National Accounting Review

NAR, 4(1): 38–55.
DOI: 10.3934/NAR.2022003
Received: 16 November 2021
Revised: 16 January 2022
Accepted: 14 February 2022
Published: 22 February 2022

http://www.aimspress.com/journal/NAR

Research article

Modelling the link between Covid-19 cases, hospital admissions and deaths in England

Terence C. Mills*

School of Business and Economics, Loughborough University, Ashby Road, Loughborough LE11 3TU, UK

*** Correspondence:** Email: t.c.mills@lboro.ac.uk.

Abstract: Analysing the mass of time series data accumulating daily and weekly from the coronavirus pandemic has become ever more important as the pandemic has progressed through its numerous phases. Econometric techniques are particularly suited to analysing this data and research using these techniques is now appearing. Much of this research has focused on short-term forecasting of infections, hospital admissions and deaths, and on generalising to stochastic settings compartmental epidemiological models, such as the well-known "susceptible (S), infected (I) and recovered or deceased (R)", or SIR, model. The focus of the present paper is rather different, however, in that it investigates the changing dynamic relationship between infections, hospital admissions and deaths using daily data from England. It does this using two approaches, balanced growth models and autoregressive distributed lag/error correction models. It is found that there has been a substantial decrease over time in the number of deaths and hospital admissions associated with an increase in infections, with patients being kept alive longer, as clinical practice has improved and the vaccination program rolled out. These responses may be tracked and monitored through time to ascertain whether such improvements have been maintained.

Keywords: Covid-19; infections, admissions and deaths; England; time series econometrics; balanced growth models; autoregressive distributed lag models; error correction

JEL Codes: C22, I10, I12

1. Introduction

Since the onset of the Covid-19 pandemic in early 2020 an enormous research effort has been underway on the modelling and prediction of various aspects of the pandemic. For accessible reviews concentrating on general features of this modelling, see, for example, Vespignani et al. (2020), Poletto et al. (2020) and Gnanvi et al. (2021), while for discussion of the growth models widely used for predicting Covid-19 infections and deaths, see Tovissodé et al. (2020) and Shen (2020). Central to this modelling, the analysis of the mass of time series data accumulating daily and weekly from the coronavirus pandemic has become ever more important as the pandemic has progressed through its numerous phases. Spiegelhalter and Masters (2021) provide an accessible introduction to such data issues, paying particular attention to the evidence emerging from the U.K.

It is becoming increasingly apparent that econometric techniques are particularly suited to analysing this data: see, for example, Li and Linton (2020), Manski and Molinari (2020) and the review by Dolton (2021). Much of the research using these techniques has focused on short-term forecasting of cases, hospital admissions and deaths, with notable examples being Doornik et al. (2020), Doornik et al. (2021) and Harvey et al. (2021). It has also been directed at generalising, to stochastic settings, compartmental epidemiological models, such as the well-known "susceptible (S), infected (I) and recovered or deceased (R)", or SIR, model, as in Korolev (2020) and Pesaran and Yang (2021).

The focus of the present paper is rather different, however, in that we investigate the changing dynamic relationship between infections, hospital admissions and deaths using daily data from England. Section 2 thus considers the relationship between hospital admissions and subsequent deaths and introduces two models that might be useful for this task: the recently proposed balanced growth model of Harvey (2020) and the more familiar autoregressive distributed lag/error correction model used widely to analyse economic time series (see, for example, Banerjee et al., 1993, for detailed development and Mills, 2019, chapters 12 and 14, for a more introductory treatment). Section 3 extends the analysis to examining the prior relationship between infections and hospital admissions, while Section 4 links the two sets of models together before discussing the advantages and disadvantages of the two modelling procedures.

2. Modelling the relationship between hospital admissions and deaths in England

Figure 1 shows daily hospital admissions and deaths in England between 19th March 2020 and 31st October 2021.[1] Both admissions and deaths show pronounced multiple wave patterns with admissions obviously leading deaths, but the shifting nature of the relationship between the two series is clearly discernible. How might this evolving and dynamic relationship be modelled? Attention is focused in this paper on two approaches: balanced growth modelling and the use of autoregressive distributed lags.

2.1. Balanced growth modelling

Let daily deaths due to Covid-19 in England be denoted y_t, $t = 1,2,\cdots,T$, with their cumulation being $Y_t = \sum_{j=1}^{t} y_j$, so that the growth rate of daily deaths is $g_{y,t} = y_t/Y_{t-1}$. Similarly, denote daily

[1]The focus here is on data from England as U.K.—wide hospital admissions rely on different definitions across the home nations.

hospital admissions due to Covid-19 by x_t, their cumulation by $X_t = \sum_{j=1}^{t} x_j$, and their growth rate by $g_{x,t} = x_t / X_{t-1}$.

Following Harvey and Kattuman (2020), we initially assume that there is balanced growth between daily deaths and hospital admissions lagged k days, which implies the regression model.

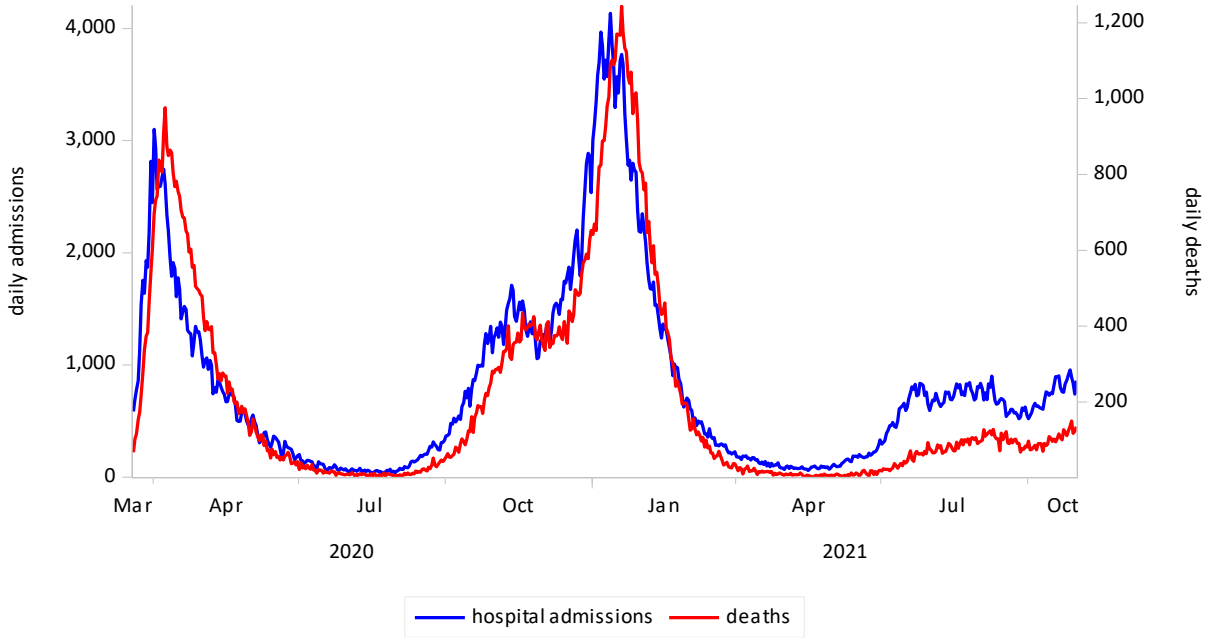

Figure 1. Daily hospital admissions and deaths for England: 19th March 2020–31st October 2021.

$$\log(g_{y,t}) = \delta + \log(g_{x,t-k}) + \varepsilon_t \qquad t = k+1, \cdots, T \tag{1}$$

where ε_t is an error term assumed to independently and identically distributed through time with zero mean and variance σ_ε^2, which is denoted $\varepsilon_t \sim$ IID$(0, \sigma_\varepsilon^2)$. The "equilibrium" relationship between the two growth rates is given by

$$g_{y,t} = \exp(\delta) g_{x,t-k} \tag{2}$$

Between daily deaths and admissions, y_t and x_t, the equilibrium is then

$$y_t = \exp(\delta)(Y_{t-1}/X_{t-k-1})x_{t-k} \tag{3}$$

Allowing for a lag structure in the leading admissions series in (1) gives

$$\log(g_{y,t}) = \delta + \sum_{j=h}^{k} \beta_j \log(g_{x,t-j}) + \varepsilon_t \tag{4}$$

where $h < k$ and $\sum \beta_j = 1$, a restriction that may be imposed by rewriting (4) as

$$\log(g_{y,t}) - \log(g_{x,t-k}) = \delta + \sum_{j=h}^{k-1} \beta_j \left(\log(g_{x,t-j}) - \log(g_{x,t-k})\right) + \varepsilon_t \tag{5}$$

so that $\beta_k = 1 - \sum_{j=h}^{k-1} \beta_j$, a restriction that ensures that there is indeed balanced growth. The corresponding equilibrium relationship is then

$$g_{y,t} = \exp(\delta) \prod_{j=h}^{k} g_{x,t-j}^{\beta_j} = \exp(\delta) \bar{g}_{x,t-k} \tag{6}$$

where $\bar{g}_{x,t-k}$ is the weighted geometric mean of $g_{x,t-h}, \cdots, g_{x,t-k}$. The levels equilibrium is thus

$$y_t = \exp(\delta)(Y_{t-1}/\bar{X}_{t-k-1})\bar{x}_{t-k} = \Delta \bar{x}_{t-k} \tag{7}$$

where $\bar{x}_{t-k}$ and $\bar{X}_{t-k-1}$ are the corresponding weighted geometric means of $x_{t-h}, \cdots, x_{t-k}$ and $X_{t-1-h}, \cdots, X_{t-1-k}$, respectively. Thus Δ measures the *long-run* response of deaths to an increase in hospital admissions: if daily admissions increase by 100 then deaths will increase by 100Δ after k days.

When the two series are not on the same growth path, the model can be extended by replacing the intercept δ with a stochastic trend:

$$\log(g_{y,t}) = \delta_t + \sum_{j=h}^{k} \beta_j \log(g_{x,t-j}) + \varepsilon_t \tag{8}$$

or

$$\log(g_{y,t}) - \log(g_{x,t-k}) = \delta_t + \sum_{j=h}^{k-1} \beta_j \left(\log(g_{x,t-j}) - \log(g_{x,t-k})\right) + \varepsilon_t \tag{9}$$

where δ_t is defined as

$$\delta_t = \delta_{t-1} - \gamma_{t-1} + \eta_t \qquad \eta_t \sim \mathrm{IID}(0, \sigma_\eta^2) \tag{10}$$

$$\gamma_t = \gamma_{t-1} + \zeta_t \qquad \zeta_t \sim \mathrm{IID}(0, \sigma_\zeta^2) \tag{11}$$

i.e., δ_t is a random walk with a drift that is itself potentially a random walk. If $\sigma_\zeta^2 = 0$ then $\gamma_t = \gamma_{t-1}$ and the drift is constant. On the other hand, if $\sigma_\eta^2 = 0$ then $\delta_t = 2\delta_{t-1} - \delta_{t-2} - \zeta_{t-1}$ and δ_t will tend to evolve very smoothly, being known as an integrated random walk, or IRW. The equilibrium relationship in (6) is

$$g_{y,t} = \exp(\delta_t) \bar{g}_{x,t-k} \tag{12}$$

so that the dynamic relationship between the two growth rates is given by $\exp(\delta_t)$. In terms of daily deaths and admissions, y_t and x_t, we have

$$y_t = \exp(\delta_t)(Y_{t-1}/\bar{X}_{t-k-1})\bar{x}_{t-k} = \Delta_t \bar{x}_{t-k} \tag{13}$$

Thus, an increase of 100 in hospital admissions will lead to an increase of $100\Delta_t$ deaths in the following k days and this long run response will shift through time. As Harvey (2020) shows, this model may be arrived at by assuming that deaths and admissions follow Gompertz processes separated by k days, but such an assumption is not necessary.

Equation (9) may be fitted by casting it into state space form and employing the Kalman filter. Estimation is carried out by maximum likelihood using the predictive error decomposition and the

estimates of δ_t used to compute the Δ_t series shown in Figures 3 and 7 below are the "smoothed" estimates, obtained by running the Kalman filter first forwards from $t = k + 1$ to $t = T$ and then backwards from $t = T$ to $t = k + 1$. Mills (2019, chapter 17) provides an introductory discussion to state space modelling and Harvey (1989) is the classic exposition. A value for the lag k must be obtained before estimation can be undertaken. This lag could be selected by considering the basic model (9) with $h = 0$, setting k to an initial value k_{max} and then running the regression (9) for $k = k_{max}, k_{max} - 1, k_{max} - 2, \cdots$, the sequence stopping when $\hat{\beta}_k$ is significant at some pre-chosen level of significance. Alternatively, clinical considerations may suggest an appropriate value for k and indeed h, and this simpler approach is also investigated. A further refinement may be to select only a subset of the lagged regressors $\log(g_{x,t}), \cdots, \log(g_{x,t-k})$, thus determining h and including only significant lags. This may be done, for example, by using a stepwise least squares algorithm or other sequential testing procedure. The balanced growth assumption may be checked by including $\log(g_{x,t-k})$ (or, indeed, any other lag) in (9) and testing for its significance.

2.2. ARDL modelling

An autoregressive distributed lag (ARDL) model linking deaths and admissions may be specified in general as

$$y_t = \phi_0 + \sum_{i=1}^{m} \phi_i y_{t-i} + \sum_{i=0}^{n} \theta_i x_{t-i} + u_t \tag{14}$$

where $u_t \sim \text{IID}(0, \sigma_u^2)$ is an error term. Typically, the lag lengths m and n will be unknown and must be determined from the data. An algebraically equivalent but often more convenient form of this ARDL(m, n) model, particularly for model specification and inference, is the *error correction model* (ECM)[2]

$$\nabla y_t = a_0 + \sum_{i=1}^{m-1} a_i \nabla y_{t-i} + \sum_{i=0}^{n-1} b_i \nabla x_{t-i} - c(y_{t-1} - d x_{t-1}) + u_t \tag{15}$$

where ∇ is the difference operator defined such that $\nabla y_t = y_t - y_{t-1}$ and where the coefficients of (14) and (15) are linked by the set of relationships

$$\begin{aligned}
\phi_0 &= a_0 \\
\phi_1 &= 1 + a_1 - c \\
\phi_i &= a - a_{i-1} \qquad i = 2, \cdots, m-1 \\
\phi_m &= -a_{m-1} \\
\theta_0 &= b_0 \\
\theta_1 &= b_1 - b_0 + cd \\
\theta_i &= b_i - b_{i-1} \qquad i = 2, \cdots, n-1 \\
\theta_n &= -b_{n-1}
\end{aligned} \tag{16}$$

[2]Banerjee et al (1993) sets out and analyses in detail the algebraic equivalencies existing between the ARDL and ECM formulations. It should be emphasised that the recasting of (14) as (15) is a purely algebraic transformation and is not predicated on any particular statistical properties of the data. With integrated data, cointegration leads from an ARDL to an ECM via Granger's representation theorem (see Engle and Granger, 1987). The series here are *not* integrated, as may be demonstrated from standard unit root tests, and so we are in a stationary world in which the ECM (15) is simply a more convenient representation for our purposes than the ARDL (14).

The *error correction term* is $y_{t-1} - dx_{t-1}$, which embodies the long-run, or equilibrium, relationship $y = dx$ between deaths and admissions, with d being termed the *long-run* or *total multiplier*. An increase in hospital admissions of 100, say, will eventually increase deaths by $100d$.

The speed at which the increase in deaths approaches the total multiplier depends on c, the *speed of adjustment* parameter. The smaller is c, the faster the speed of adjustment and the quicker the total multiplier is arrived at. The actual time path of adjustment depends upon the lag coefficients ψ_i in the distributed lag

$$y_t = \psi_0 + \sum_{i=0}^{\infty} \psi_i x_{t-i} \tag{17}$$

where

$$\psi_i = \sum_{j=1}^{\min(i,m)} \phi_j \psi_{i-j} + \theta_i \qquad 0 \le i \le n \tag{18}$$

$$\psi_i = \sum_{j=1}^{\min(i,m)} \phi_j \psi_{i-j} \qquad i > n$$

This result is most easily obtained by utilising the lag operator B, defined such that $B^j z_t \equiv z_{t-j}$ (note that the difference operator introduced in (15) may then be written as $\nabla = 1 - B$). This allows (14), on ignoring the error term as we are only interested in the systematic dynamics here, to be written as

$$\phi(B) y_t = \phi_0 + \theta(B) x_t \tag{19}$$

where the *lag polynomials* $\phi(B)$ and $\theta(B)$ are defined as

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_m B^m \tag{20}$$

$$\theta(B) = \theta_0 + \theta_1 B + \cdots + \theta_n B^n \tag{21}$$

Equation (19) can then be expressed as

$$y_t = \phi^{-1}(B)\phi_0 + \phi^{-1}(B)\theta(B) x_t = \psi_0 + \psi(B) x_t \tag{22}$$

where

$$\psi_0 = \phi_0 / (1 - \phi_1 - \cdots - \phi_m) \tag{23}$$

$$\psi(B) = \phi^{-1}(B)\theta(B)$$

Thus, the lag coefficients in (18) are obtained by equating coefficients of powers of B in $\psi(B)\phi(B) = \theta(B)$. The total multiplier is then given by the sum of these lag coefficients, i.e., $d = \sum_{i=1}^{\infty} \psi_i$ and the increase in deaths after l days is given by the l^{th} *interim multiplier* $d_l = \sum_{i=1}^{l} \psi_i \qquad l = 1, 2, \cdots$ which will converge to d as l increases, i.e., $d_l \to d$ as $l \to \infty$.

2.3. Fitting balanced growth and ARDL/ECM models to the relationship between hospital admissions and deaths in England

The logarithms of the growth rates of daily hospital admissions and subsequent deaths between 21st March 2020 and 31st October 2021 are shown in Figure 2. The series are evidently not on the same growth path so that a model of the form (9) rather than (5) is clearly required. Both clinical

considerations and exploratory sequential testing along the lines suggested in section 2.1 above suggest setting $h = 0$, thus allowing for the death of some patients on the day of their admission to hospital, and $k = 7$, so that there is a one week delay between hospital admission and death.

Column (1) of Table 1 reports estimates of (9) with $\theta \log(g_{x,t-7})$ included as an additional term. If θ is non-zero then balanced growth does not hold: with this term included,

$$\beta_7 = 1 - \sum_{j=0}^{6} \beta_j + \theta \tag{24}$$

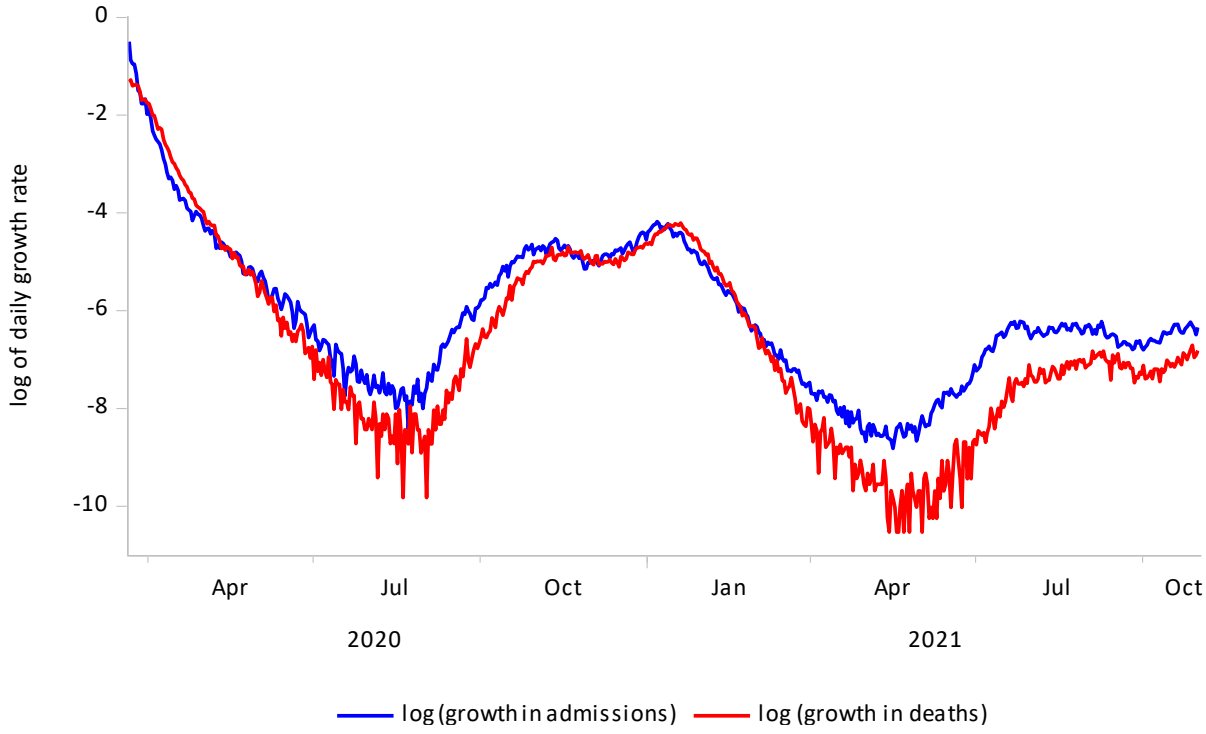

Figure 2. Logarithms of growth rates in daily deaths and admissions in England: 21st March 2020–31st October 2021.

Table 1. Estimates of (9) for daily hospital admissions and deaths in England: 21st March 2020–31st October 2021; standard errors in parentheses ( ); prob-values in brackets [ ].

	(1)	(2)
β_0	0.057 (0.069) [0.408]	0
β_1	−0.071 (0.072) [0.323]	0
β_2	0.327 (0.068) [0.000]	0.364 (0.048) [0.000]
β_3	0.039 (0.072) [0.593]	0
β_4	0.148 (0.068) [0.030]	0.168 (0.052) [0.001]
β_5	−0.063 (0.076) [0.406]	0
β_6	0.185 (0.072) [0.010]	0.201 (0.061) [0.001]
θ	−0.129 (0.083) [0.118]	0
σ_η	0.000 (3.497) [0.999]	0
σ_ζ	0.002 (0.001) [0.000]	0.002 (0.000) [0.000]
σ_ε	0.212 (0.004) [0.000]	0.213 (0.004) [0.000]

Several of the coefficients are estimated to be insignificantly different from zero, including θ, so that balanced growth is confirmed. Column (2) of Table 1 reports the estimates of the model with $\beta_0, \beta_1, \beta_3, \beta_5$ and θ all restricted to be zero on using a 10% significance level; the remaining coefficients are all significantly positive (β_7 is calculated to be 0.267 (0.057)). It is also found that $\sigma_\eta^2 = 0$, so that δ_t will be a smoothly evolving IRW. Figure 3 shows the resulting estimate of Δ_t, the response of deaths to an increase in hospital admissions, calculated using the geometric mean

$$\bar{X}_{t-8} = X_{t-3}^{0.364} X_{t-5}^{0.168} X_{t-7}^{0.201} X_{t-8}^{0.267} \tag{25}$$

and it is indeed seen to evolve smoothly. 95% confidence interval upper and lower bounds for Δ_t are also shown, and these indicate that Δ_t is indeed estimated precisely.

Δ_t reaches a maximum at 0.38 on 27th April 2020, just under three weeks after the peak in daily deaths. There are further local maxima of 0.28 on 5th December 2020 and 0.31 on 26th December 2020, both occurring during the second wave of deaths in late 2020/early 2021. Δ_t reaches a minimum of 0.06 on 9th June 2021, the previous minimum being 0.13 on 10th August 2020, and at the end of October 2021 it stood at 0.15. This suggests that, while Δ_t is positively related to the number of daily deaths, its magnitude has been declining relative to the number of deaths as clinical practice has improved over the course of the pandemic and the vaccination program has been rolled out. The long run response of deaths to an increase of 100 in hospital admissions has fallen from a maximum of 37 deaths (95% confidence interval $(34, 40)$) during the first wave of the pandemic in April 2020, to around 30 $(27, 33)$ during the second wave at the turn of the year and down to just 6 $(5, 7)$ by the summer of 2021, although this had increased to 15 $(13, 17)$ by the end of October.

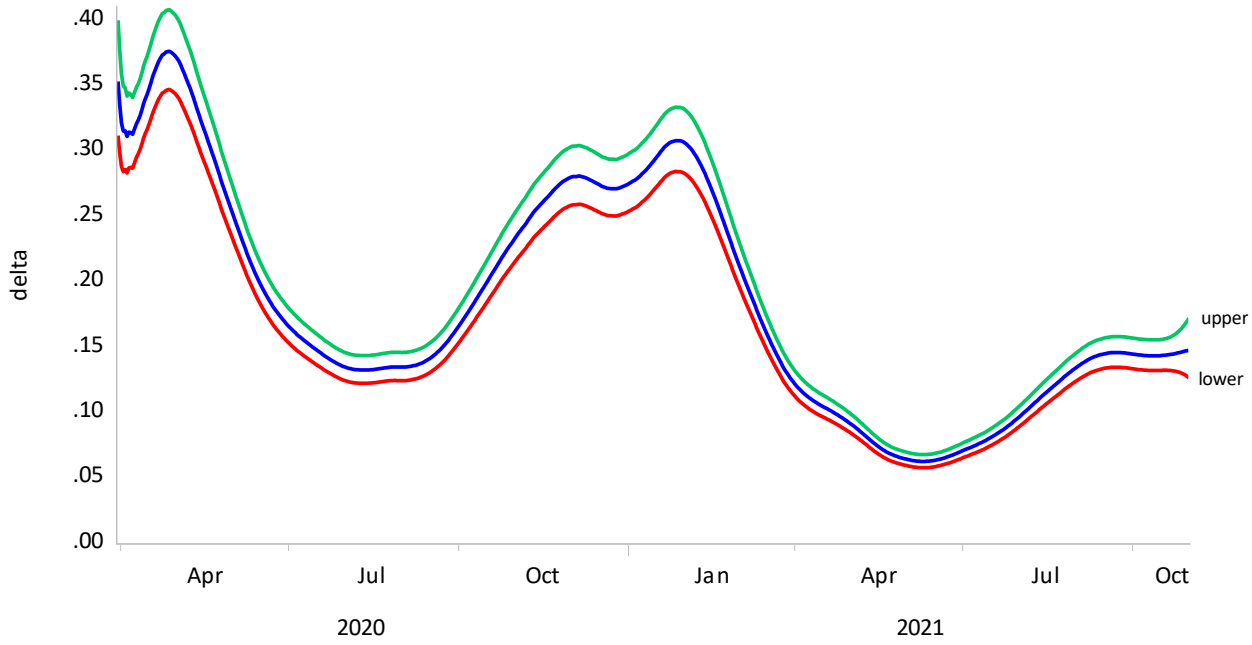

Figure 3. Estimated response Δ_t for deaths in England, 30th March 2020–31st October 2021, with 95% confidence interval upper and lower bounds.

ARDL/ECM models were developed to characterise the relationship between daily deaths and admissions for the complete sample period to 31st October 2021 and for two sub-periods, the first ending on 31st January 2021, the second beginning on 1st February 2021, the "break point" being chosen to reflect the increasing vaccination uptake at this point. The models, estimated by nonlinear least squares, were initially selected using the AIC information criteria with insignificant coefficients (at the 10% level) then being sequentially removed to obtain a parsimonious ECM specification. Details of the chosen specifications are given in Table 2. HAC standard errors are shown in parentheses to accommodate any remaining autocorrelation and heteroskedasticity in the residuals. The most notable features of the models are the much lower estimate of the equilibrium parameter d in the second sub-period than in the first, which is accompanied by a much larger estimate of c. Note that the estimates of these parameters are all highly significant and, in terms of goodness of fit, the model for the second sub-period has much the superior performance, with a higher R^2 and a lower equation standard error than the model for the first sub-period.

The time paths of the interim multipliers d_l for the two sub-periods are shown in Figure 4. The paths have been smoothed to ensure that they are monotonically non-declining, since the number of deaths that follow a given increase in hospital admissions cannot fall with time! With $\hat{d} = 0.328\ (0.035)$ in the first sub-period ending on 31st January 2021, an increase of 100 in hospital admissions will therefore eventually lead to a further 33 deaths (95% confidence interval $(26, 40)$). From Figure 4 it is seen that approximately 95% of these deaths (31) occur within 14 days of admission to hospital.

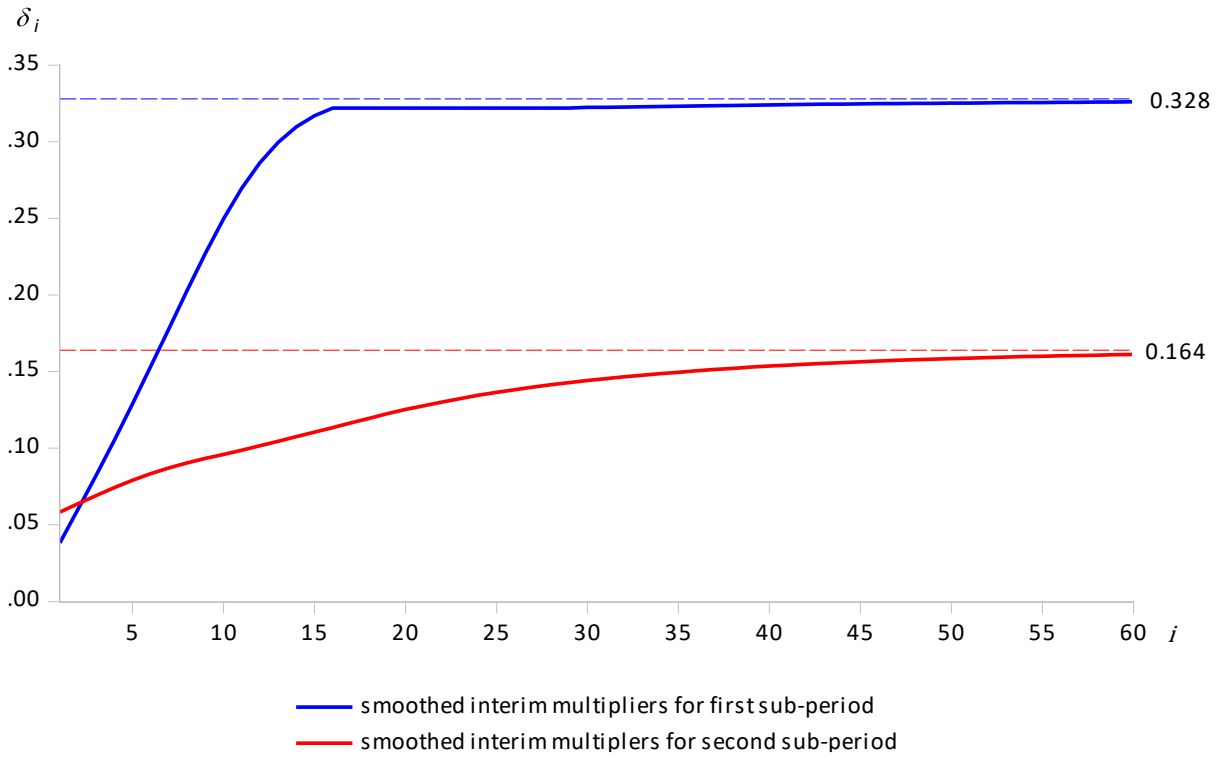

Figure 4. Time paths of interim multipliers for deaths calculated from ECMs.

Table 2. ECM estimates of the relationship between deaths and hospital admissions: standard errors in parentheses ( ); prob-values in brackets [ ].

	31st March 2020–31st October 2021	31st March 2020–31st January 2021	1st February 2021–31st October 2021
c	-0.052 (0.014) [0.000]	-0.056 (0.020) [0.005]	-0.181 (0.020) [0.000]
d	0.308 (0.025) [0.000]	0.328 (0.035) [0.000]	0.164 (0.010) [0.000]
a_0	-3.080 (1.052) [0.004]	-2.925 (1.369) [0.033]	-3.679 (0.846) [0.000]
a_1	-0.438 (0.056) [0.000]	-0.374 (0.055) [0.000]	-0.549 (0.078) [0.000]
a_2	-0.344 (0.055) [0.000]	-0.400 (0.068) [0.000]	-0.346 (0.068) [0.000]
a_3	–	–	-0.263 (0.093) [0.005]
a_4	–	–	-0.339 (0.092) [0.000]
a_5	0.119 (0.054) [0.028]	–	–
a_6	0.108 (0.064) [0.095]	–	–
a_7	0.091 (0.053) [0.089]	–	0.107 (0.058) [0.067]
a_8	–	–	0.141 (0.062) [0.024]
a_9	–	–	0.193 (0.079) [0.015]
a_{10}	–	–	–
a_{11}	–	–	-0.110 (0.061)[0.075]
b_0	0.055 (0.012) [0.000]	0.056 (0.013) [0.000]	0.043 (0.018) [0.015]
b_1	0.041 (0.013) [0.002]	0.034 (0.016) [0.029]	0.036 (0.019) [0.068]
b_2	–	–	–
b_3	–	–	–
b_4	–	–	0.036 (0.017) [0.041]
b_5	0.027 (0.015) [0.069]	0.041 (0.015) [0.008]	–
b_6	0.030 (0.014) [0.030]	0.041 (0.010) [0.000]	–
b_7	0.027 (0.015) [0.067]	0.042 (0.013) [0.002]	–
b_8	0.031 (0.016) [0.058]	0.037 (0.015) [0.011]	–
b_9	0.060 (0.015) [0.000]	0.076 (0.016) [0.000]	-0.047 (0.022) [0.034]
b_{10}	0.058 (0.012) [0.000]	0.062 (0.013) [0.000]	–
b_{11}	0.050 (0.013) [0.000]	0.065 (0.015) [0.000]	–
R^2	0.469	0.512	0.597
σ_u	17.91	20.80	11.13
T	580	307	273
m	8	3	12
n	12	12	10

In contrast, for the second sub-period beginning on 1st February 2021, $\hat{d} = 0.164\ (0.010)$, so that now an increase of 100 in hospital admissions eventually leads to just 16 further deaths (95% confidence interval $(14, 18)$). Moreover, only 69% of these deaths (11) occur within 14 days of admission to hospital. It would thus appear that, as the vaccination programme was rolled out along with other improvements in clinical practice, so the hospital death rate was more than halved with patients being kept alive for longer.

3. Modelling the relationship between Covid-19 cases and hospital admissions

As well as the linkage from hospital admissions to deaths, there is also the prior link from testing positive (i.e., becoming infected) for Covid-19 to admission into hospital. This has become particularly important to analyse since the roll-out of the vaccination program in the U.K: is there evidence that vaccination has "broken the link between cases and admissions", as has been stated several times by the government? To investigate whether this is indeed the case, we analyse English data on daily positive cases and hospital admissions from 1st September 2020 to 31st October 2021, as shown in Figure 5. Earlier data have been excluded both because of the limited extent of testing during the early months of the pandemic and the hiatus in cases and admissions during the summer months of 2020.

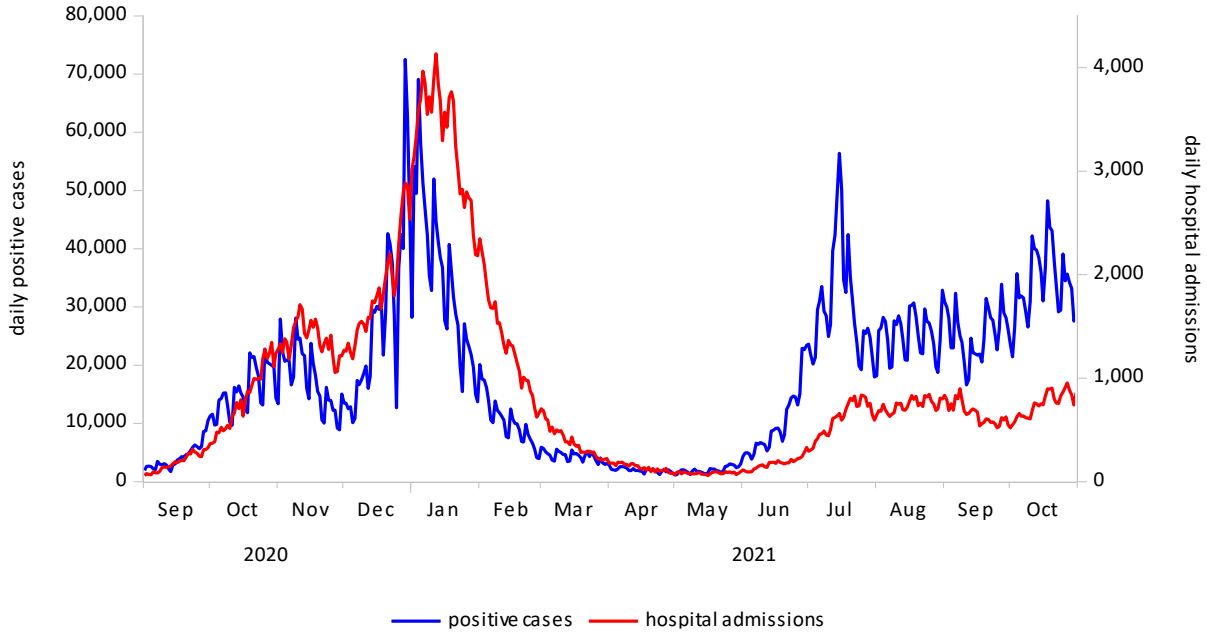

Figure 5. Positive cases and hospital admissions for England: 1st September 2020–31st October 2021.

We first fit balanced growth models to the logarithms of hospital growth, $g_{x,t}$, and case growth, $g_{w,t} = w_t/W_{t-1}$, where w_t and W_t are daily and cumulative cases, respectively. These logarithms are shown in Figure 6. From both clinical considerations and exploratory sequential testing a lag of $k = 14$ was chosen, along with $h = 1$.

Estimates of the balanced growth Equation (9) are shown in Table 3 after the elimination of insignificant coefficients. Here $\sigma_\eta > 0$ so that, with both σ_ζ and γ_t estimated to be very small, $\nabla\delta_t \approx \eta_t$, a driftless random walk. The consequent Δ_t series, estimated under the assumption of balanced growth (which is questionable here since θ appears to be significantly negative: $\hat{\theta} = -0.321\ (0.064)$), is shown in Figure 7, along with 95% confidence interval upper and lower bounds. These show that Δ_t is estimated extremely precisely. The trend in Δ_t values is generally upwards until the middle of February 2021, reaching a maximum of 0.12 (95% confidence interval $(11, 13)$) on 12th December 2020 (i.e., an additional 120 hospital admissions result from an increase of 1000 positive test cases), after which it turns down, reaching a minimum of 0.02 $(0.019, 0.021)$ on 21st July 2021 (20 additional admissions from an increase of 1000 cases), and was still around this value at the end of October. This

suggests that there is indeed evidence that the rollout of the vaccination program has had an impact on the relationship between cases and subsequent hospital admissions.

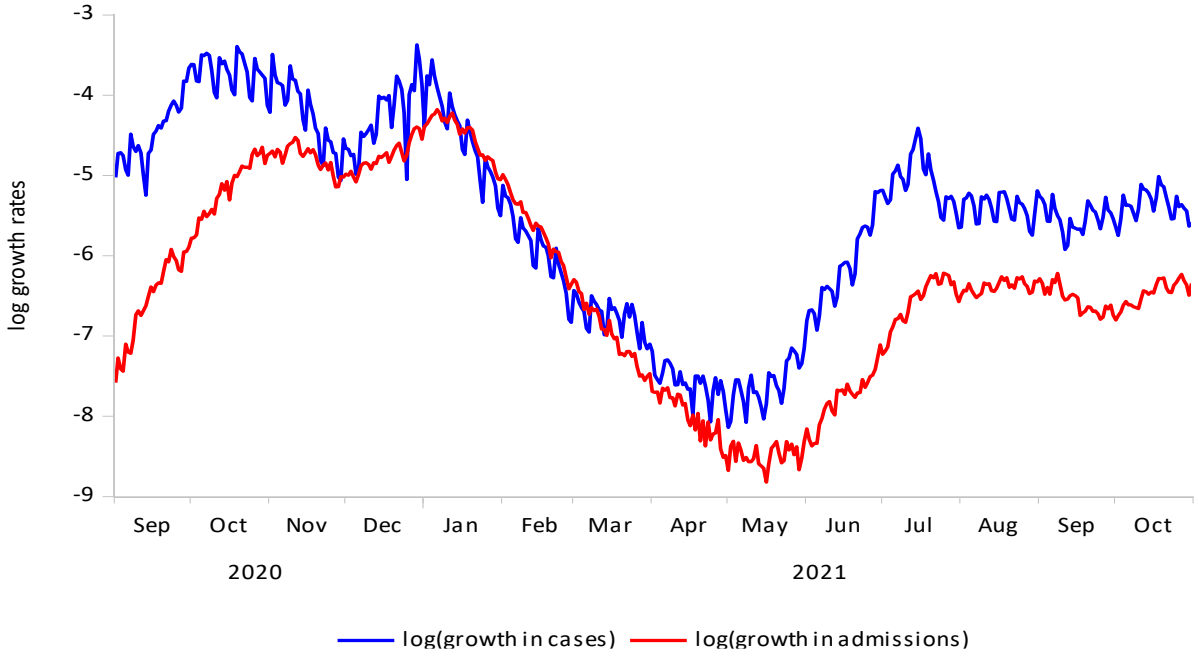

Figure 6. Logarithms of growth rates in daily cases and admissions in England: 1st September 2020–31st October 2021.

Table 3. Estimates of (9) for daily positive tests and hospital admissions in England: 1st September 2020–31st October 2021; standard errors in parentheses ( ); prob-values in brackets [ ].

	value
β_1	0.086 (0.024) [0.000]
β_2	0
β_3	0.108 (0.018) [0.000]
β_4	0
β_5	0.118 (0.025) [0.000]
β_6	0.150 (0.029) [0.000]
β_7	0.094 (0.029) [0.001]
β_8	0.053 (0.023) [0.020]
β_9	0.096 (0.020) [0.000]
β_{10}	0
β_{11}	0
β_{12}	0.045 (0.025) [0.073]
β_{13}	0.067 (0.031) [0.031]
β_{14}	0.183 (0.028) [0.000]
σ_η	0.046 (0.004) [0.000]
σ_ζ	0.001 (0.0006) [0.061]
σ_ε	0.054 (0.003) [0.000]
θ	−0.321 (0.064) [0.000]

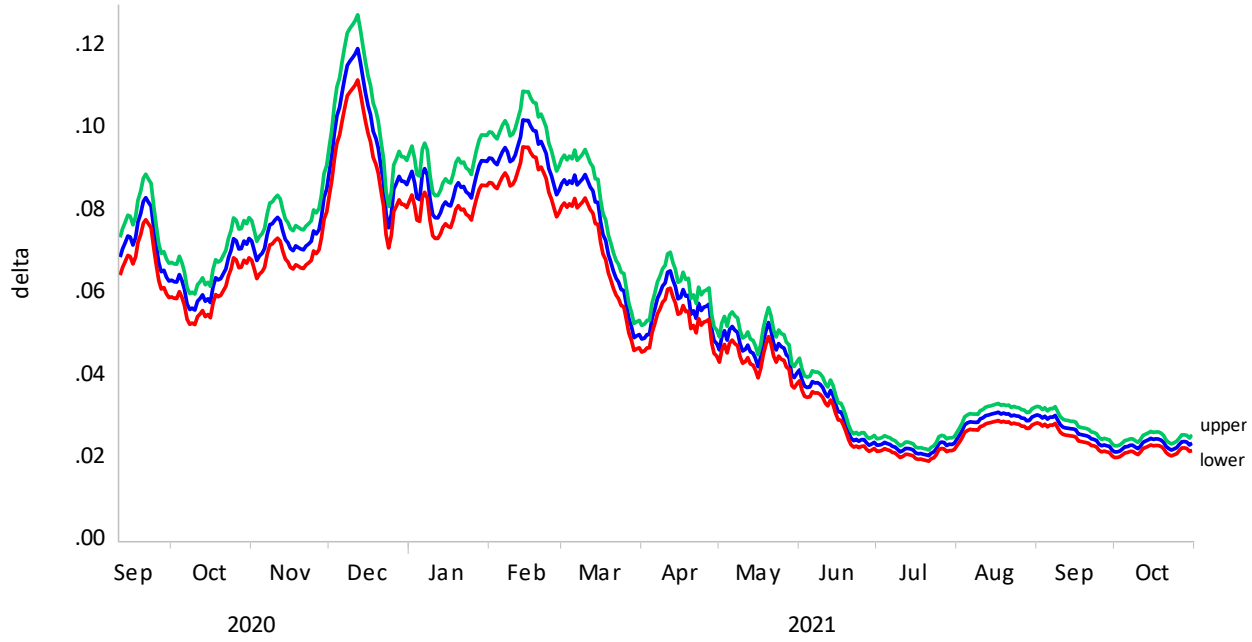

Figure 7. Estimated response Δ_t for hospital admissions in England, 11th September 2020 –31st October 2021, with 95% confidence interval lower bounds.

ARDL/ECM models were fitted to the sub-periods 1st September 2020–31st January 2021 and 1st February–31st October 2021, with the resulting estimates and statistics shown in Table 4. The estimate of the long-run parameter d is 0.087 (0.006) in the early period but just 0.027 (0.001) in the later period, i.e., up to the end of January 2021 an increase of 1000 positive test cases ultimately led to 87 (95% confidence interval $(75, 99)$) additional hospital admissions, whereas from February such an increase led to only an additional 27 (95% confidence interval $(25, 29)$) admissions. The estimate of the speed of adjustment parameter, c, on the other hand, has remained almost the same across the two periods, just increasing from 0.103 to 0.118. Again, the fit of the second sub-period model is vastly superior to that of the first.

The interim multipliers for admissions are plotted in Figure 8. In the period up to the end of January 2021, approximately 85% of hospital admissions occur within 28 days of a positive test. In contrast, for the second sub-period beginning on 1st February 2021, only 79% of these admissions occur within 28 days of a positive test. It would thus appear to be clear that, as the vaccination programme was rolled out, along with other improvements in medical practice, so fewer people were admitted to hospital.

Table 4. ECM estimates of the relationship between hospital admissions and cases; standard errors in parentheses ( ); prob-values in brackets [ ].

	9th September 2020–31st October 2021	9th September 2020–31st January 2021	1st February 2021–31st October 2021
c	−0.010 (0.005) [0.071]	−0.103 (0.020) [0.000]	−0.118 (0.016) [0.000]
d	0.038 (0.014) [0.006]	0.087 (0.006) [0.000]	0.027 (0.001) [0.000]
a_0	−	−	−
a_1	−0.241 (0.035) [0.000]	−0.198 (0.095) [0.038]	−0.488 (0.049) [0.000]
a_2	−	−	−0.236 (0.064) [0.000]
a_3	−0.178 (0.066) [0.007]	–0.287 (0.088) [0.001]	−0.351 (0.086) [0.000]
a_4	–	−0.121 (0.068) [0.075]	−0.117 (0.059) [0.050]
a_5	−	–	–
a_6	0.131 (0.031) [0.000]	–	−
a_7	0.304 (0.046) [0.000]	0.294 (0.121) [0.017]	0.213 (0.061) [0.001]
a_8	0.265 (0.024) [0.000]	0.254 (0.088) [0.005]	0.110 (0.057) [0.055]
a_9	−	−0.230 (0.081) [0.005]	−
a_{10}	0.128 (0.051) [0.013]	–	–
a_{11}	0.115 (0.049) [0.020]	–	−0.129 (0.065) [0.050]
a_{12}	−0.134 (0.037) [0.000]	−0.234 (0.094) [0.014]	0.152 (0.049) [0.002]
b_0	0.003 (0.000) [0.000]	0.005 (0.001) [0.001]	0.002 (0.001) [0.001]
b_1	−	−0.005 (0.002) [0.009]	−
b_2	–	–	–
b_3	0.003 (0.001) [0.005]	–	–0.002 (0.001) [0.077]
b_4	–	–	−0.002 (0.001) [0.003]
b_5	0.004 (0.001) [0.000]	−	−0.002 (0.001) [0.020]
b_6	0.005 (0.001) [0.008]	−	–
b_7	0.004 (0.001) [0.003]	−	0.002 (0.001) [0.016]
b_8	0.005 (0.002) [0.004]	0.003 (0.001) [0.044]	–
b_9	0.002 ± 0.001	−	–
b_{10}	−	−	0.002 (0.001) [0.089]
R^2	0.445	0.442	0.628
σ_u	72.71	112.37	33.18
T	419	146	273
m	13	13	13
n	10	9	11

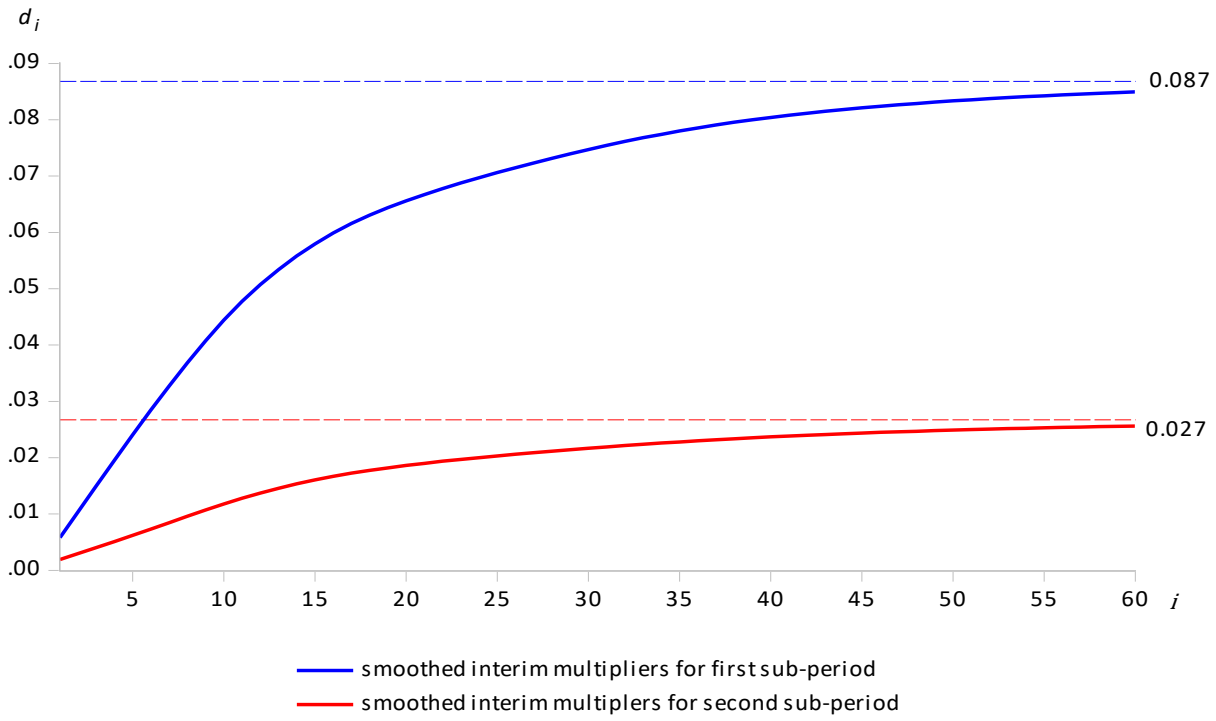

Figure 8. Time paths of interim multipliers for admissions in England calculated from ECMs.

4. Conclusions

Given the "causal structure" inherent in the relationship between cases, hospital admissions and deaths, the models in Sections 2 and 3 may be linked together. If the equilibrium responses from the balanced growth models are now denoted $\Delta_{y,t}$ and $\Delta_{x,t}$, to use an obvious extension of notation, then the response of deaths to an increase in cases is given by the product $\Delta_{y,t} \times \Delta_{x,t-7}$ to ensure an appropriate timing match-up, and this series is shown in Figure 9.[3] This product reached a maximum of 33 deaths per thousand cases on 18th December 2020 (95% confidence interval $(28, 38)$) and by the end of October 2021 had declined to just 3 $(2, 4)$.

Similarly, denoting d_x and d_y as the total multipliers from the ARDL/ECM models, with $d_{x,l}$ and $d_{y,l}$ being the accompanying interim multipliers, then the products $d_x \times d_y$ and $d_{x,l} \times d_{y,l}$ provide the total multiplier and set of interim multipliers for the response of deaths to an increase in cases, these being shown for the two sub-periods in Figure 10. The total multipliers are 28 deaths per thousand cases (95% confidence interval $(18, 38)$) for the first sub-period and 4 $(3, 5)$ for the second, both consistent with the balanced growth estimates.

Both modelling approaches provide a consistent story: as clinical practice has evolved and the vaccination program rolled out, so there has been a substantial decline in hospital admissions and subsequent deaths for a given level of infection. Are there grounds for preferring one approach over the other? Each have their advantages and disadvantages.

[3]Using (7) we have $y_t = \Delta_{y,t} x_{t-7}$ and $x_t = \Delta_{x,t} w_{t-14}$, so that $y_t = \Delta_{y,t} \Delta_{x,t-7} w_{t-21}$.

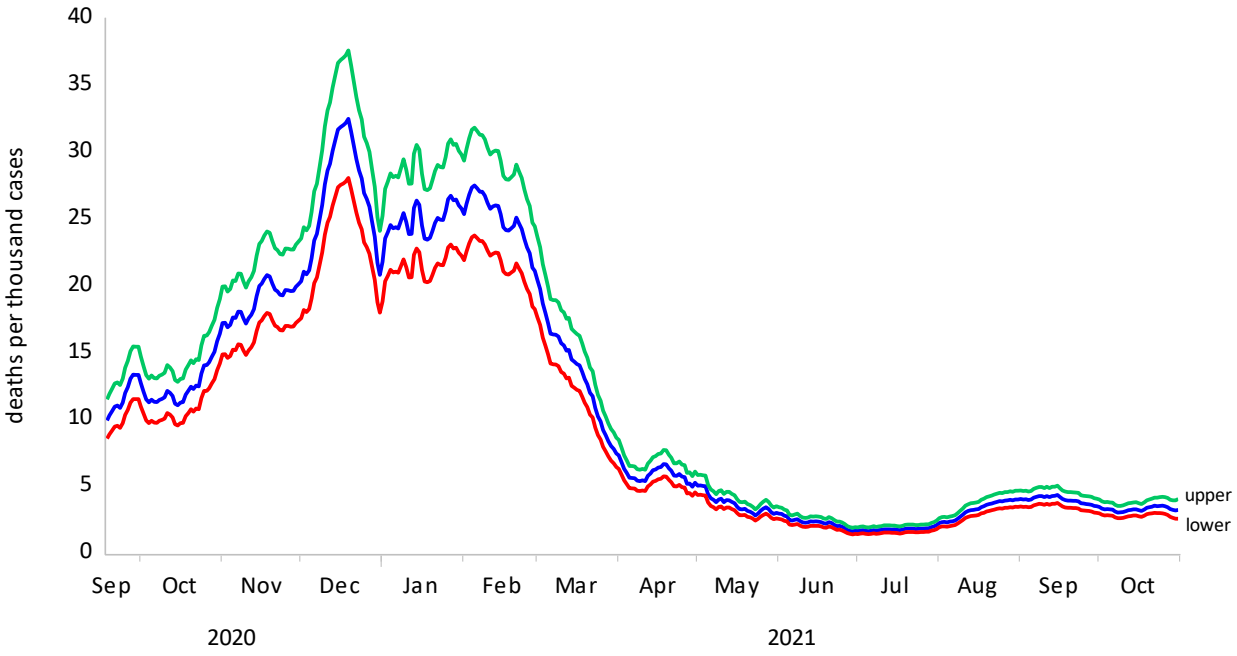

Figure 9. Estimated response $\Delta_{y,t} \times \Delta_{x,t-7}$ expressed as deaths per thousand cases, with 95% confidence interval upper and lower bounds.

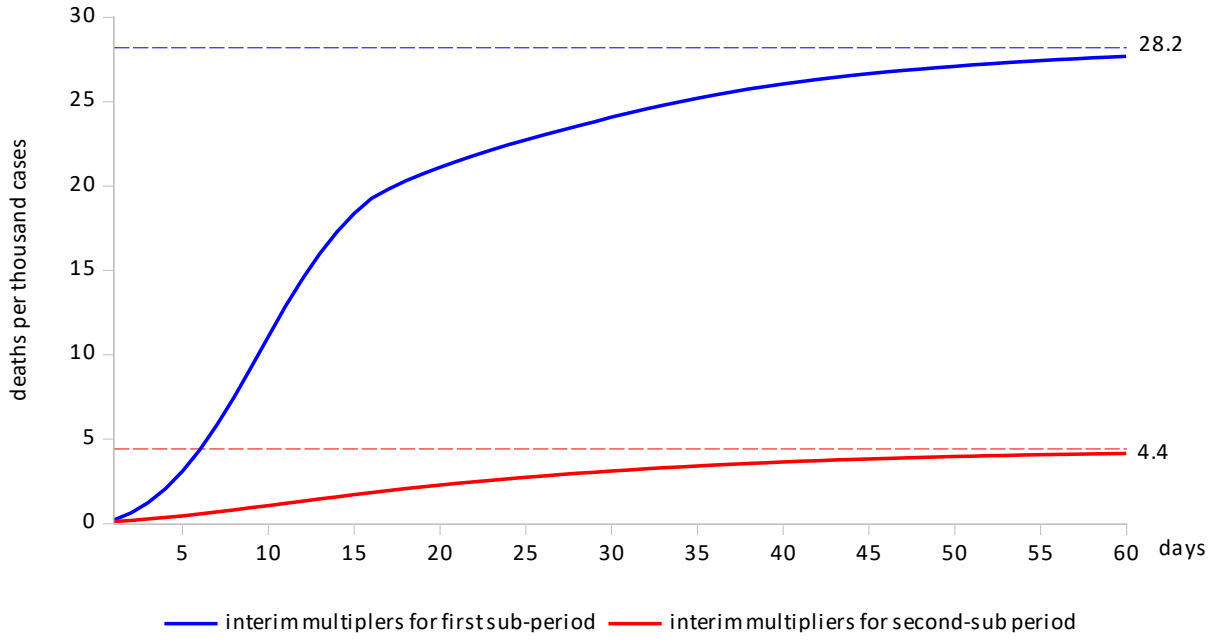

Figure 10. Time paths of interim multipliers $d_{x,l} \times d_{y,l}$ with total multipliers $d_x \times d_y$, expressed as deaths per thousand cases.

Balanced growth models have the advantage of providing a time varying equilibrium response which can therefore be tracked through time. They have the disadvantages of requiring balanced growth to hold, which might not be the case, and a fixed adjustment period dependent on the setting of the lag k. Their specification is also very precise with little flexibility in the setup and, moreover, estimation requires specialised software. ARDL/ECM models, on the other hand, have greater

flexibility in their specification and may be estimated by routine regression software. The adjustment to the long-rum response is freely estimated rather than fixed but the multiplier is not time varying, with evolving relationships having to be investigated through fitting the models over sub-periods of the data, which have to be selected by the investigator or by using break-point procedures and tests, such as those developed by Bai and Perron (1998). Given the limited time period available and the clear waves in the data, we have chosen not to follow this latter route here. However, as more data becomes available, and particularly given the more recent wave of the pandemic associated with the omicron variant, formal testing for break-points would be a useful extension of the ECM modelling approach. Given these competing benefits and drawbacks, it would thus seem sensible to keep an open mind and to use both models to track the behaviour of key Covid variables through time.

While this paper has focused on the relationship between infections, hospital admissions and deaths in England for the period of the pandemic up to the end of October 2021, it is clear that the models may be used for similar data from other countries and time periods. They may also be used for data on infections, admissions and deaths disaggregated into age groups and regions, if such data are available, as it is for England, where such research is ongoing.

Conflict of interest

The author declares that there are no conflicts of interest in this paper.

References

Bai BY, Perron P (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66: 47–78. https://doi.org/10.2307/2998540

Banerjee A, Dolado J, Galbraith JW, et al. (1993) *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data*, Oxford: Oxford University Press.

Dolton P (2021) The statistical challenges of modelling Covid-19. *Natl Inst Econ Rev* 257: 46–82. https://doi.org/10.1017/nie.2021.22

Doornik JA, Castle JL, Hendry DF (2020) Short-term forecasting of the Coronavirus pandemic. *Int J Forecasting* 38: 453–466. https://doi.org/10.1016/j.ijforecast.2020.09.003

Doornik JA, Hendry DF, Castle JL (2021) Statistical short-term forecasting of the COVID-19 pandemic. *J Clin Immunol Immunother* 6.

Engle RF, Granger CWJ (1987) Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55: 251–276. https://doi.org/10.2307/1913236

Gnanvi JE, Salako KV, Kotanmi GB, et al. (2021) On the reliability of prediction of Covid-19 dynamics: a systematic and critical review of modelling techniques. *Infect Dis Model* 6: 258–272. https://doi.org/10.1016/j.idm.2020.12.008

Harvey AC (1989) *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Harvey AC (2020) Time series models for epidemics: leading indicators, control groups and policy assessment. NIESR Discussion Paper 517.

Harvey AC, Kattuman P (2020) Time series models based on growth curves with applications to forecasting Coronavirus. *Harvard Data Sci Rev.* https://doi.org/10.1162/99608f92.828f40de

Harvey AC, Kattuman P, Thamotheram C (2021) Tracking the mutant: forecasting and nowcasting COVID-19 in the UK in 2021. *Natl Inst Econ Rev* 256: 110–126. https://doi.org/10.1017/nie.2021.12

Korolev I (2020) Identification and estimation of the SEIRD epidemic model for COVID-19. *J Econometrics* 220: 63–85. https://doi.org/10.1016/j.jeconom.2020.07.038

Li S, Linton O (2020) When will the Covid-19 pandemic peak? *J Econometrics* 220: 130–157. https://doi.org/10.1016/j.jeconom.2020.07.049

Manski CF, Molinari F (2020) Estimating the COVID-19 infection rate: anatomy of an inference problem. *J Econometrics* 220: 181–192. https://doi.org/10.1016/j.jeconom.2020.04.041

Mills TC (2019) *Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting*, London: Academic Press.

Pesaran MH, Yang CF (2021) Matching theory and evidence on Covid-19 using a stochastic network SIR model. arXiv: 2109.00321v1.

Poletto C, Scarpino SV, Volz EM (2020) Applications of predictive modelling early in the COVID-19 epidemic. *Lancet Digital Health* 2: E498–E499. https://doi.org/10.1016/S2589-7500(20)30196-5

Sen CY (2020) Logistic growth modelling of COVID-19 proliferation and its international implications. *Int J Infect Dis* 96: 582–589. https://doi.org/10.1016/j.ijid.2020.04.085

Spiegelhalter D, Masters A (2021) *Covid by Numbers. Making Sense of the Pandemic with Data*, London: Penguin Random House.

Tovissodé CF, Lokonen BE, Kakai RG (2020) On the use of growth models to understand epidemic outbreaks with applications for COVID-19 data. *PLoS ONE* 15: e0240578. https://doi.org/10.1371/journal.pone.0240578

Vespignani A, Tian H, Dye C, et al. (2020) Modelling COVID-19. *Nat Rev Phys* 2: 279–281. https://doi.org/10.1038/s42254-020-0178-4

AIMS Press