univariate ARMA and other forecasting rules, and conclude that multivariate models outperform univariate ones.

During the Covid-19 pandemic, the timeliness of forecast of ICU demand has become more and more important. Most papers focus on infected, deaths and recovered patients. Examples include Jombart et al. (2020), Petropoulos and Makridakis (2020), Benvenuto et al. (2020), Peracchi (2020), Remuzzi and Remuzzi (2020), Grasselli et al. (2020), Deasy et al. (2020), which focus on predicting the inception phase of the epidemic only. No other paper has yet apparently adopted the present approach to jointly model the evolution of ICU demand and hospitalized patients with Covid-19 symptoms in a time series context.

Several of these real-time models foresee re-specification to adapt to different evolution of the pandemic, which makes them ill-suited for forecast comparisons. Adaptive methods with regular models update include Egidi and Torelli (2020), Davis and Fard (2020) using Poisson generalized linear mixed models, StatGroup-19 (2020), with Poisson autoregressions, Shoukat et al. (2020) using negative-binomial models or IHME COVID-19 health service utilization forecasting team (2020) with non-linear mixed effects curve-fitting models.

StatGroup-19 (2020) proposed to forecast *IC* by means of an ensemble model, with each component being an univariate model (INteger valued AutoRegressive, INAR model with Poisson innovations). In addition, Branas et al. (2020) and USCOVID (2020) publish daily forecasts of new confirmed cases, deaths and patients needing ICU care in US States using a meta-population Susceptible-Exposed-Infectious-Recovered (SEIR) model.

The approach in this paper, instead, rests in a pure multivariate time series approach (VECM), which can be readily codified, and hence lends itself to forecast comparisons. Early results on the application of the present approach to the first seven weeks of data up to mid-April 2020 for Italian regions, Switzerland and Spain are reported in Berta et al. (2020).

2. The data

This section describes the type of data analyzed, both for Italian regions and for Scotland and the Department of Paris. Let $x_t = (ic_t, hws_t)' = (\log(IC_t), \log(HwS_t))'$ and denote with $\Delta = 1 - L$ the difference operator, with L the lag operator, $Lx_t = x_{t-1}$. Fig. 1 graphs the time series x_t (natural logarithm of counts) for the four Italian regions for the subsets of the consecutive days in February 24 - August 5, 2020 for which $IC > 5$. Fig. 2 reports the same graph for the Scotland and the French health Department of Paris. All data were last downloaded on August 7, 2020.

The data covers different calendar periods. Data for Italy started on February 24. Some regions like Piemonte, did not have any person in ICU in the beginning; some other regions, like Veneto, had a few or zero cases in ICU since late July. Because of this, data were here selected so as to cover consecutive periods of days with $IC > 5$. The resulting periods are February 24 to August 6 for Lombardy, March 3 to July 24 for Piemonte, February 25 to June 1 for Veneto, February 27 to July 23 for Emilia Romagna.

The epidemic started later in France and Scotland. The corresponding data cover the periods March 18 to July 21 for Scotland and March 18 to August 5 for the Department of Paris.

The (first) peak of IC_t is attained at different times; for Lombardy, this happened at $t = 40$ corresponding to April 3 with $IC_t = 1381$; the first peak for HwS_t is at $t = 42$ corresponding to April 5 with $HwS_t = 12009$ and a second peak is at $t = 51$ corresponding to April 14 with $HwS_t = 12077$.

For all the shown cases, the levels of the two time-series x_t show a common trend. This appears a common feature, that goes beyond the six data sets analysed here.[8] The time evolution in some cases appears to change, possibly reflecting changes in treatment practice and in the spread of the disease.

The first difference Δx_t are reported for some regions in Fig. 3. The first differences Δx_t show persistent declining behaviour in the first part of the sample, corresponding to the initial outbreak of the COVID-19 epidemic.

The second differences $\Delta^2 x_t$ are reported for Lombardy in the left panel of Fig. 4; the remaining regions are similar to Lombardy. The second differences do not show persistent behaviour, but possibly time-varying heteroskedasticity.

A first question regards the level of integration of x_t. A common rule of thumb in the Box-Jenkins tradition is to choose the order of differencing for which the standard deviation of the series is lowest. Approximating the standard deviation with the sample range of the time series, this rule can be applied to the scales of Fig. 1, 3, 4; this rule would suggest that x_t is integrated of order 1, indicated $x_t \sim I(1)$.

The assumption $x_t \sim I(1)$ does not explain the persistent declining behaviour of Δx_t in the first part of the sample in Fig. 3. However, this behaviour would be consistent with Δx_t being stationary – i.e. $x_t \sim I(1)$ – but starting from an initial value that lies far from its average. As an illustration of this, the right panel in Fig. 4 reports one simulation of a univariate Autoregressive (AR) process with mean c, $w_t - c = \rho(w_{t-1} - c) + \varepsilon_t$ with $\rho = 0.94$, $c = -0.1$, $w_0 = 0.6$, where ε_t are independent and identically distributed (i.i.d.) $N(0, \sigma^2)$ random variables, with $\sigma = 0.01$.

The chosen initial value $w_0 = 0.6$ is way above the average of w_t equal to $c = -0.1$, and the process is pushed towards its mean $c = -0.1$ (return-to-the-mean) over a rather long stretch of time. The speed of the return-to-the-mean transition is governed by ρ: the larger the value of ρ, the longer the time the process takes for the return-to-the-mean; this behaviour is also described as a 'large half-life', see e.g. Fanelli and Paruolo (2010) are references therein. The observed behaviour of Δx_t in Fig. 3 appears consistent with a large real autoregressive root.

Formal univariate unit root test on the two time series in x_t were not performed for a number of

[8] Some early occasional look at data from Switzerland and regions in Australia showed the same feature. A full analysis of these other data sets was not pursued, not only because of limitations in data access or changes in definitions of variables, but also because the present paper does not attempt to cover all possible (sub-)national cases.

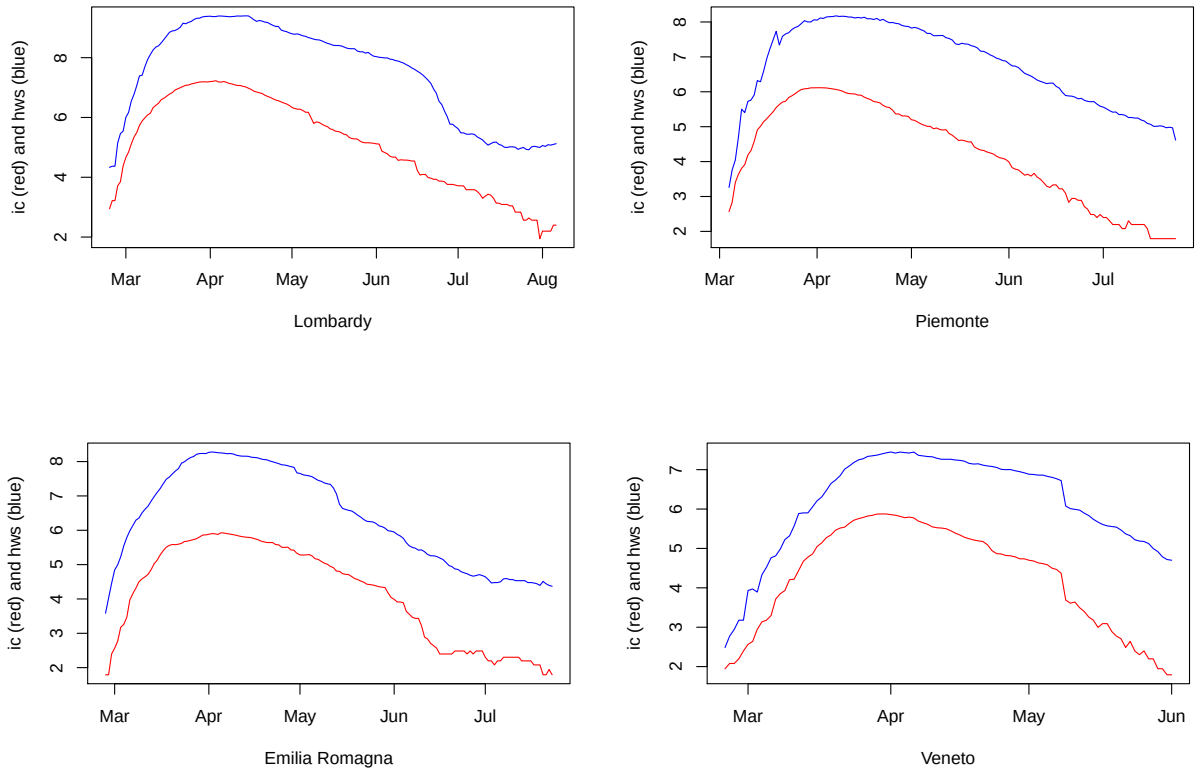

Figure 1: Italian data for $ic = \log(IC)$ and $hws = \log(HwS)$. Lombardy (upper left), Piemonte (upper right), Emilia Romagna (lover left), Veneto (lower right). Sample periods are the subsets of the consecutive days in February 24 - August 5, 2020 for which $IC > 5$. Red: ic. Blue: hws.

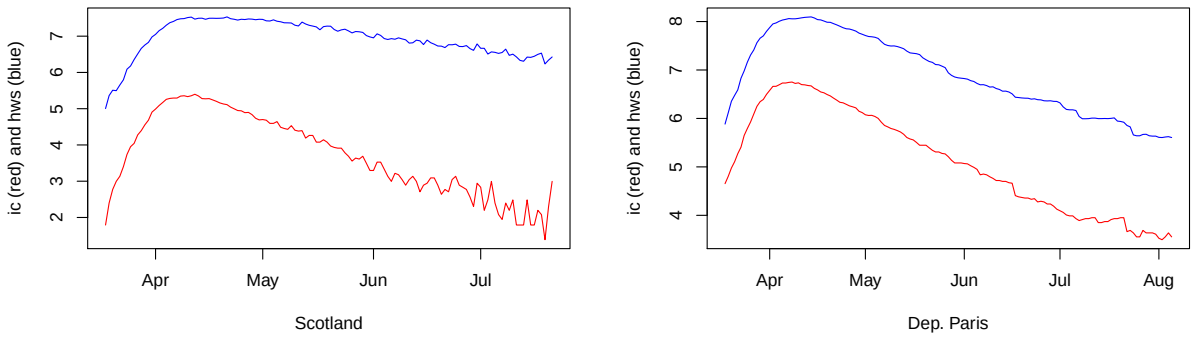

Figure 2: Scottish and French data for $ic = \log(IC)$ and $hws = \log(HwS)$. Left: Scotland. Right: Department of Paris. Red: ic. Blue: hws.

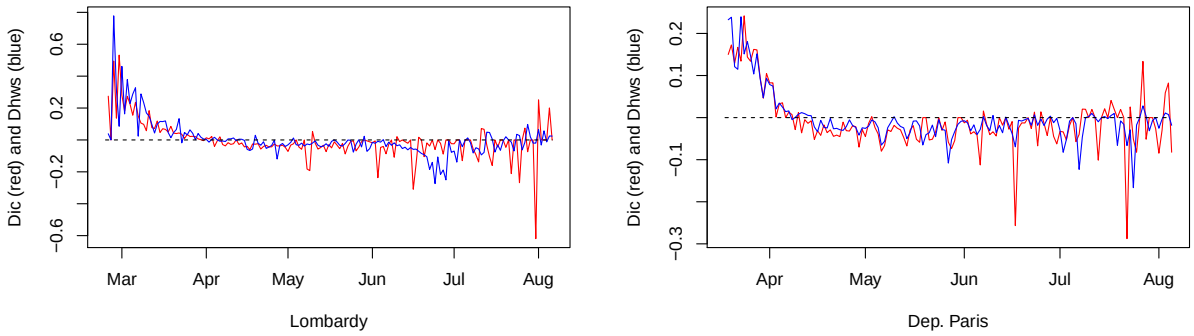

Figure 3: Data for $\Delta ic = \Delta \log(IC)$ and $\Delta hws = \Delta \log(HwS)$. Left: Lombardy. Right: Dep. of Paris. Red: *ic*. Blue: *hws*.

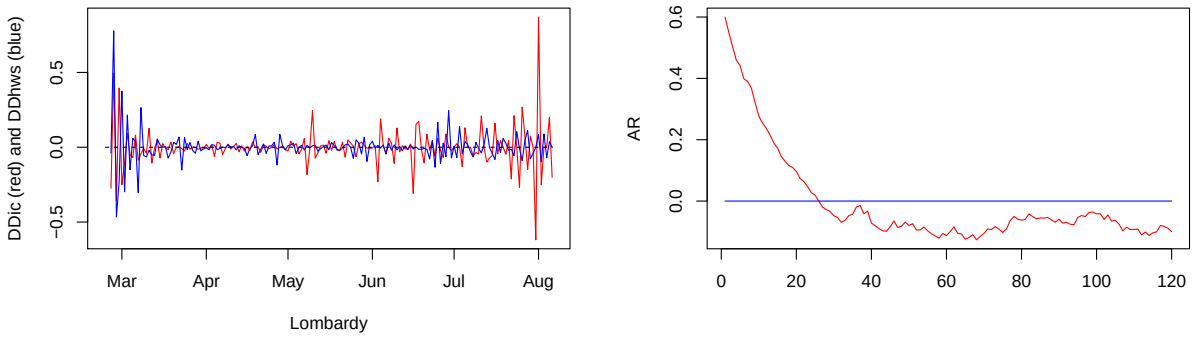

Figure 4: Left: data for $\Delta^2 ic = \Delta^2 \log(IC)$ and $\Delta^2 hws = \Delta^2 \log(HwS)$ for Lombardy, February 24 - August 6, 2020. Red: *ic*. Blue: *hws*. Right: simulation of a univariate AR process with mean c, $w_t - c = \rho(w_{t-1} - c) + \varepsilon_t$ with $\rho = 0.94$, $c = -0.1$, $w_0 = 0.6$, ε_t i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.01$.

reasons. The first one is that the approach proposed in this paper is multivariate, and as suggested in Johansen (1996), this would rule out univariate unit root tests. A second reason not to perform these tests is that the declining behaviour in the first part of the sample, corresponding to the initial outbreak of the COVID-19 epidemic, would probably distort test results. Overall, the proposed approach is to analyse the data as being I(1) and to use the resulting forecasts.

A second question is the possible presence of breaks in the mechanisms generating these data. Some likely anomalies in the ic_t time series can be observed for all regions. It is possible that hospital behaviour evolved along the epidemic, with a different pass-through of hospitalised COVID patients into ICU care as time goes by. Other likely anomalies may arise because of reporting lags and mis-dating of events; several of these events were reported in the press in Italy for instance. All these events may cause the presence of breaks in the process, especially in the deterministic components.

The approach of this paper is to model x_t as a Vector AutoRegressive (VAR) process with I(1) variables and with a common trend and a Vector Error Correction VECM representation. In order to accommodate the slow return-to-the-mean of Δx_t from an initial value away from its mean (see the right panel of Fig. 4) the I(1) VAR is expected to present a large AR root and initial conditions away from the mean. Moreover, breaks in the deterministic components of the processes are expected to play a role. The relevant class of processes that can be used to incorporate these features is briefly summarised in the next section.

3. Cointegrated VAR

This section recalls the essential elements of VAR and VECM with I(1) variables, see Hendry (1995) and Johansen (1996). It also discusses breaks in the deterministic components in line with Clements and Hendry (1996) and Johansen et al. (2000) in relation to the observed features of data shown in Section 2.

3.1. VECM

Let x_t be the $p \times 1$ vector of variables, satisfying the equations $A(L)x_t = \mu D_t + \varepsilon_t$ with $A(L) = I - \sum_{i=1}^{k} A_i L^i$; x_t is called a Vector AutoRegressive process of order k, VAR(k). Here k is the number of lags, A_i are $p \times p$ matrices of coefficients, D_t is an $n_D \times 1$ vector of deterministic variables (such as an intercept, time trend, seasonal or intervention dummy variables), ε_t is a $p \times 1$ vector white noise with mean 0 and positive definite variance-covariance matrix Ω. A special case is when ε_t is assumed to be i.i.d. Gaussian;[9] this assumption is made explicit when made in the following.

The AR characteristic polynomial $\det A(z)$ is assumed to have zeros outside the unit disc or at $z = 1$; this is labelled the 'unit and stable roots' assumption in the following. A necessary and sufficient condition for $\det A(z)$ to have a root at $z = 1$ is that $A(1)$ is singular, where $A(1) = I - \sum_{i=1}^{k} A_i$. This unit root is responsible for the random-walk behaviour of the process. Under the 'unit and stable roots' assumption, one hence has $A(1) = -\alpha\beta'$, where α and β are $p \times r$ matrices, with $0 \leq r < p$[10]. Note that in the present case of $p = 2$, r can be 0 or 1. Under the $A(1) = -\alpha\beta'$, the process for x_t can be rewritten in the form

$$\Gamma(L)\Delta x_t = \alpha\beta' x_{t-1} + \mu D_t + \varepsilon_t, \qquad \Gamma(L) = I - \sum_{i=1}^{k-1} \Gamma_i L^i \tag{2}$$

where $\Gamma_i = -\sum_{j=i+1}^{k} A_j$, as can be verified by re-arranging terms. Johansen (1991) showed that a necessary and sufficient condition for x_t to be I(1) under the 'unit and stable roots' assumption is that

$$\operatorname{rank}(\alpha'_\perp \Gamma \beta_\perp) = p - r, \tag{3}$$

where $\alpha_\perp$ (respectively $\beta_\perp$) is a basis of the orthogonal complement with respect to the space spanned by the columns of α (respectively β) and $\Gamma = \Gamma(1) = I - \sum_{i=1}^{k-1} \Gamma_i$; see Johansen (1991, 1996). For this reason condition (3) is called the 'I(1) condition'.

[9] This assumption may not be strictly valid for measurements x_t that contain logs of count variables.

[10] When discussing properties of the process, it is convenient to assume that α and β are of full column rank.

Under the 'I(1) condition' (3), it is shown that β' defines r a stationary linear combinations among the non-stationary variables x_{t-1} (corrected for D_t), i.e. that the r rows in β' define r cointegrating relations in the sense of Engle and Granger (1987). Eq. (2) defines the adjustment of Δx_t that correct the system toward equilibrium via the adjustment coefficients α. A process x_t satisfying (2), the 'unit and stable roots' assumption as well as the 'I(1) condition' is called an I(1) vector Error (or Equilibrium) Correction process (VECM).

The I(1) VECM process can be inverted into the so-called 'common trends' (CT) representation for x_t of the form

$$x_t = C \sum_{i=1}^{t} (\varepsilon_i + \mu D_i) + C_1(L)(\varepsilon_t + \mu D_t) + a, \qquad C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp \tag{4}$$

with $\beta' a = 0$ where a depends on k initial values of x_t.[11] Taking differences of (4) one obtains the MA representation for Δx_t,

$$\Delta x_t = C(\varepsilon_t + \mu D_t) + C_1(L)(\Delta \varepsilon_t + \mu \Delta D_t). \tag{5}$$

where $C_1(z)$ is an infinite order matrix polynomial convergent for all z within the a circle with radius $1 + \delta$ with $\delta > 0$.

The 'common trends' representation (4) shows that the p time series x_t share the $p - r$ common stochastic trends (random walks) $\alpha'_\perp \sum_{i=1}^{t} \varepsilon_i$, and that the r linear combinations $\beta' x_t$ are stationary (around their mean), because $\beta' C = 0$ in (4). The 'common trends' representation also shows that $\beta' x_t$ are cointegrating relations and β is the matrix of cointegration coefficients; $\beta' x_t - \mathbb{E}(\beta' x_t)$ represent deviations from equilibrium.

Eq. (4) and (5) can be used to calculate the expectations $\mathbb{E}(\cdot)$ of x_t, $\beta' x_t$ and Δx_t; this gives

$$\mathbb{E}(x_t) = C\mu \sum_{i=1}^{t} D_i + C_1(L)\mu D_t + a, \qquad \mathbb{E}(\beta' x_t) = \beta' C_1(L)\mu D_t, \tag{6}$$

$$\mathbb{E}(\Delta x_t) = C\mu D_t + C_1(L)\mu \Delta D_t, \tag{7}$$

where $\beta' C = 0$, $\beta' a = 0$ from (4).

The coefficient $C_1 := C_1(1)$ can be expressed in terms of the coefficients in the VECM equation (2) as

$$C_1 = -\bar{\beta}\bar{\alpha}' - C\Gamma\bar{\beta}\bar{\alpha}' - \bar{\beta}\bar{\alpha}'\Gamma C - C\Gamma\bar{\beta}\bar{\alpha}'\Gamma C - C\Phi C \tag{8}$$

where $\bar{a} = a(a'a)^{-1}$ for any full column rank matrix a, $\Gamma := \Gamma(1) = I - \sum_{i=1}^{k-1} \Gamma_i$ and $\Phi := \sum_{i=1}^{k-1} i\Gamma_i$, see eq. (7) in Johansen (2009) or eq. (3.8) in Franchi and Paruolo (2019). When $\mu_1 = \alpha\rho_1$, one finds $C_1\mu_1 = -(I + C\Gamma)\bar{\beta}\rho_1$ from (8).

3.2. Deterministic components

D_t may contain any type of deterministic terms; the specifications considered in this paper allow for broken linear trends, along the lines proposed in Johansen et al. (2000). Denote the

[11] Namely $\{x_t\}_{t=-k+1}^{0}$. For a proof see Theorem 4.2 in Johansen (1996), where $\beta' x_t$ is assigned its stationary distribution.

sample as $1, \ldots, T$. The deterministic terms D_t are split in two subsets, $D_t = (D_t^{(1)\prime}, D_t^{(2)\prime})'$ where $D_t^{(1)}$ is taken to include a constant 1 and a linear trend t or both $(1, t)$ interacted with period dummies $\mathbb{1}(t > T_1)$. Here $\mathbb{1}(\cdot)$ denotes the indicator function, which takes value 1 when the argument is true and 0 otherwise, and T_i are in-sample dates, $1 < T_i < T$.

As in Assumption 2.2 in Kurita et al. (2011)[12], define q breaks $1 < T_1 < \ldots, T_q < T$ and, for each break, consider the linear trend variable $(t - T_i)\mathbb{1}(t > T_i) = (t - T_i)_+$, which takes value 0 until $t = T_i$ and values $1, 2, \ldots$ starting from $t = T_i + 1$ onwards. A changing trend slope at T_i can be represented as a linear combination of the elements in $t_{\text{vec}} = (t, (t - T_1)_+, \ldots, (t - T_q)_+)'$. The first difference of t_{vec} is indicated here as $1_{\text{vec},t} = \Delta t_{\text{vec}}$, and it contains the period dummies $1_{\text{vec},t} = (1, \mathbb{1}(t > T_1), \ldots, \mathbb{1}(t > T_q))'$. It is simple to see that $\sum_{i=1}^{t} 1_{\text{vec},i} = t_{\text{vec}}$.

The second set $D_t^{(2)}$ within D_t is taken to contain de-meaned cyclical deterministic components and other dummies that do not generate trends when cumulated, i.e. $|\sum_{i=1}^{t} D_i^{(2)}| < c$ for all t for some constant $c > 0$. Examples include day-of-week dummies $s_{i,t} = \mathbb{1}(t \mod 7 = i) - \frac{1}{7}$, $i = 1, \ldots, 7$, or blip dummies $d_{i,t} = \Delta\mathbb{1}(t = i)$. Note that these variables are bounded when cumulated, where for instance $0 = \sum_{j=1}^{7} s_{i,j}$ and hence $|\sum_{j=1}^{t} s_{i,j}| < \frac{6}{7}$ is a bounded sequence.[13] Therefore, the variables in $D_t^{(2)}$ do not generate trends when cumulated.

3.3. Trends

This section describes the deterministic trends induced in the process by the chosen specification of D_t. Partition $\mu = (\mu^{(1)}, \mu^{(2)})$ conformably with $D_t = (D_t^{(1)\prime}, D_t^{(2)\prime})'$, and $\mu^{(1)} = (\mu_0, \mu_1)$ conformably with $D_t^{(1)} = (1'_{\text{vec},t}, t'_{\text{vec}})'$. Next observe that

$$\mathbb{E}(x_t) = C\mu_1 \sum_{i=1}^{t} i_{\text{vec}} + (C\mu_0 + C_1\mu_1)\, t_{\text{vec}} + (C_1\mu_0 + C_2\mu_1)\, 1_{\text{vec},t} + c_{1t}, \tag{9}$$

$$\mathbb{E}(\beta' x_t) = \beta' C_1\mu_1 t_{\text{vec}} + (\beta' C_1\mu_0 + \beta' C_2\mu_1)\, 1_{\text{vec},t} + c_{2t} \tag{10}$$

$$\mathbb{E}(\Delta x_t) = C\mu_1 t_{\text{vec}} + (C\mu_0 + C_1\mu_1)\, 1_{\text{vec},t} + c_{3t}, \tag{11}$$

where c_{it} are such that $|c_{it}| < c$ for all t,[14] and C_1 and C_2 are defined by $C_1(L) = C_1 + C_2(1 - L) + C_3(L)(1 - L)^2$. Eq. (9) shows that $\mathbb{E}(x_t)$ contains broken quadratic trends unless $C\mu_1 = 0$, eq. (10) shows that $\mathbb{E}(\beta' x_t)$ contains linear unless $\beta' C_1\mu_1 = 0$ and eq. (11) shows that $\mathbb{E}(\Delta x_t)$ contains broken linear trends unless $C\mu_1 = 0$. In view of the interpretation of Δx_t as a stationary process with initial value away from the average, the last restriction $C\mu_1 = 0$ appears of interest.

In this context, Johansen (1996) (eq. (5.13) to (5.17)) and Johansen et al. (2000) discussed the type of restrictions on (μ_0, μ_1) that generate different trends, as in (9), (10), (11).[15] In line with this approach, the notation $d(i, j; q)$ is employed here to indicate different classes of VECM processes, i.e. different models; details are given in the following. The integer i in $d(i, j; q)$ is associated with the degree of the polynomial in the cointegration relation, while the integer j relates to the one in the adjustment equations, with $i \geq j$. The last integer q indicates the number of breaks; when

[12] The break dates denoted T_i here correspond to dates $T_i + 1$ in Kurita et al. (2011).

[13] In the following, the notation $|c_t| < c$ is used to denote bounded deterministic sequences c_t, possibly different for different expressions.

[14] $c_{3t} = \Delta c_{1t}$

[15] The same approach is used in Kurita et al. (2011) in a I(2) VECM setting.

	Two stages EG		VECM		Model in Johansen (1996)
model	first stage	second stage	levels	differences	symbol and eq. therein
$d(0,0;q)$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$H_2(r)$ eq. (5.17)
$d(1,0;q)$	$1_{\text{vec},t}$	$\varnothing$	1_{vec}	$\varnothing$	$H_1^*(r)$ eq. (5.16)
$d(1,1;q)$	$1_{\text{vec},t}$	1_{vec}	$\varnothing$	$1_{\text{vec},t}$	$H_1(r)$ eq. (5.15)
$d(2,1;q)$	$(1'_{\text{vec},t}, t'_{\text{vec}})'$	$1_{\text{vec},t}$	t_{vec}	1_{vec}	$H^*(r)$ eq. (5.14)
$d(2,2;q)$	$(1'_{\text{vec},t}, t'_{\text{vec}})'$	$(1'_{\text{vec},t}, t'_{\text{vec}})'$	$\varnothing$	$(1'_{\text{vec},t}, t'_{\text{vec}})'$	$H(r)$ eq. (5.13)

Table 1: Models $d(i, j; q)$ and their correspondence to models in Johansen (1996).

$q = 0$, the simplified notation $d(i, j; 0) = d(i, j)$ is used. The correspondence of the models $d(i, j; q)$ with the models in Johansen (1996) section 5.7 is provided in Table 1.

More specifically, if $\mu_1 \neq \alpha\rho_1$ for any non-zero ρ_1, then $\mathbb{E}(x_t)$ is seen to contain a broken quadratic trend, see (9). In terms of models, this case corresponds to μ_0 and μ_1 unrestricted, which is labelled here as $d(2, 2; q)$, where q indicates the number of breaks in $D_t^{(1)}$.

If $\mu_1 = \alpha\rho_1$ for some non-zero ρ_1 one has that $\mathbb{E}(x_t)$ contains a broken linear trend, $\mathbb{E}(\beta' x_t)$ also contains a broken linear trend and $\mathbb{E}(x_t)$ contain a (broken) constant. This case is labeled $d(2, 1; q)$. If $\mu_1 = 0$ and $\mu_0 \neq \alpha\rho_0$ for any $\rho_0 \neq 0$ one has that $\mathbb{E}(x_t)$ contains a broken linear trend but $\mathbb{E}(\beta' x_t)$ does not; again $\mathbb{E}(\Delta x_t)$ contains a broken constant. This case is labeled $d(1, 1; q)$ in the following. Note that the difference between the $d(2, 1; q)$ and the $d(1, 1; q)$ cases lies with the (broken) linear trend in $\mathbb{E}(\beta' x_t)$ or not. This appears relevant in the present application, where the cointegrating relation may present a broken linear trend.

If $\mu_1 = 0$ and $\mu_0 = \alpha\rho_0$ for some non-zero ρ_0 one has that $\mathbb{E}(x_t)$ and $\mathbb{E}(\beta' x_t)$ contain a broken constant, and $\mathbb{E}(\Delta x_t)$ has a zero mean, up to a bounded remainder c_t. This case is labeled $d(1, 0; q)$ in the following. The final case, labeled $d(0, 0)$ corresponds to $\mu_1 = \mu_0 = 0$ for which $\mathbb{E}(x_t)$, $\mathbb{E}(\beta' x_t)$ and $\mathbb{E}(\Delta x_t)$ have zero mean, up to a bounded remainder c_t. Note that in this case the number of breaks q is irrelevant.

Remark that the present context, the cases $d(2, 1; q)$ and $d(1, 1; q)$ are of special interest; for these cases, in fact, $\mathbb{E}(\Delta x_t)$ is a (broken) constant.

4. Estimation and testing

Classes of processes of the type reviewed above need to be taken to the data. Subsections 4.1 and 4.2 discuss model estimation, while Subsection 4.3 provides interpretation for the cointegrating relation. Subsection 4.4 discusses tests on the adjustment coefficients.

4.1. Inference

Inference based on the two stage approach in Engle and Granger (1987) (EG) is justified as follows: the first stage cointegrating regression is super consistent, see Stock (1987). In the second stage, inference is asymptotically normal, with standard errors computed as for known cointegrating coefficients being consistent.

Maximum Likelihood (ML) in model (2) corresponds to looking for the canonical linear combination of the levels x_{t-1} that is most correlated with Δx_t, see Johansen (1991). Because in the

first half of the sample one observes a downward movement in Δx_t, it was found that ML estimates of $\beta' x_{t-1}$ tend to reproduce the downward movements in Δx_t.

While ML is asymptotically the most efficient method in large samples, in the present context of a small sample with downward movements in Δx_t in its beginning, it was found not to be optimal for the forecasting angle of this paper.

4.2. EG

Note that the term $\beta' x_{t-1}$ in (2) cointains r (in the present case $r = 1$) cointegrating relations among x_{t-1}, and α measures adjustment towards equilibrium. The long-run equilibrium (cointegrating) relation $\beta' x_{t-1}$, properly corrected for D_t, is stationary. In the present case of $r = 1$, the long-run equilibrium relation β' can be normalised setting the coefficient to $x_{1,t-1}$ equal to 1, $\beta' = (1, -\beta_2)'$, and the cointegrating relation reads $\beta' x_{t-1} = x_{1t-1} - \beta_2' x_{2,t-1}$. Hence, correcting for D_t,

$$x_{1,t-1} = \beta_2' x_{2,t-1} + \varphi' H_1' D_t + w_{t-1}. \tag{12}$$

The β_2 coefficients (only one coefficient in the present case) describe the proportionality factor between x_{1t} and x_{2t} in the long-run, where in (2) x_{1t} and x_{2t} may be trending (non-stationary) while their deviation w_{t-1}, called 'ecm', is stationary. In (2), H_1' is a design matrix that can select a subset of D_t in $H_1' D_t$.

The EG first stage is performed by estimating (12) by regression of x_{1t-1} on $x_{2,t-1}$ and $H_1' D_t$; this generates $\widehat{\beta}' x_{t-1}$ or, equivalently $\widehat{w}_{t-1}$. These quantities are substituted in (2) in the second stage, which estimates the remaining parameters by regression of Δx_t on $\widehat{w}_{t-1}$, Δx_{t-j} for $j = 1, \ldots, k-1$ and $H_2' D_t$, where H_2' is a design matrix that selects a subset of D_t in $H_2' D_t$.

The specification of the deterministic terms in $H_i' D_t$, $i = 1, 2$, gives the rationale for the notation $d(i, j; q)$, where i refers to the first stage and j to the second one. Consider first i; when $i = 2$, both the constant and the trend t are included in $H_1' D_t$, while when $i = 1$ only the constant is included. Finally when $i = 0$, no deterministic component is included in (12). Hence, the index i is associated to the degree $i - 1$ of the polynomial fitted in the first stage. In case $q > 0$, a number of break points in the constant and trend are allowed for.

The same convention is used for j in relation to $H_2' D_t$ in the second stage, i.e. the index j is associated to the degree $j - 1$ of the polynomial fitted in the second stage.

4.3. Interpretation

When $x_{1,t} = ic_t$ (the log of IC_t) and $x_{2,t} = hws_t$ (the log of HwS_t), the interpretation of the coefficients in the cointegrating relation is as follows. The cointegrating relation (12) reads $ic_{t-1} = ic^\star_{t-1} + w_{t-1}$ where $ic^\star_{t-1} = \beta_2\, hws_{t-1} + a_0' 1_{\text{vec}} + a_1' t_{\text{vec}}$ for $i = 2$, where for $j = 1$, $a_1 = 0$ and for $j = 0$, $a_1 = a_2 = 0$. Taking exponentials,

$$IC^\star_{t-1} = \gamma_t HwS^{\beta_2}_{t-1} \qquad \text{where} \qquad \gamma_t = \exp(a_0' 1_{\text{vec}} + a_1' t_{\text{vec}}). \tag{13}$$

Here γ_t is the 'pass-through' coefficient from $HwS^{\beta_2}_{t-1}$ to $IC^\star_{t-1}$ and $\omega = \beta_2$ is the 'elasticity' coefficient. γ_t is allowed to be time-varying, reflecting changing practice in hospitals in the transition to ICU admission.

Eq. (13) is a relation of semi-proportionality when $0 < \beta_2 < 1$ i.e. $IC^\star$ moves less than proportionally with respect to HwS. On the contrary, when $\beta_2 = 1$, $IC^\star$ is directly proportional to HwS.

4.4. Adjustment

The adjustment eq. (2) describes how $\Delta x_{1,t} = \Delta ic_t$ and $\Delta x_{2,t} = \Delta hws_t$ adjust to deviations from equilibrium w_{t-1} (equilibrium-correction), as well as to past changes. In particular, the equation for the i-th (block of) variable(s), reads

$$\Delta x_{i,t} = \alpha_i \beta' x_{t-1} + \sum_{\ell=1}^{k-1} \Gamma_{ij,\ell} \Delta x_{j,t-\ell} + \sum_{\ell=1}^{k-1} \Gamma_{ii,\ell} \Delta x_{i,t-j} + c_i' \mu D_t + \varepsilon_{i,t} \tag{14}$$

where c_i is an appropriate selection matrix. The following hypotheses can be considered for parameters in eq. (14)

$$H_{01}: \quad \alpha_i = 0, \tag{15}$$

$$H_{02}: \quad \Gamma_{ij,\ell} = 0, \quad \ell = 1, \ldots, k-1 \tag{16}$$

If both hypotheses hold, there is lack of Granger causality of $x_{j,t}$ for $x_{i,t}$. In this case $x_{i,t}$ can be interpreted as 'autonomous' i.e. as not reacting to past values of x_{jt}, and it can be predicted on the basis of its own past only.

5. Full sample estimation

This section reports estimation of the model on the full-sample data described in Section 2. The following Section 7 reports the real-time forecasting exercise. The first 7 days of data were used as 'burn-in' period and the last 7 days at the end of the sample were kept for out-of-sample validation. This left observations from $T_0 = 8$ (corresponding to March 2) to observation $T_{q+1} = 114$ (corresponding to June 16, 2020) as full estimation sample.

Indicate the estimation sample as $t = T_0, \ldots T_{q+1}$, and let $b = (T_1, \ldots, T_q)'$ be the vector of possible q break dates. Denote also the sum of squares residuals from the EG first stage regression (12) as $S(b)$, which is a function of the break dates b. Remark that the first stage regression is the same for models $d(2, j; q)$, which correspond to a single $S(b)$ for all $j = 0, 1, 2$. Following Zivot and Andrews (1992), b was selected so as to minimize $S(b)$, namely

$$(T_1, \ldots, T_q) = \underset{b \in U_q}{\arg\min}\, S(b), \tag{17}$$

where U_q is an appropriate set of break date vectors. In practice this set was chosen as follows, denoting by g the number of gap days between breaks

$$U_q = \{(b_1, \ldots, b_h, \ldots, b_q) : b_{h-1} + g + 1 \leq b_h \leq b_h^{\max}, h = 1, \ldots, q\}, \tag{18}$$

where $b_0 = T_0$ and $b_h^{\max} = T_{q+1} - (q - h + 1)g + \mathbb{1}(h \neq q)$. Note that $b_1 < \cdots < b_h < \cdots < b_q$ in U_q, and that the spacings between breaks in U_q are at least equal to g gap days, $b_h - b_{h-1} > g$. In the implementation the values $q = 0, 1, 2$ were considered.[16] In the case of Lombardy, $q = 4$ was chosen, see below, and break dates were set without recourse to the minimisation in (17).

[16] Too small values of g would result in perfectly collinear regressors: for example, if $q = 1$ and $1_{\text{vec},t} = (1, \mathbb{1}(T > T_1))'$, $t_{\text{vec}} = (t, (t - T_1)\mathbb{1}(T > T_1))'$ choosing $T_1 = 1$ would generate perfect collinearity, because in this case $(t - T_1)\mathbb{1}(T > T_1) + 1 - t = 0$.

Region	$\widehat{\omega} = \widehat{\beta}_2$	$\min \widehat{\gamma}_t$	mean $\widehat{\gamma}_t$	$\max \widehat{\gamma}_t$	break dates	ρ_2, ρ_3, ρ_4
Lombardy	0.5344	0.21	3.71	10.18	38, 75, 120, 145	$0.92, 0.68 \pm 0.31i$
Emilia Romagna	0.8728	0.15	0.27	0.44	78, 106	$0.85, 0.76 \pm 0.20i$
Piemonte	0.4851	2.0	5.4	13.8	28, 116	$0.82, 0.64 \pm 0.41i$
Veneto	0.9482	0.052	0.257	0.441	19, 68	$0.86, -0.33 \pm 0.57i$
Scotland	1.3988	8.7e-05	3.2e-03	1.0e-02	92, 108	$0.13 \pm 0.77i, 0.70$
Dep. of Paris	1.0750	0.095	0.131	0.168	71, 112	$0.83, 0.59 \pm 0.20i$

Table 2: Elasticity $\widehat{\omega} = \widehat{\beta}_2$, pass trough estimates $\widehat{\gamma}_t$ and eigenvalues ρ_i of the companion matrix $\widehat{A}$ for the various regions. ρ_i are ordered such that $1 = |\rho_1| \geq |\rho_2| \geq \ldots, |\rho_s|$.

Consider the full samples described in Section 2; model $d(2, 1; q)$ was fitted to the full estimation sample, excluding 7 days of data at the beginning for 'burn-in' period, and 7 days of data at the end of out-of-sample forecast. Except for Lombardy, q was set for all regions equal to 2 and the break dates were found by minimisation of (17).

The identified break dates are reported in Table 2. Lombardy seems to have undergone more than 2 breaks; q was hence set equal to 4, and the break dates were identified as follows. In an early stage of the writing of this paper, approximately 100 observations were available, and the identified break dates were found to be 38, 75 by (17). As more data became available, the additional break dates 120, 145 were added.

The EG first stage regressions for the four Italian regions are reported in Fig. 5 and the ones of Scotland and the Dept. of Paris are reported in Fig. 6. The estimated elasticity parameters $\omega = \beta_2$, and the pass-trough coefficient γ_t are summarised in Table 2. It can be seen that the estimated elasticity parameters $\omega = \beta_2$ ranges from 0.48 to 1.4. The elasticity parameters and the pass-trough coefficient γ_t appear to compensate, in the sense that lower values of $\omega = \beta_2$ are associated with higher values of γ_t and vice versa.

In the specification of the second EG stage, no short term dummies $D_t^{(2)}$ were added, and the coefficients of the period dummies $\mathbb{1}(t > T_i)$ were constrained to 0, leaving only the constant in 1_{vec} unrestricted. Table 2 reports the obtained largest eigenvalues ρ_i of the companion matrix A in eq. (19), see the following section. It can seen that, usually, the estimated system have a large persistent root close to 0.9, as anticipated in Section 2.

Table 3 reports the estimates of the adjustment coefficients α_i, $i = 1, 2$ in eq. (14), as well as the list of lags ℓ for which the $\Gamma_{ij,\ell}$ coefficients were significant at 5% level for both adjustment equations in eq. (14), see (15) and (16). It is seen that the error-correction coefficient α_1 is always strongly significant, implying that Δic_t adjusts to disequilibrium errors. In a couple of regions, also α_2 is marginally significant at 5% level, implying that also Δhws_t adjusts to disequilibrium errors.

Only two lags ℓ are identified for Emilia Romagna and for Scotland for which Δic_t is significantly influenced by $\Delta hws_{t-\ell}$, see (15). Vice versa, the lags ℓ for which Δhws_t is significantly influenced by $\Delta ic_{t-\ell}$ are rather abundant. One would hence conclude that $(ic_t, hws_t)'$ are jointly dependent, i.e., neither variable appears 'autonomous'. Focusing on the Δic_t equation, the ability of hws_{t-j} to predict Δic_t appears to come mainly from the 'ecm' term w_{t-1}.

Overall, the $d(2, 1; q)$ model shows a good fit on the full estimation sample data. The following sections summarize the forecasting characteristics of this class of models.

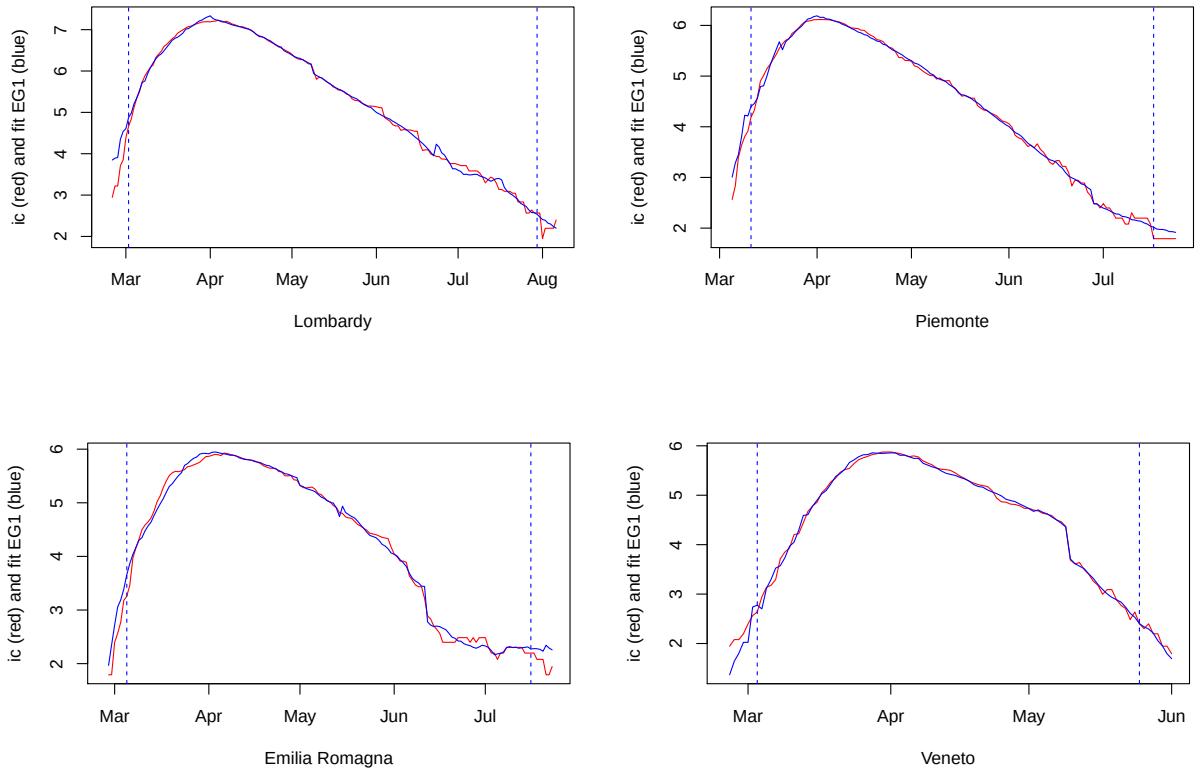

Figure 5: First stage EG regression for Italian regions. Red: *ic*. Blue: fitted values from (12). Estimation sample between vertical dashed lines.

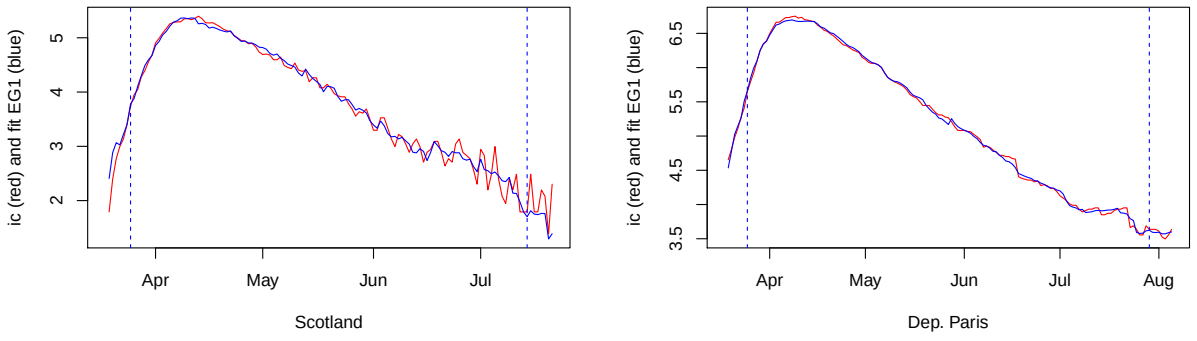

Figure 6: First stage EG regression for Italian regions. Red: *ic*. Blue: fitted values from (12). Estimation sample within vertical dashed lines.

			sig.lag ℓ in Δic_t eq.		sig.lag ℓ in Δhws_t eq.	
Region	$\widehat{\alpha}_1$(pvalue)	$\widehat{\alpha}_2$(pvalue)	$\Delta ic_{t-\ell}$	$\Delta hws_{t-\ell}$	$\Delta ic_{t-\ell}$	$\Delta hws_{t-\ell}$
Lombardy	−0.256 (0e-04)	0.1645 (6e-04)	1,3			1,2,3
Emilia Romagna	−0.1807 (6e-04)	0.025 (0.4631)	1, 3	2	1	1
Piemonte	−0.5586 (0e-04)	0.1685 (0.1265)	1,2,3		3	2,3
Veneto	−0.757 (0.0022)	0.1542 (0.6627)			2	
Scotland	−0.9146 (2e-04)	0.1448 (0.1265)	3	3	2,3	2,3
Dep. of Paris	−0.5497 (0e-04)	−0.1434 (0.0447)	1,2		2	1

Table 3: Adjustment equations. Estimates of α_i, $i = 1, 2$ and list of lags ℓ for which the $\Gamma_{i,j,\ell}$ coefficients were significant at 5% level for both adjustment equations in eq. (14), see (15), (16).

6. Forecasting

The VECM class of processes is a subset of the VAR class, whose forecasting properties are well known both in the stationary and non-stationary cases, see Lütkepohl (2005). This section summarises how point and interval forecast are derived, see Subsection 6.2, together with predictors for turning points, see Subsection 6.3, making use of the companion form of the VAR and of the VECM, see Subsection 6.1. Appendix A proves claims on their relation. Finally, indices of point forecast performance are reviewed in Subsection 6.4, while Subsection 6.5 discusses of performance indicators for interval forecasts.

The notation $y_{t+h|t}$ is used to denote the point forecast $y_{t+h|t} = \mathbb{E}_t(y_{t+h})$ where $\mathbb{E}_t(\cdot)$ indicates expectation conditional on all observed information on the past of the process x_t, i.e. the sigma field generated by $\{x_{t-s}, s \geq 0\}$. Here y_{t+h} denotes any random variable at time $t+h$, like $y_{t+h} = \Delta x_{t+h}$ or $y_{t+h} = x_{t+h}$.

6.1. The companion form

This subsection describes how (2) can be expressed in companion form. Define the state vector $X_t = (\Delta x_t', x_{t-1}', \Delta x_{t-1}', \ldots, \Delta x_{t-k-2}')'$ with dimension $s \times 1$, $s = pk$, and observe that (2) can be written as

$$X_t = AX_{t-1} + J_{s,p}(\mu D_t + \varepsilon_t)$$

i.e.

$$\begin{pmatrix} \Delta x_t \\ x_{t-1} \\ \Delta x_{t-1} \\ \Delta x_{t-2} \\ \vdots \\ \Delta x_{t-k+2} \end{pmatrix} = \begin{pmatrix} \Gamma_1 + \alpha\beta' & \alpha\beta' & \Gamma_2 & \ldots & \Gamma_{k-2} & \Gamma_{k-1} \\ I_p & I_p & & & & \\ I_p & & & & & \\ & & I_p & & & \\ & & & \ddots & & \\ & & & & I_p & \end{pmatrix} \begin{pmatrix} \Delta x_{t-1} \\ x_{t-2} \\ \Delta x_{t-2} \\ \Delta x_{t-3} \\ \vdots \\ \Delta x_{t-k+1} \end{pmatrix} + J_{s,p}(\mu D_t + \varepsilon_t), \tag{19}$$

where $J_{s,p}$ with $s \geq p$, indicates the first p columns of I_s, of the form $J_{s,p} = (I_p, 0)'$, and A is called the companion matrix of the AR process x_t. The eigenvalues of A are the reciprocals of the roots of $\det(A(z)) = 0$, which are assumed to be equal to 1 or with modulus greater than 1.[17]

[17] For a proof of this, see e.g. Johansen (1996) proof of Theorem 2.2.

The state-space X_t in (19) is not minimal. In fact a reduced state-space vector W_t can be defined, which contains only the stationary part of X_t, namely $W_t = (\Delta x_t', x_{t-1}'\beta, \Delta x_{t-1}', \ldots, \Delta x_{t-k-2}')'$. The reduced state-space vector W_t is of dimension $q = s - (p - r)$, and one has $W_t = G'X_t$ where $G = \text{diag}(I_p, \beta, I_{(k-1)p})$. Appendix A shows that W_t satisfies the companion form $W_t = BW_{t-1} + J_{q,p}(\mu D_t + \varepsilon_t)$ with $B = G'A\bar{G}$. The set of eigenvalues of A is shown to be the union of the set of eigenvalues of B and 1 (taken $p - r$ times).

6.2. *Forecasts*

Set $v_t = J_{s,p}\mu D_t$ and $u_t = J_{s,p}\varepsilon_t$, so that eq. (19) reads $X_{t+1} = AX_t + v_{t+1} + u_{t+1}$. One finds

$$X_{t+h} = A^h X_t + \sum_{i=0}^{h-1} A^i (v_{t+h-i} + u_{t+h-i}) \tag{20}$$

and $X_{t+h|t} := \mathbb{E}_t(X_{t+h}) = A^h \mathbb{E}_t(X_t) + \sum_{i=0}^{h-1} A^i \mathbb{E}_t(v_{t+h-i} + u_{t+h-i}) = A^h X_t + \sum_{i=0}^{h-1} A^i v_{t+h-i}$. $X_{t+h|t}$ can be computed recursively as $X_{t+j|t} = AX_{t+j-1|t} + v_{t+j}$ with $j = 1, \ldots, h$, starting at $j = 1$ with $X_{t|t} = X_t$.

Interest lies in the forecast of the level and the first differences $y_{t+h} := (\Delta x_{t+h}', x_{t+h}')'$. Indicate their point forecast as $y_{t+h|t} := (\Delta x_{t+h|t}', x_{t+h|t}')'$; these can be retrieved from $X_{t+h|t}$ as follows

$$y_{t+h|t} = \begin{pmatrix} \Delta x_{t+h|t} \\ x_{t+h|t} \end{pmatrix} = M' X_{t+h|t}, \qquad M' = K' J_{s,2p}', \qquad K' = \begin{pmatrix} I_p & 0 \\ I_p & I_p \end{pmatrix} \tag{21}$$

where $J_{s,2p}' X_{t+h|t} = (\Delta x_{t+h|t}', x_{t+h-1|t}')'$. The forecast variance-covariance matrix for $y_{t+h} := (\Delta x_{t+h}', x_{t+h}')'$, denoted by Σ_h, is found to be

$$\Sigma_h = \sum_{i=0}^{h-1} V_i \Omega V_i' = \Sigma_{h-1} + V_{h-1} \Omega V_{h-1}', \qquad V_i = M' A^i J_{s,p}, \tag{22}$$

where $\Omega = \mathbb{E}(\varepsilon_t \varepsilon_t')$.[18] This expression follows from (20) and (21) using the equality $X_{t+h} - X_{t+h|t} = \sum_{i=0}^{h-1} A^i u_{t+h-i}$.

Forecast intervals can be constructed as $a' y_{t+h|t} \pm \kappa \sqrt{a' \Sigma_h a}$ where a is any pre-specified selection vector. For instance a can be chosen to the j-th column e_j of the I_{2p} to select variable number j within $y_{t+h|t}$. If the errors are Gaussian, choosing for instance $\kappa = 2$ (respectively 3) gives a 95.45% (respectively 99.73%) coverage rate for these forecast intervals. If errors are not Gaussian, Chebyshev's inequality can be used to derive a conservative bound for the coverage of these intervals, $\Pr(|a' y_{t+h|t} / \sqrt{a' \Sigma_h a}| > \kappa) < \kappa^{-2}$. For instance, for $\kappa = 2$ (respectively 3) the coverage is at least $\Pr(|a' y_{t+h|t} / \sqrt{a' \Sigma_h a}| < \kappa) \geq 1 - \kappa^{-2} = 0.75$ (respectively 88.89%).

6.3. *Turning points*

The process (2) can generate peaks and troughs, i.e. turning points; this is also reflected in forecasts. For instance, the growth rates Δx_t in the first outbreak phase of the epidemic are usually

[18] As noted in Lütkepohl (2005), some of the elements in Σ_h may diverge as $h \to \infty$, as at least one of the eigenvalues of A is equal to 1. However, this is not a concern for any finite horizon h.

large and positive. The forecasts of Δx_t may predict a decline of future growth rates Δx_{t+h} to small positive and eventually negative rates; this generates the prediction of the first peak.

Recall that $(x_{1,t}, x_{2,t})' = (ic_t, hws_t)'$, and assume one wishes to forecast a downward turning point in variable i, such as $i = 1$. In particular, one can be interested in the first forecast horizon for which m consecutive predictions of $\Delta x_{i,t}$ are negative. In order to do so, consider all forecast horizons j at which there is a change of sign in the prediction of $\Delta x_{i,t}$, namely $C_{i,h} = \{j : \text{sign}(\Delta x_{i,t+j|t}) \neq \text{sign}(\Delta x_{i,t+j-1|t}), j = 1, 2, \ldots, h\}$. Given the set $C_{i,h}$, one can define the subset for which m consecutive forecasts of $\Delta x_{i,t}$ are negative

$$C_{i,h}^{m-} = \{j \in C_{i,h} : \text{sign}(\Delta x_{i,t+j|t}) = \cdots = \text{sign}(\Delta x_{i,t+j+m|t}) = -1\}, \qquad h_i^\star = \inf C_{i,h}^{m-} \tag{23}$$

and define the minimal horizon $h_i^\star = \inf C_{i,h}^{m-}$ at which this occurs. Note that, if m consecutive negative prediction of $\Delta x_{i,t+j+m|t}$ do not exist, then $\inf \varnothing = \infty$.

The interpretation of $h_i^\star$ is the first forecast horizon for which m consecutive predictions of $\Delta x_{i,t}$ are negative; this is hence a possible definition for turning point for $x_{i,t}$ during the upward trending phase. A similar definition applies for upward turning points in a decreasing trending phase where $C_{i,h}^{m+}$ is defined replacing -1 with $+1$ in the previous definitions. Here the focus is on the first downward turning point.

The horizon $h_i^\star$ in (23) depends of the companion matrices A or B in the companion form for W_t or X_t in (20) or (A.1), as well as on the last value W_t or X_t at which the forecast is made. In particular, if the expected growth rate in (7) is negative, long-term forecasts of $\Delta x_{i,t}$ will tend to the overall expectation $\mathbb{E}(\Delta x_{i,t})$, and this implies a certain forecast horizon $h_i^\star$ at which predictions of $\Delta x_{i,t}$ turn negative.

6.4. Accuracy of point forecasts

Let $y_{t+j|t}^{(i)}$ indicate the point forecast of y_{t+j} performed in t for model i over time periods $t+1, \ldots, t+h$. Several measures of ex-post accuracy for the point forecast can be considered, starting from the Mean Absolute Error (MAE). Following Makridakis et al. (2020), the symmetric mean absolute percentage error (sMAPE) and mean absolute scaled error (MASE) were also considered, along with the mean absolute percentage error (MAPE) and the mean arctangent absolute percentage error (MAAPE), see Kim and Kim (2016). The standard formulation of these indicators is the following

$$\begin{aligned}
\text{MAE}_i &= \frac{1}{h}\sum_{j=1}^{h} |y_{t+j} - y_{t+j|t}^{(i)}|, & \text{sMAPE}_i &= \frac{2}{h}\sum_{j=1}^{h} \frac{|y_{t+j} - y_{t+j|t}^{(i)}|}{|y_{t+j}| + |y_{t+j|t}^{(i)}|}, \\
\text{MAPE}_i &= \frac{1}{h}\sum_{j=1}^{h} \left|\frac{y_{t+j} - y_{t+j|t}^{(i)}}{y_{t+j}}\right|, & \text{MAAPE}_i &= \frac{1}{h}\sum_{j=1}^{h} \arctan\left|\frac{y_{t+j} - y_{t+j|t}^{(i)}}{y_{t+j}}\right|, \\
\text{MASE}_i &= \frac{1}{h}\frac{\sum_{j=1}^{h} |y_{t+j} - y_{t+j|t}^{(i)}|}{\frac{1}{t-n}\sum_{j=n+1}^{t} |y_j - y_{j-n}|},
\end{aligned} \tag{24}$$

For any of these indices, like MAE, one can consider their relative version with respect to the Random Walk model in the spirit of MASE. Specifically the following relative indices are considered here: $\text{MAER}_i = \text{MAE}_i/\text{MAE}_0$, $\text{MAPER}_i = \text{MAPE}_i/\text{MAPE}_0$, $\text{sMAPER}_i = \text{sMAPE}_i/\text{sMAPE}_0$, $\text{MAAPER}_i = \text{MAAPE}_i/\text{MAAPE}_0$.

In the definition of MASE, Makridakis et al. (2020) select $n = 1$ for daily data, as in the original proposal of Hyndman and Koehler (2006). In the present application, the standard transposition of MASE shown in (24) has some limitations. First, the idea in Hyndman and Koehler (2006) was to scale the absolute forecast error of the point forecast $y^{(i)}_{t+j|t}$ with the average absolute forecast error of a näive forecast based on the Random Walk model, here indexed as model 0. The above formula of MASE for $n = 1$ compares the absolute forecast error for a j-step ahead forecast $y^{(i)}_{t+j|t}$ with the one of a one-step ahead of the Random Walk model, hence losing comparability of the forecast horizons between numerator and denominator.

Secondly, the above formula of MASE for $n = 1$ compares the mean absolute error of the *out-of-sample* forecast $y^{(i)}_{t+j|t}$ in the numerator with the *in-sample* mean absolute error of the Random Walk model. Of course this comparison is appropriate when there is time-homogeneity between the in-sample and out-of-sample periods. This is not the case in the present application, where the mean absolute error of the Random Walk model decreases as time passes from the onset of the epidemic, because of the slow decline in growth rates in the first part of the epidemic, see Section 2. For these reasons, this indicator is not considered in the present application.

A possible solution to these shortcomings of MASE that provides a solution to the above limitations is given by $\text{MAER}_i = \text{MAE}_i/\text{MAE}_0$, where the average forecast error for model i, namely $\sum_{j=1}^{h} |y_{t+j} - y^{(i)}_{t+j|t}|$ is compared with the same average $\sum_{j=1}^{h} |y_{t+j} - y_t| = \sum_{j=1}^{h} |y_{t+j} - y^{(0)}_{t+j|t}|$ for the forecast error for the Random Walk model for the same forecast horizon j. In this sense, the forecast horizons are balanced in the numerator and denominator. Moreover, both forecast errors are computed out-of-sample, hence circumventing the slow decline in growth rates at the onset of the epidemic.

6.5. Accuracy of forecasts intervals

Consider the forecast intervals $(l_{i,t+j}, u_{i,t+j})$ built for y_{t+j} by model i at $1 - \alpha$ confidence level over horizons $j = 1, \ldots, h$. To measure the accuracy of forecast intervals, define the following indices

$$\text{AIS} = \frac{1}{h}\sum_{j=1}^{h} s_{i,t+j}, \qquad s_{i,j} = u_{i,j} - l_{i,j} + \frac{2}{\alpha}\left((l_{i,j} - y_j)\mathbb{1}(y_j < l_{i,j}) + (y_j - u_{i,j})\mathbb{1}(y_j > u_{i,j})\right) \tag{25}$$

$$\text{MSIS} = \frac{1}{\frac{1}{t-n}\sum_{j=n+1}^{t} |y_j - y_{j-n}|}\frac{1}{h}\sum_{j=1}^{h} s_{i,t+j}, \qquad \text{BASIS}_i = \frac{1}{h}\sum_{j=1}^{h}\frac{s_{i,t+j}}{s_{0,t+j}}$$

Here $s_{i,t+j}$ is the interval scoring proposed by Gneiting and Raftery (2007) for model i. The factor $2/\alpha$ given a penalty to deviations from the confidence interval. In the present context, the confidence level α was computed both under Gaussianity or using Checyshev's approximation.

AIS gives the Average Interval Score, while MSIS is standard transposition of the Mean Scaled Interval Score in Makridakis et al. (2020). The latter suffers from the same limitation of MASE. a possible modification of MSIS is introduced here, called BASIS for 'Balanced Average Scaled Interval Score', where the score for model i is compared with the one for the Random Walk model.

Comparing AIS_i with the value for the Random Walk model gives $\text{AISR}_i = \text{AIS}_i/\text{AIS}_0$. Note that BASIS is a measure already relative to Random Walk model, so that its relative version $\text{BASIS}_i/\text{BASIS}_0 = \text{BASIS}_i/1 = \text{BASIS}_i$ would give the same result.

7. Real-time forecasting

The forecasting problem is defined using daily data organized in weeks, assuming to observe data for $t = 1, \ldots, T$, where $T = 7n$ and n indicates weeks. The exercise consists in forecasting x_{T+j} for $j = 1, \ldots, 7$ using data up to x_T.

In order to mimic a real-time forecasting situation as closely as possible, the forecast exercise is started as soon as there is enough data available to compute estimates, namely $n = 2$ weeks of data.

No use of $D_t^{(2)}$ dummies was made, nor of judgemental choices on model selection in real time. This provides a fair comparison between models; moreover, it provides a lower bound on the performance of the models with respect to the one that can be obtained in real time. In real situation, in fact, additional forecaster's analysis and wisdom may improve the performance of the models.

More precisely, models $d(1, 1; q)$ and $d(2, 1; q)$ were fitted to the data for $t = 1, \ldots, 7n$ starting for week n, and forecasts were produced for the following week, namely for $t = 7n + 1, \ldots, 7n + 7$. The number of breaks q was set equal to 0 up to week 5, equal to 1 from week 6 to 10, and equal to 2 from week 11 onwards. Indices of forecast accuracy were then computed, providing a measure of forecast performance of the associated models. For weeks 2 and 3, no 'burn-in' period was used, while starting form week 4, the first week of data was discarded.

Fig. 7 reports the real-time forecasts for $ic = \log(IC)$ of model $d(2, 1; q)$ for Italian regions, while Fig. 8 is relative to Scotland and the health Department of Paris. Graphs for model $d(1, 1; q)$ were similar and showed a marginally lower forecast performance; they are not reported here for brevity. It can be seen that the forecast performance of weeks 2 and 3 is rather poor, possibly reflecting the large estimation uncertainty due to the small sample size. For later weeks the forecast performance is reasonable.

Tables 4 and 5 report MAE and MAER for models $d(1, 1; q)$ and $d(2, 1; q)$ respectively, for all regions. Recall that the values of MAE can be taken as an approximation to the MAPE for the count of *IC*; these table show that the relative percentage error for *IC* is substantial in the weeks 2 and 3, as high as 50 %. It is also usually higher than the Random Walk model. Starting from week 4 onwards, the relative percentage error for *IC* gets often lower that 10%. Comparing MAE and MAER values for $d(1, 1; q)$ and $d(2, 1; q)$, one can see that $d(1, 1; q)$ tends to do better up to week 7, but it is later dominated by $d(2, 1; q)$.

Tables 6 and 7 report AIS and AISR for models $d(1, 1; q)$ and $d(2, 1; q)$ respectively, for all regions. The patterns for this index are similar with respect to what observed for MAE and MAER above, with low performance in weeks 2 and 3 and improved performance thereafter. In some weeks where *IC* tends to stay constant, the Random Walk model outperforms the proposed models also in later weeks. Comparing AIS and AISR values for $d(1, 1; q)$ and $d(2, 1; q)$, one can see that $d(1, 1; q)$ tends to do better up to week 7, but it is later dominated by $d(2, 1; q)$. Similar patterns were found for the other forecast performance indicators considered above, such as sMAPE, MAAPE, BASIS, etc., which are not reported here for brevity.

Overall, the $d(1, 1; q)$ and $d(2, 1; q)$ models appears to have a reasonable performance in real-time.

	MAE						MAER					
week	L	ER	P	V	S	DP	L	ER	P	V	S	DP
2	0.64	0.12	0.36	0.14	0.59	0.33	1.62	0.25	1.03	0.20	2.04	1.74
3	0.09	0.50	0.12	0.24	0.10	0.06	0.35	4.73	0.67	0.67	1.78	1.65
4	0.06	0.62	0.02	0.02	0.15	0.07	0.62	5.22	1.33	0.13	1.91	0.49
5	0.04	0.05	0.14	0.04	0.19	0.03	3.26	2.51	1.39	0.69	1.09	0.24
6	0.04	0.01	0.19	0.08	0.11	0.09	0.56	0.16	1.47	0.43	0.65	0.60
7	0.10	0.01	0.27	0.09	0.18	0.10	0.76	0.11	1.22	0.43	0.89	0.53
8	0.01	0.02	0.11	0.17	0.23	0.11	0.08	0.13	0.82	0.67	0.98	0.60
9	0.03	0.05	0.05	0.08	0.31	0.09	0.17	0.32	0.50	0.74	1.23	0.83
10	0.12	0.04	0.27	0.46	0.33	0.04	0.59	0.23	1.06	0.94	2.63	0.26
11	0.13	0.41	0.25	0.86	0.35	0.12	0.79	3.14	1.21	2.04	1.08	0.73
12	0.05	0.20	0.14	0.06	0.08	0.06	0.37	1.50	0.77	0.12	0.61	0.64
13	0.02	0.08	0.16	0.23	0.17	0.17	0.21	0.32	1.03	0.78	1.02	0.58
14	0.15	0.33	0.37		0.21	0.04	0.56	0.84	1.30		0.72	0.33
15	0.20	0.69	0.33		0.57	0.03	2.24	1.18	1.06		1.06	0.26
16	0.25	0.15	0.30		0.28	0.26	0.63	2.96	0.97		0.87	5.92
17	0.18	0.06	0.04		0.45	0.11	1.57	1.15	0.16		0.82	1.65
18	0.28	0.11	0.34			0.35	3.82	1.22	2.96			1.11
19	0.09	0.12	0.23			0.08	0.71	2.55	0.63			1.62
20	0.16	0.13					0.63	0.64				
21	0.20						0.72					
22	0.15						0.75					

Table 4: MAE and MAER for model $d(1,1;q)$. Left: MAE. Right: MAER. L: Lombardy, ER: Emilia Romagna, P: Piemonte, V: Veneto, S: Scotland, DP: Department of Paris.

	MAE						MAER					
week	L	ER	P	V	S	DP	L	ER	P	V	S	DP
2	0.56	0.52	0.58	0.14	0.59	0.43	1.42	1.07	1.65	0.21	2.01	2.25
3	0.06	0.42	0.22	0.38	0.09	0.05	0.25	4	1.27	1.05	1.52	1.38
4	0.08	0.02	0.23	0.01	0.14	0.10	0.86	0.17	15.66	0.05	1.81	0.69
5	0.05	0.09	0.23	0.04	0.14	0.02	4.38	4.50	2.33	0.68	0.80	0.17
6	0.03	0.02	0.08	0.21	0.05	0.03	0.51	0.32	0.63	1.10	0.30	0.19
7	0.10	0.02	0.13	0.09	0.05	0.12	0.75	0.21	0.58	0.45	0.27	0.59
8	0.01	0.06	0.05	0.12	0.25	0.08	0.09	0.42	0.36	0.49	1.05	0.44
9	0.01	0.07	0.05	0.02	0.26	0.06	0.09	0.48	0.53	0.22	1	0.53
10	0.09	0.04	0.05	0.36	0.19	0.05	0.46	0.26	0.20	0.73	1.52	0.28
11	0.13	0.33	0.08	0.53	0.21	0.11	0.77	2.52	0.38	1.25	0.66	0.69
12	0.10	0.03	0.03	0.20	0.10	0.03	0.70	0.20	0.15	0.40	0.81	0.29
13	0.03	0.11	0.05	0.12	0.24	0.23	0.27	0.43	0.30	0.43	1.45	0.76
14	0.12	0.35	0.10		0.22	0.02	0.46	0.90	0.37		0.74	0.20
15	0.04	0.66	0.14		0.52	0.09	0.42	1.13	0.44		0.97	0.69
16	0.17	0.47	0.21		0.28	0.30	0.43	9.45	0.70		0.86	6.96
17	0.22	0.07	0.11		0.46	0.04	1.86	1.31	0.48		0.83	0.56
18	0.29	0.07	0.22			0.21	3.92	0.80	1.89			0.66
19	0.32	0.21	0.25			0.08	2.56	4.71	0.69			1.58
20	0.19	0.20					0.73	1.01				
21	0.08						0.29					
22	0.16						0.77					

Table 5: MAE and MAER for model $d(2,1;q)$. Left: MAE. Right: MAER. L: Lombardy, ER: Emilia Romagna, P: Piemonte, V: Veneto, S: Scotland, DP: Department of Paris.

	AIS						AISR					
week	L	ER	P	V	S	DP	L	ER	P	V	S	DP
2	2.97	0.75	1.27	0.70	3.76	1.92	1.51	0.39	0.87	0.43	2.39	1.69
3	0.25	3.20	0.59	1.12	0.46	0.22	0.21	2.16	0.53	0.71	0.45	0.29
4	0.20	3.10	0.55	0.55	0.39	0.24	0.20	2.55	0.59	0.41	0.45	0.39
5	0.19	0.49	0.47	0.51	0.52	0.24	0.22	0.46	0.59	0.44	0.69	0.44
6	0.18	0.46	0.44	0.40	0.37	0.25	0.23	0.48	0.61	0.38	0.54	0.50
7	0.30	0.42	0.88	0.46	0.51	0.32	0.41	0.49	1.31	0.47	0.78	0.69
8	0.20	0.39	0.42	0.51	0.75	0.29	0.29	0.49	0.64	0.56	1.08	0.63
9	0.20	0.39	0.44	0.45	1.14	0.25	0.30	0.51	0.70	0.52	1.54	0.57
10	0.60	0.42	0.89	2.56	1.18	0.26	0.84	0.58	1.51	1.09	1.72	0.60
11	0.32	1.86	0.76	4.25	1.11	0.34	0.50	2.62	1.32	3.96	1.14	0.80
12	0.29	0.49	0.46	0.68	0.69	0.32	0.47	0.72	0.76	0.53	0.88	0.76
13	0.32	0.50	0.48	1	0.90	0.81	0.52	0.72	0.84	0.97	1.06	0.68
14	0.41	1.23	1.43		0.88	0.33	0.55	0.93	1.91		0.80	0.75
15	0.68	3.96	1.06		2.44	0.33	1.13	1.52	1.10		1.22	0.77
16	1.01	0.64	0.90		0.91	0.99	0.62	0.87	1		0.63	2.34
17	0.43	0.61	0.47		2.18	0.37	0.69	0.85	0.76		0.96	0.88
18	0.88	0.63	1.45			1.72	1.43	0.87	2.36			1.30
19	0.50	0.64	0.83			0.44	0.83	0.90	0.74			0.96
20	0.55	0.63					0.87	0.90				
21	0.58						0.65					
22	0.69						0.75					

Table 6: AIS and AISRfor model $d(1,1;q)$. Confidence level computed using Chebyshev's inequality. L: Lombardy, ER: Emilia Romagna, P: Piemonte, V: Veneto, S: Scotland, DP: Department of Paris.

	AIS						AISR					
week	L	ER	P	V	S	DP	L	ER	P	V	S	DP
2	3.30	2.83	3.92	0.64	3.71	2.62	1.68	1.48	2.70	0.40	2.35	2.31
3	0.24	2.68	0.82	2.31	0.39	0.22	0.20	1.81	0.74	1.48	0.38	0.29
4	0.22	0.47	0.56	0.57	0.39	0.30	0.21	0.39	0.61	0.43	0.45	0.49
5	0.19	0.51	0.58	0.51	0.41	0.24	0.22	0.48	0.73	0.43	0.54	0.43
6	0.18	0.42	0.30	0.81	0.37	0.23	0.23	0.44	0.41	0.77	0.53	0.45
7	0.29	0.38	0.33	0.32	0.38	0.34	0.40	0.44	0.49	0.33	0.58	0.72
8	0.19	0.36	0.33	0.41	0.81	0.29	0.27	0.45	0.51	0.45	1.16	0.62
9	0.19	0.37	0.33	0.40	0.71	0.27	0.29	0.49	0.53	0.46	0.96	0.60
10	0.36	0.39	0.32	2.10	0.65	0.22	0.50	0.53	0.53	0.89	0.94	0.53
11	0.33	1.36	0.23	1.84	0.78	0.34	0.50	1.91	0.40	1.71	0.81	0.78
12	0.31	0.24	0.26	0.69	0.59	0.25	0.49	0.35	0.43	0.54	0.74	0.60
13	0.28	0.31	0.24	0.73	1.20	0.96	0.46	0.46	0.42	0.71	1.42	0.80
14	0.31	1.49	0.39		1.04	0.36	0.42	1.12	0.52		0.95	0.81
15	0.30	3.77	0.49		2.19	0.36	0.49	1.45	0.51		1.10	0.82
16	0.61	2.51	0.83		0.84	1.23	0.38	3.41	0.92		0.58	2.90
17	0.81	0.59	0.37		1.83	0.38	1.29	0.83	0.61		0.81	0.90
18	1.18	0.61	0.78			1.01	1.92	0.86	1.27			0.76
19	1.27	0.62	0.93			0.43	2.11	0.88	0.83			0.93
20	0.59	0.66					0.93	0.94				
21	0.31						0.35					
22	0.81						0.88					

Table 7: AIS and AISRfor model $d(2,1;q)$. Confidence level computed using Chebyshev's inequality. L: Lombardy, ER: Emilia Romagna, P: Piemonte, V: Veneto, S: Scotland, DP: Department of Paris.

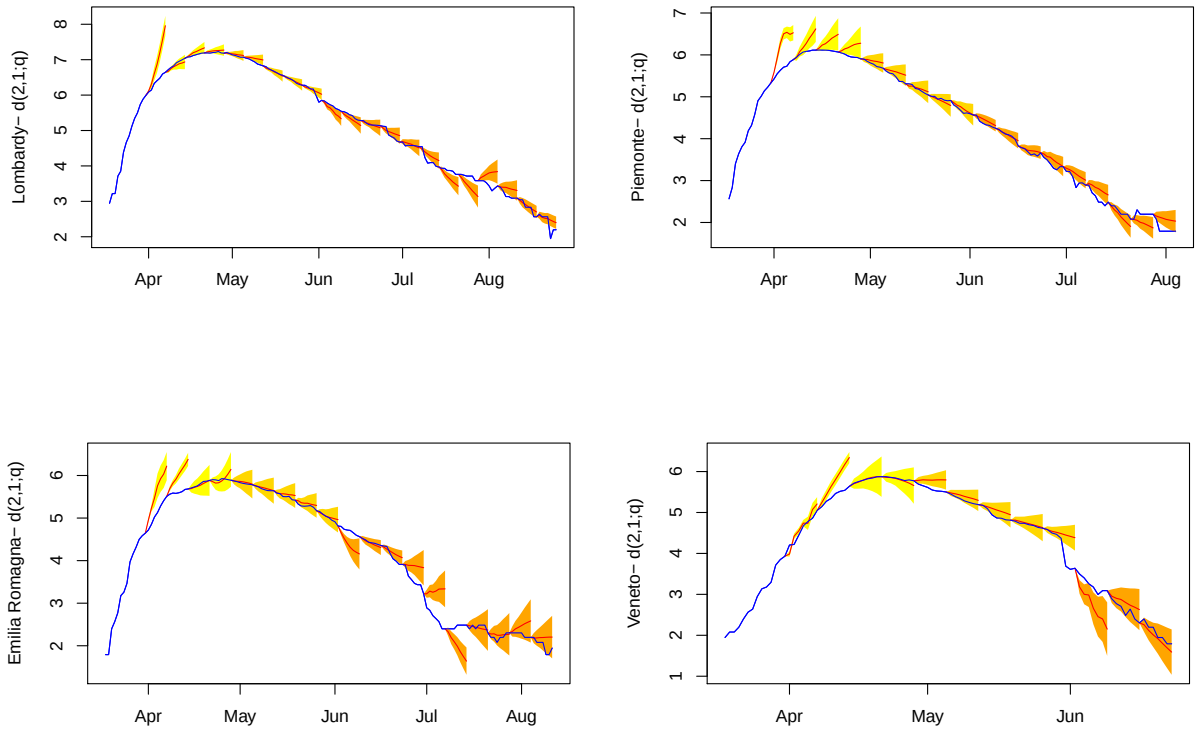

Figure 7: Real-time forecasts for $ic = \log(IC)$ of model $d(2,1;q)$ for Italian regions. Lombardy (upper left), Piemonte (upper right), Emilia Romagna (lover left), Veneto (lower right). Yellow: $q = 0$. Gold: $q = 1$. Orange: $q = 2$.

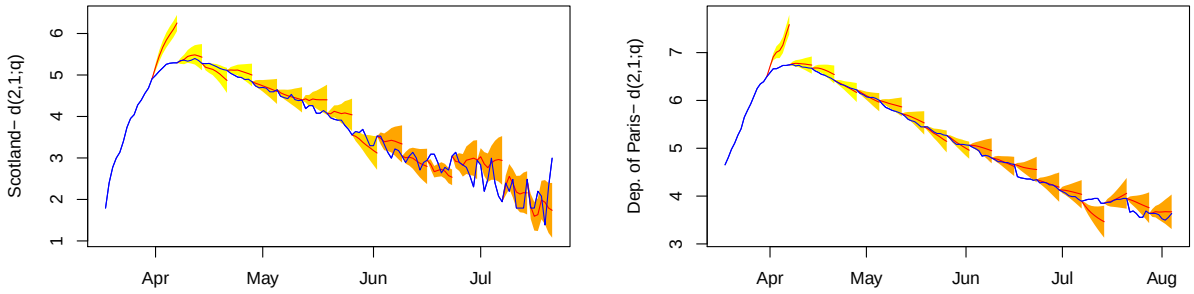

Figure 8: Real-time forecasts for $ic = \log(IC)$ of model $d(2,1;q)$ for Scotland and the Department of Paris. Left: Scotland. Right: Department of Paris. Yellow: $q = 0$. Gold: $q = 1$. Orange: $q = 2$.

8. Conclusions

Daily time series data of Covid-19 hospitalized patients and of Covid-19 patients admitted in ICU are used to forecast demand for ICU in real time. The proposed bivariate VECM model exploits the link between the series in levels, which reflects the fact that group of hospitalised patients is the natural population at risk for intensive care.

This link is found to translate empirically into a cointegrating relation. The associated parameters can be grouped in a pass through multiplicative coefficient and in an elasticity parameter. It is found that both the pass through coefficient and an elasticity parameter are region-specific, reflecting differences in health care response to the outbreak.

The VECM, endowed with broken linear trends, is able to capture changes in hospital practice, the effects of containment measures and the evolution of the epidemic. These broken deterministic components affect the pass through coefficient, and are found to be very relevant empirically.

These models were applied to the real-time forecast if ICU demand for the four most affected Italian regions; this analysis showed that the models can provide real-time reasonably reliable forecasts. The same findings are obtained for French data from the health Department of Paris and for Scotland. These results suggest that the proposed approach may be useful to predict Covid-19 related ICU demand for other heath-care systems beyond the ones analysed here.

The VECM models presented here do not consider issues relate to reduced availability of ICU beds due to the healthcare system reaching or even surpassing its capacity (Ji et al., 2020). In practice ICU capacity utilization never approached full capacity in the analysed regions, due to the creation of new ICU, possibly via the retrofitting of other hospital facilities; hence this does not appear to be a problem from the empirical perspective.

Further improvements of the approach may come from incorporating additional sources of information, such as ICU saturation or size, waiting times, lengths of stay in ICU, patient distributions by age and co-morbidities in the area etc. The multivariate nature of the VECM has the potential to accommodate the modeling of this additional information, which may also lower the need to recourse to broken deterministic components.

References

Abaluck, J., J. A. Chevalier, N. A. Christakis, H. P. Forman, E. H. Kaplan, A. Ko, and S. H. Vermund (2020). The case for universal cloth mask adoption and policies to increase supply of medical masks for health workers. Technical report, Available at SSRN 3567438.

Aboagye-Sarfo, P., Q. Mai, F. M. Sanfilippo, D. B. Preen, L. M. Stewart, and D. M. Fatovich (2015). A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. *Journal of Biomedical Informatics* 57, 62–73.

Adda, J. (2016). Economic activity and the spread of viral diseases: evidence from high frequency data. *The Quarterly Journal of Economics 131*(2), 891–941.

Angelo, S. A., E. F. Arruda, R. Goldwasser, M. S. Lobo, A. Salles, et al. (2017). Demand forecast and optimal planning of intensive care unit (icu) capacity. *Pesquisa Operacional 37*(2), 229–245.

Benvenuto, D., M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, https://doi.org/10.1016/j.dib.2020.105340.

Berta, P., P. Paruolo, S. Verzillo, and P. G. Lovaglio (2020). Adapting the health care system's response to the spread of Covid-19. Mimeo, submitted.

Bollon, J., M. Paganini, C. R. Nava, N. De Vita, R. Vaschetto, L. Ragazzoni, F. Della Corte, and F. Barone-Adesi (2020). Predicted effects of stopping Covid-19 lockdown on Italian hospital demand. *Disaster Medicine and Public Health Preparedness*, 1–15.

Branas, C. C., A. Rundle, S. Pei, W. Yang, B. G. Carr, S. Sims, A. Zebrowski, R. Doorley, N. Schluger, J. W. Quinn, et al. (2020). Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (sars-cov2) cases in us counties. *medRxiv*.

Buonanno, P., S. Galletta, and M. Puca (2020). Estimating the severity of Covid-19: Evidence from the Italian epicenter. Technical report, http://dx.doi.org/10.2139/ssrn.3567093.

Buzaev, I. V., V. V. Plechev, I. E. Nikolaeva, and R. M. Galimova (2016). Artificial intelligence: Neural network model as the multidisciplinary team member in clinical decision support to avoid medical mistakes. *Chronic Diseases and Translational Medicine 2*(3), 166–172.

Chernozhukov, V., H. Kasahara, and P. Schrimpf (2020). Causal impact of masks, policies, behavior on early Covid-19 pandemic in the US. *Covid Economics 35*(7 July 2020), https://cepr.org/file/9287/download?token=Ba6cin7P.

Clements, M. P. and D. F. Hendry (1996). Intercept corrections and structural change. *Journal of Applied Econometrics 11*(5), 475–494.

Costa, A., S. Ridley, A. Shahani, P. R. Harper, V. De Senna, and M. Nielsen (2003). Mathematical modelling and simulation for planning critical care capacity. *Anaesthesia 58*(4), 320–327.

Davidson, J. E. H., D. F. Hendry, F. Srba, and S. Yeo (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the united kingdom. *The Economic Journal 88*(352), 661–692.

Davis, S. and N. Fard (2020). Theoretical bounds and approximation of the probability mass function of future hospital bed demand. *Health Care Management Science 23*(1), 20–33.

Deasy, J., E. Rocheteau, K. Kohler, D. J. Stubbs, P. Barbiero, P. Liò, and A. Ercole (2020). Forecasting ultra-early intensive care strain from covid-19 in england. Technical report, https://europepmc.org/article/ppr/ppr118533.

ECDC (2020). Coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK? seventh update. Technical report, European Centre for Disease Prevention and Control, https://www.ecdc.europa.eu/sites/default/files/documents/RRA-seventh-update-Outbreak-of-coronavirus-disease-COVID-19.pdf.

Egidi, L. and N. Torelli (2020). Graphical Bayesian analysis of Covid-19 data across Italian regions. Technical report, Technical Report, https://www.leonardoegidi.com/covid-19.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica 55*(2), 251–276.

Fanelli, L. and P. Paruolo (2010). Speed of adjustment in cointegrated systems. *Journal of Econometrics 158*(1), 130–141.

Franchi, M. and P. Paruolo (2019). A general inversion theorem for cointegration. *Econometric Reviews 38*(10), 1176–1201.

Giordano, G., F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Granger, C. W. (2004). Time series analysis, cointegration, and applications. *American Economic Review 94*(3), 421–425.

Grasselli, G., A. Pesenti, and M. Cecconi (2020). Critical care utilization for the covid-19 outbreak in lombardy, italy: early experience and forecast during an emergency response. *JAMA 323*(16), 1545–1546.

Harper, P. R. and A. Shahani (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society 53*(1), 11–18.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford University Press, USA.

Hsiang, S., D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, A. Hultgren, L. Y. Huang, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, and T. Wu (2020). The effect of large-scale anti-contagion policies on the coronavirus (Covid-19) pandemic. Technical report, https://doi.org/10.1101/2020.03.22.20040642.

Hyndman, R. J. and A. B. Koehler (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting 22*(4), 679–688.

IHME COVID-19 health service utilization forecasting team (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. Technical report, MedRxiv, 26 March 2020, https://doi.org/10.1101/2020.03.27.20043752.

Ji, Y., Z. Ma, M. P. Peppelenbosch, and Q. Pan (2020). Potential association between COVID-19 mortality and health-care resource availability. *The Lancet Global Health 8*(4), http://doi.org/10.1016/S2214–109X(20)30068–1.

Jilani, T., G. Housley, G. Figueredo, P.-S. Tang, J. Hatton, and D. Shaw (2019). Short and long term predictions of hospital emergency department attendances. *International Journal of Medical Informatics* 129, 167–174.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica 59*(6), 1551–1580.

Johansen, S. (1996). *Likelihood-based Inference in Cointegrated Vector Auto-Regressive Models*. Oxford University Press.

Johansen, S. (2005). Interpretation of cointegrating coefficients in the cointegrated vector autoregressive model. *Oxford Bulletin of Economics and Statistics 67*(1), 93–104.

Johansen, S. (2009). Representation of cointegrated autoregressive process with application to fractional process. *Econometric Reviews* 28, 121–145.

Johansen, S., R. Mosconi, and B. Nielsen (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal* 3, 216–249.

Jombart, T., E. Nightingale, M. Jit, O. de Waroux, G. Knight, S. Flasche, R. Eggo, A. J. Kucharski, C. Pearson, and S. Procter (2020). Forecasting critical care bed requirements for COVID-19 patients in england. https://cmmid.github.io/topics/covid19/ICU–projections.html.

Jones, S. S., R. S. Evans, T. L. Allen, A. Thomas, P. J. Haug, S. J. Welch, and G. L. Snow (2009). A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of Biomedical Informatics 42*(1), 123–139.

Kadri, F., F. Harrou, S. Chaabane, and C. Tahon (2014). Time series modelling and forecasting of emergency department overcrowding. *Journal of Medical Systems 38*(9), 107.

Kim, S. and H. Kim (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting 32*(3), 669–679.

Kreke, J. E., A. J. Schaefer, and M. S. Roberts (2004). Simulation and critical care modeling. *Current Opinion in Critical Care 10*(5), 395–398.

Kurita, T., H. B. Nielsen, and A. Rahbek (2011). An I(2) cointegration model with piecewise linear trends. *The Econometrics Journal 14*(2), 131–155.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting 36*(1), 54–74.

Milne, E. and P. Whitty (1995). Calculation of the need for paediatric intensive care beds. *Archives of Disease in Childhood 73*(6), 505–507.

Pearson, G. A., F. Reynolds, and J. Stickley (2012). Calculating the need for intensive care beds. *Archives of Disease in Childhood 97*(11), 943–946.

Peracchi, F. (2020). The Covid-19 pandemic in Italy. Working paper, EIEF.

Petropoulos, F. and S. Makridakis (2020). Forecasting the novel coronavirus covid-19. *PloS One 15*(3).

Remuzzi, A. and G. Remuzzi (2020). COVID-19 and Italy: what next? *The Lancet*, https://doi.org/10.1016/S0140–6736(20)30627–9.

Richardson, S. and D. Spiegelhalter (2020). Coronavirus statistics: what can we trust and what should we ignore? *The Guardian* (April 12).

Ridge, J., S. Jones, M. Nielsen, and A. Shahani (1998). Capacity planning for intensive care units. *European Journal of Operational Research 105*(2), 346–355.

Rodriguez-Llanes, J., R. Delgado, M. Pedersen, M. Meneghini, and P. González (2020). COVID-19: Surging critical care capacity across Europe. Technical report, https://www.researchgate.net/publication/340528505.

Sadki, A., X. Xie, and F. Chauvin (2013). Planning oncologists of ambulatory care units. *Decision Support Systems 55*(2), 640–649.

Shoukat, A., C. R. Wells, J. M. Langley, B. H. Singer, A. P. Galvani, and S. M. Moghadas (2020). Projecting demand for critical care beds during covid-19 outbreaks in canada. *CMAJ 192*(19), E489–E496.

StatGroup-19 (2020). Technical note on short-term predictions of daily intensive care unit (icu) hospitalizations due to covid-19. Technical report, SIS, https://statgroup-19.blogspot.com/p/short-term-predictions-of-daily.html.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica 55*(5), 1035–1056.

USCOVID (2020). COVID-19 forecasts for the united states. Technical report, https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html.

Zhang, R., Y. Li, A. L. Zhang, Y. Wang, and M. J. Molina (2020). Identifying airborne transmission as the dominant route for the spread of covid-19. *Proceedings of the National Academy of Sciences*.

Zhu, T., L. Luo, X. Zhang, Y. Shi, and W. Shen (2015). Time-series approaches for forecasting the number of hospital daily discharged inpatients. *IEEE Journal of Biomedical and Health Informatics 21*(2), 515–526.

Zhu, Z., B. H. Hen, and K. L. Teow (2012). Estimating icu bed capacity using discrete event simulation. *International Journal of Health Care Quality Assurance 25*(2), 134–144.

Zivot, E. and D. W. K. Andrews (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics 10*(3), 251–270.

Appendix A. Stationary companion form

W_t satisfies the stationary companion form $W_t = BW_{t-1} + J_{q,p}(\mu D_t + \varepsilon_t)$ with $B = G'A\bar{G}$, namely

$$\begin{pmatrix} \Delta x_t \\ \beta' x_{t-1} \\ \Delta x_{t-1} \\ \Delta x_{t-2} \\ \vdots \\ \Delta x_{t-k+2} \end{pmatrix} = \begin{pmatrix} \Gamma_1 + \alpha\beta' & \alpha & \Gamma_2 & \dots & \Gamma_{k-2} & \Gamma_{k-1} \\ \beta' & I_r & & & & \\ I_p & & & & & \\ & & I_p & & & \\ & & & \ddots & & \\ & & & & I_p & \end{pmatrix} \begin{pmatrix} \Delta x_{t-1} \\ \beta' x_{t-2} \\ \Delta x_{t-2} \\ \Delta x_{t-3} \\ \vdots \\ \Delta x_{t-k+1} \end{pmatrix} + J_{q,p}(\mu D_t + \varepsilon_t). \tag{A.1}$$

The fact that $B = G'A\bar{G}$ can be shown as follows. Let $N = (G, F)'$ where $F = (0, \beta_\perp, 0)$ conformably with G. Observe that N is invertible and that $NX_t = (W_t', f_t')' = R_t$ where $f_t = \beta_\perp' x_{t-1}$. Premultiplying $X_t = AX_{t-1} + J_{s,p}(\mu D_t + \varepsilon_t)$ by N and observe that $NX_t = NAN^{-1}NX_{t-1} + NJ_{s,p}(\mu D_t + \varepsilon_t)$ i.e. $R_t = NAN^{-1}R_{t-1} + J_{s,p}(\mu D_t + \varepsilon_t)$ because $NJ_{s,p} = J_{s,p}$ as can be verified directly. Because N has orthogonal blocks one also sees that $N^{-1} = (\bar{G}, \bar{F})$, and hence

$$NAN^{-1} = \begin{pmatrix} G'A\bar{G} & G'A\bar{F} \\ F'A\bar{G} & F'A\bar{F} \end{pmatrix} = \begin{pmatrix} G'A\bar{G} & 0 \\ (\beta_\perp', 0) & I_{p-r} \end{pmatrix}$$

where $G'A\bar{F} = 0$ follows because $\bar{F}$ acts on the second block of columns, and all entries in $G'A$ in the second block end with β'. Similarly one finds $F'A\bar{G} = (\beta_\perp', 0)$ and $F'A\bar{F} = I_{p-r}$. The equation of $f_t = \beta_\perp' x_{t-1}$ reads $f_t = \beta_\perp' J_{q,p}' W_{t-1} + f_{t-1} = \beta_\perp' \Delta x_{t-1} + f_{t-1}$, i.e. it is an identity (which loads on the stationary part W_{t-1}). This shows that $B = G'A\bar{G}$ follows because $\beta_\perp' x_{t-2}$ is never loaded in Δx_t. This shows that the dynamics of W_t is autonomous with respect to the one of f_t, and that f_t depends on the dynamics of W_t.

This also shows that the eigenvalues of A are equal to the eigenvalues of B and of I_{p-r}, where the latter are $p - r$ eigenvalues all equal to 1. In the present context with $p = 2$ and $r = 1$, one has $p - r = 1$.