cemmap

centre for microdata methods and practice

When will the Covid-19 pandemic peak?

Oliver Linton

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper
CWP11/20

Economic and Social Research Council

When will the Covid-19 pandemic peak?

Oliver Linton*
University of Cambridge, 2nd April 2020

Abstract

We carry out some analysis of the daily data on the number of new cases and number of new deaths by (191) countries as reported to the European CDC. We work with a quadratic time trend model applied to the log of new cases for each country. This seems to accurately describe the trajectory of the epidemic in China. We use our model to predict when the peak of the epidemic will arise in terms of new cases or new deaths in other large countires. We find that for the UK, this peak is mostly likely to occur within the next two weeks. The total number of cases per day will peak at around 8000 yielding a little more than 255,000 cases in total.

1 Purpose and Main Findings

We compare the progress of COVID-19 on countries worldwide with the aim of finding evidence of future turnaround of the upward trends. We will provide regular updates.

We find that many countries are approaching the peak number of cases per day after which a slow decline should set in. The exceptions are Iran, Russia, and Romania, where it is still too early to estimate when the peak time will occur. For fatalities, which lag behind cases, there are more countries where it is not yet possible to determine an estimate of the peak time, these include: the USA, Germany, Sweden, Austria, Ecuador, Ireland, Romania, Algeria, India, and Peru. The UK does have an estimated peak day for cases, currently April 17th, but this is subject to (large) uncertainty. We are forecasting over a quarter of a million total cases in the UK by the beginning of June when our models are predicting the epidemic will have faded out.

*Thanks to Vasco Carvalho and Giancarlo Corsetti for comments. This is of course work in progress and subject to errors given the timescale in which the work has been done. Cambridge INET and Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: `obl20@cam.ac.uk`.

2 Trend Modelling

We use daily data on new cases and fatalities downloaded from the website of the European Centre for disease protection and control. According to that website, the first case worldwide was recorded as December 31st 2019 (day 1). We have the daily number of (new) cases and the number of (new) deaths upto April 2nd 2020, which is 93 days since day 1. These are nominally count data but the counts get quite big over time and so we consider regressions of the form

$$\log\left(y_{it}+1\right)=m_i(t)+\varepsilon_{it}, \tag{1}$$

where m_i is the trend in mean and y_{it} is either the number of new cases or the number of new deaths in country i; the error term ε_{it} is mean zero.[1] We adopt a general to specific methodology. We first show the nonparametric (rolling window) local linear kernel regression estimate for the number of cases (Figure 1) and number of deaths (Figure 2) in China, which is the country with the longest exposure..

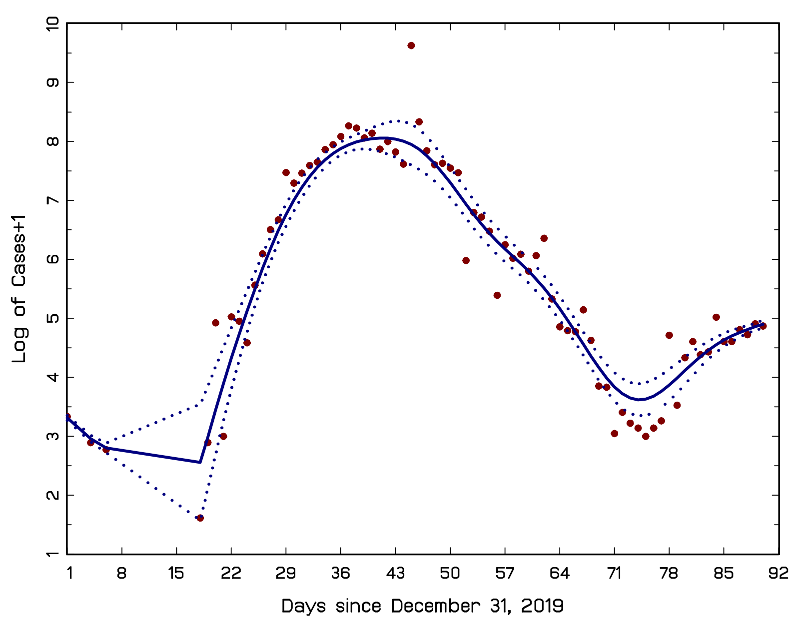

Figure 1. Fitted curve and 95% pointwise confidence band in blue, data points in red.

[1] Adding one to the count is only necessary for some countries with sparse data records or at early stages of the epidemic when zero counts were common.

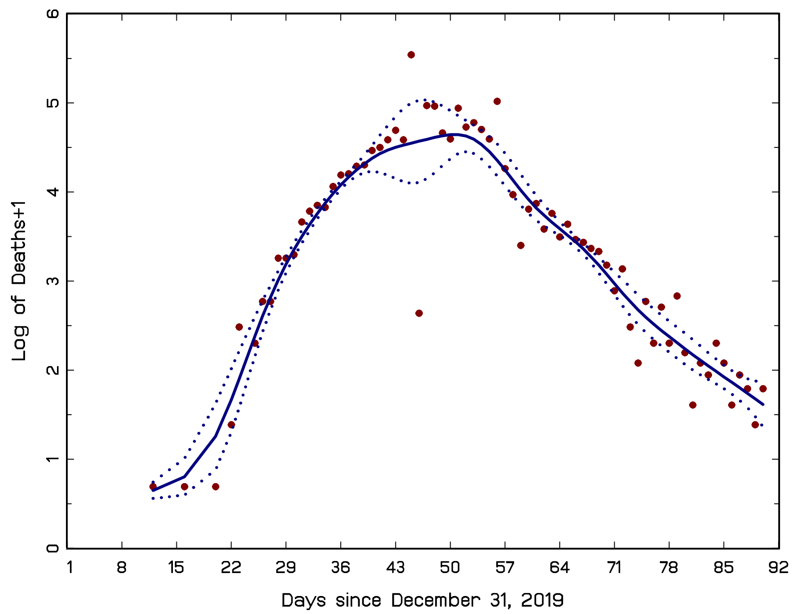

Figure 2. Fitted curve and 95% pointwise confidence band in blue, data points in red.

As one can see from the graphic, China has passed its (first) peak in both cases and deaths and the first part at least of both curves seems well approximated by a quadratic. The central government imposed a lockdown on the city of Wuhan on January 23rd, day 24. In Figure 3 we show the case graph for the UK, which is clearly lagging behind China; the UK has not achieved yet its peak.

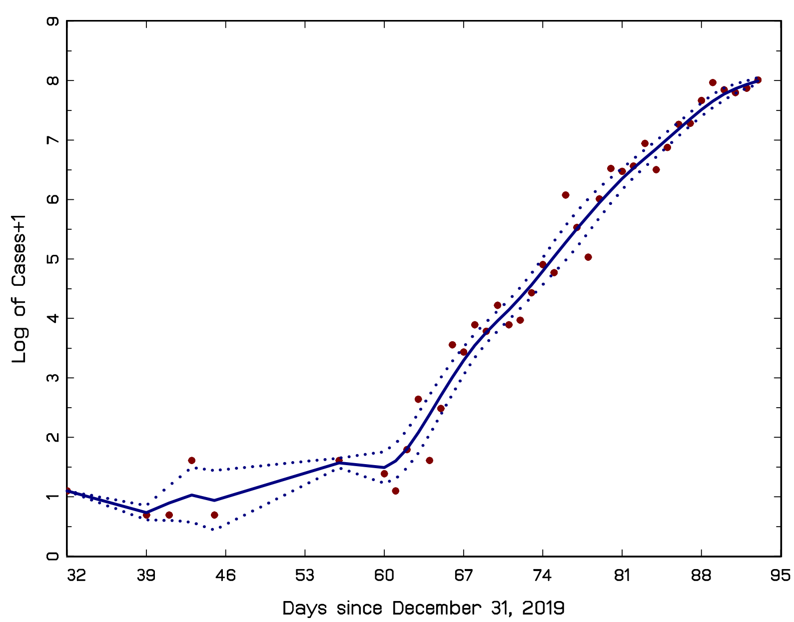

Figure 3. Fitted curve and 95% pointwise confidence band in blue, data points in red.

3 Quadratic Regression

We consider a quadratic regression fit, that is, we take

$$m(t) = \alpha + \beta t + \gamma t^2. \tag{2}$$

A quadratic is the simplest function that reflects the main features of the Chinese data, i.e., the possibility of a turning point; it can also be interpreted as a local approximation to a smooth trend function. Most media reports show log linear trends. We actually work with the count data normalized by population $(y_{it} + 1)/n_i$, where n_i is the population of country i and rescaled time so that $t \to t/T$. Division by population only affects the constant term, but is done to aid comparability across countries. If time is rescaled to the unit interval, then the average trend is $\beta + \gamma$. A well defined maximum of m occurs if $\gamma < 0$ and occurs at the time $t_{\max} = -\beta/2\gamma$, which results in the maximal value of cases per day of $m(t_{\max}) = \alpha - \beta^2/4\gamma$; finally, the value of t after which no cases would be reported (the end of the epidemic) is the larger root of $m(t) = 0$. We explain in more detail in the Appendix how we estimate $t_{\max}$ and provide standard errors for it. Intuitively, we can identify the turning point from the curvature of the regression function when we approach such a turning point.

We consider the thirty countries with the largest number of cases (excluding China) and with at least twenty one days of data. We then fit the quadratic regression on each country with the *most recent twenty one datapoints* (and using rescaled time) and report the results below in Tables 1 (cases) and Table 2 (fatalities).[2] We fully expect the parameters to change over time and to vary across country, but as the peak is approached in a given country, its γ parameter should become significantly negative. The regressions generally have high R^2 (low R^2 may partly reflects data quality and country size). The first order serial correlation in the error term implies some short term predictability relative to the trend and this varies across countries; it would affect somewhat the standard errors although we have not adjusted for this. There is also some heteroskedasticity although this is generally decreasing over time. The contemporaneous correlation across country residuals is quite variable with some large and positive and some large and negative, perhaps reflecting different lags in reporting and time zone effects. There are quite a few negative γ parameters in Table 1, which are indicative of approaching future turning points, although mostly they are not statistically significant, yet. The α and β parameters are all strongly significant so we do not show the standard errors. There is quite a bit of heterogeneity across the parameters β, γ (with less so for α) consistent with different countries being at different stages of the cycle and taking different approaches to

[2] The UK imposed school closures from March 13th, which is now more than 21 days ago. Other countries adopted similar measures around the same time.

managing the epidemic.

Country	α	β	γ	se(γ)	R^2	$\rho(1)$	Total Cases	Population
USA	-14.3125	9.4268	-4.6052	0.5373	0.9850	0.3467	216721	327167434
Italy	-11.0931	5.6633	-4.2315	2.3604	0.2893	-0.3111	110574	60431283
Spain	-11.0982	4.8379	-2.3761	0.5548	0.9415	0.0859	102136	46723749
Germany	-12.1490	5.5127	-3.0343	1.0746	0.8152	0.0137	73522	82927922
France	-11.7121	3.0561	-0.7537	0.6586	0.9094	-0.3451	56989	66987244
Iran	-11.1379	-0.6147	1.8394	0.5673	0.8356	0.6064	47593	81800269
UK	-13.3646	5.0703	-1.4802	0.9269	0.9246	-0.1067	29474	66488991
Switzerland	-11.5958	5.9227	-3.3310	4.1608	0.2442	-0.2045	17070	8516543
Belgium	-11.9426	4.6134	-1.6726	0.9162	0.8935	0.3095	13964	11422068
Netherlands	-12.1888	5.2700	-2.8212	0.4492	0.9608	0.1977	13614	17231017
Austria	-11.6981	6.1328	-4.1233	0.6960	0.8845	-0.1412	10711	8847037
South Korea	-13.4930	2.3713	-3.0801	3.0653	0.1084	-0.4494	9976	51635256
Sweden	-14.0774	5.0261	-1.1989	1.0232	0.9200	0.0185	9595	37058856
Portugal	-13.2761	7.6712	-3.8620	0.6789	0.9627	-0.3571	8251	10281762
Brazil	-16.5366	7.3719	-3.4665	1.2763	0.8842	0.1202	6836	209469333
Israel	-13.5891	6.6592	-2.6750	2.6257	0.6518	-0.5045	5591	8883800
Australia	-14.1573	7.7634	-5.1349	1.2466	0.8007	-0.5185	4976	24992369
Sweden	-11.2265	-1.1258	2.5130	0.9716	0.7054	0.0655	4947	10183175
Norway	-12.3926	7.1143	-5.0232	3.1392	0.3031	-0.5062	4665	5314336
Czech R	-12.9820	5.0408	-2.7103	1.2042	0.7556	-0.0285	3589	10625695
Ireland	-12.6255	6.2997	-3.5000	0.6953	0.9309	-0.1633	3447	4853506
Denmark	-11.0749	-1.8215	3.1077	1.2686	0.5847	-0.0151	3107	5797446
Chile	-14.7922	7.4519	-3.7148	0.9324	0.9293	-0.3530	3031	18729160
Malaysia	-14.5640	9.2521	-8.4414	3.3779	0.2643	-0.4478	2908	31528585
Russia	-16.6184	2.2419	1.8998	3.2103	0.5969	-0.1240	2777	144478050
Ecuador	-15.9481	11.7626	-7.4208	1.8898	0.8215	0.1415	2758	17084357
Poland	-14.7839	4.9477	-2.0883	0.5328	0.9591	-0.5520	2554	37978548
Romania	-14.0305	3.0317	0.0117	1.1811	0.8482	0.2316	2460	19473936
Luxembourg	-12.6352	13.2933	-9.6232	2.2552	0.7310	-0.2782	2319	607728
Phillipines	-16.5176	2.9432	-0.1036	5.5562	0.1796	-0.3447	2311	106651922

Table 1. Estimation using daily counts data 20200313-20200402

Country	α	β	γ	se(γ)	R^2	$\rho(1)$	Total Deaths	Population
Italy	-12.9006	4.0756	-2.4557	0.4785	0.9071	-0.3053	13157	60431283
Spain	-14.9671	8.1665	-4.0182	1.4636	0.8678	-0.6713	9053	46723749
USA	-17.9453	4.7118	0.7879	2.7086	0.7787	-0.2524	5138	327167434
France	-15.9321	6.8616	-2.7310	0.8552	0.9499	-0.1035	4032	66987244
Iran	-13.8389	1.8565	-1.4295	0.3379	0.6603	0.5645	3036	81800269
UK	-16.9846	6.3599	-1.3422	1.8467	0.8588	-0.2187	2532	66488991
Netherlands	-16.7969	10.3246	-5.3589	1.1746	0.9362	-0.1844	1173	17231017
Germany	-18.2122	5.0956	0.1851	2.4304	0.7993	-0.1128	872	82927922
Belgium	-17.0691	8.2353	-2.3820	1.6071	0.9156	-0.0571	828	11422068
Switzerland	-15.8739	6.4381	-3.1291	2.1130	0.6670	-0.0762	378	8516543
Brazil	-19.8830	6.0275	-1.5849	0.8013	0.9618	0.1975	241	209469333
Sweden	-16.1076	2.2629	1.5538	1.7932	0.8000	-0.0696	239	10183175
Portugal	-16.8513	5.2122	-0.9995	0.9854	0.9378	0.0801	187	10281762
South Korea	-17.0576	4.0781	-3.4406	1.9292	0.1853	-0.2472	169	51635256
Indonesia	-19.2383	3.1119	-0.3419	2.3334	0.5397	0.0884	157	267663435
Austria	-16.0776	2.3448	0.8469	2.3260	0.6199	-0.1665	146	8847037
Ecuador	-16.4738	0.1784	2.9445	1.8745	0.7287	-0.2337	146	17084357
Canada	-17.6726	3.8961	-1.2054	1.8156	0.6420	-0.3177	109	37058856
Denmark	-15.9087	3.7856	-0.8277	1.7464	0.7025	-0.4326	104	5797446
Phillipines	-18.1054	2.1168	-0.5594	2.7179	0.2121	-0.5559	96	106651922
Ireland	-15.2887	-0.9702	4.0080	1.2455	0.8616	-0.2129	85	4853506
Romania	-16.9669	0.2331	3.0008	1.2541	0.8651	0.0408	85	19473936
Algeria	-16.6761	-2.3490	3.5724	2.0318	0.3620	-0.3090	58	42228429
Japan	-17.5776	0.1684	-0.4877	1.4760	0.0488	-0.1553	57	126529100
Czech R	-16.3248	0.3400	2.2411	1.7256	0.6809	-0.1770	57	10627165
Iraq	-17.9563	5.4633	-4.1703	1.6886	0.4090	-0.0590	50	38433600
India	-20.5433	-1.4750	3.3001	1.4918	0.6373	-0.1865	50	1352617328
Greece	-16.3289	3.2257	-1.5306	1.4804	0.5167	0.0491	50	10727668
Peru	-17.1544	-0.9264	3.1541	1.2375	0.7761	-0.0184	47	31989256
Egypt	-18.4450	2.2408	-0.6069	1.6517	0.4449	0.0872	46	98423595

Table 2. Estimation using daily fatality data 20200313-20200402

The count data is subject to many errors, perhaps the main one is the undercount due to under-testing in some countries. The fatality data is perhaps more accurate although this is also subject to some errors. Some countries such have reached significantly negative γ already in Table 2, although the USA, Germany, Sweden, Austria, Ecuador, Ireland, Romania, Algeria, India, and Peru have not yet.

4 Turning Point Estimation

We present the estimated turnaround time for selected countries (which have $\gamma < 0$) along with the 95% confidence interval in Tables 3 and 4. Surprisingly, the USA is predicted to turnaround in around 5 days (plus or minus 6 days, so with high confidence before two weeks) in terms of case; however, it does not have a prediction yet in terms of deaths as the curve has some way to go to turnaround. The UK is predicted to turnaround in around 15 days (plus or minus 31 days) in cases and sooner in terms of deaths. The peak in cases should precede the peak in deaths but this is not imposed in our estimation, and data issues may lead to violations of this.

Country	Turnaround in Days	$\pm Days$
USA	0.4932	2.5280
Spain	0.3783	5.0076
France	21.5757	55.3872
UK	14.9678	31.4511
Belgium	7.9602	19.8954
Canada	23.0185	56.5466
Brazil	1.3297	8.5763
Israel	5.1383	30.1871
Chile	0.0633	5.2301
Poland	3.8768	7.2129
Phillipines	277.1621	30789.2551

Table 3. Predicted turnaround in number of cases in days from 20200402

Country	Turnaround in Days	$\pm Days$
Spain	0.3397	7.7847
France	5.3808	9.7803
UK	28.7553	106.9068
Belgium	15.3019	34.3364
Switzerland	0.6038	14.7777
Brazil	18.9335	29.3836
Portugal	33.7526	86.4308
Indonesia	74.5718	1154.9988
Canada	12.9381	69.5772
Denmark	27.0234	156.6458
Phillipines	18.7332	280.4359
Greece	1.1291	22.1521
Egypt	17.7686	151.8613

Table 4. Predicted turnaround in number of deaths in days from 20200402

In Figure 4 we show the level extrapolation curve for the UK using the most recent data; curve in blue, data points in red. The extrapolation curve is a scaled Gaussian density function as mandated by our curve model. A peak number of 8000 cases per day is forecast with over 255000 total cases. The peak is achieved around 17th April, although the confidence interval around that is still quite wide.

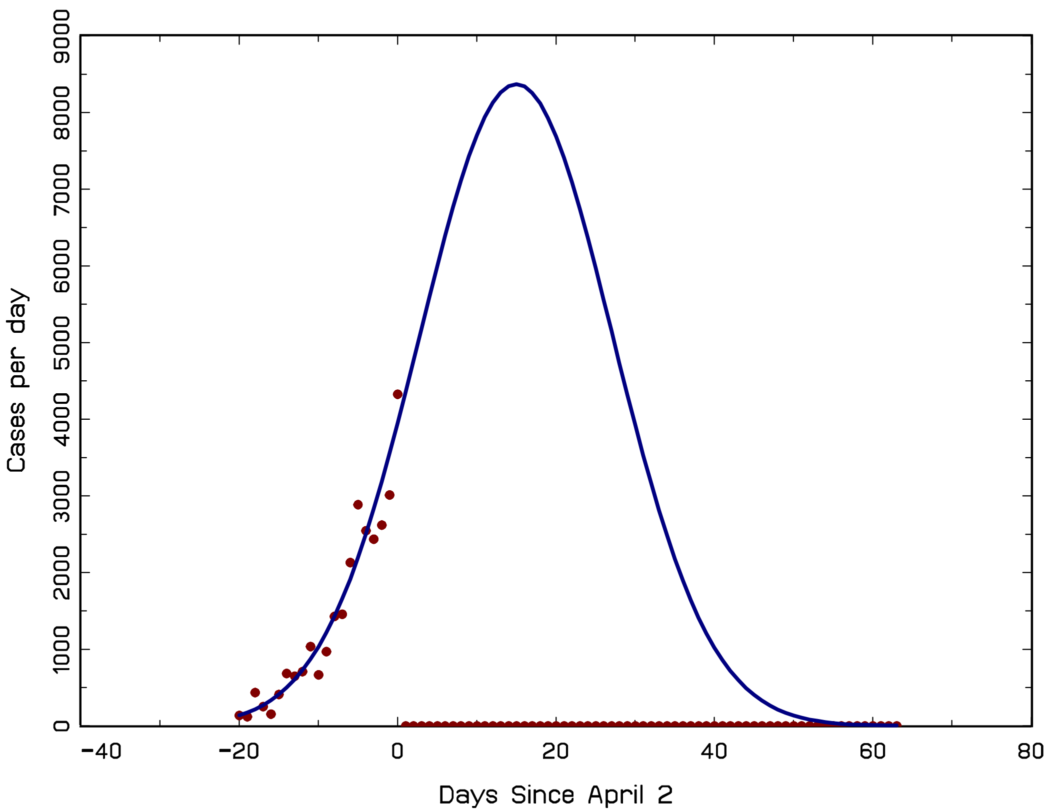

Figure 4. Fitted and extrapolated curve in blue, data points in red.

5 Combining Case and Fatality Models

Fatalities should follow cases and total fatalities should be a fraction of the total cases reported. For this reason we consider the following model, which imposes that the fatality curve is a delayed and shifted (because this is the log of cases) version of the case curve. Let y_{it}^d denote deaths and y_{it}^c denote cases, where:

$$\log\left(y_{it}^d + 1\right) = m_i^d(t) + \varepsilon_{it}^d,$$

$$\log\left(y_{it}^c + 1\right) = m_i^c(t) + \varepsilon_{it}^c.$$

We suppose that for some $\theta_i < 0$ and $k_i \geq 0$

$$m_i^d(t) = \theta_i + m_i^c(t-k). \tag{3}$$

This imposes restrictions across the coefficients of the two quadratic equations. The turning point for m occurs k periods after the turning point for g, which is defined as before in terms of the γ, β parameters. We discuss in the appendix how we estimate this model. In Table 5 we report estimates of θ, β, γ along with their standard errors, we use the delay factor of $k = 10$ days.

Country	θ	se(θ)	β	se(β)	γ	se(γ)	R^2	Total Deaths	Population
Italy	-1.2110	0.2324	2.2492	0.4734	-0.7766	0.6301	0.7987	11591	60431283
Spain	-1.0009	0.1486	5.5035	0.3027	-2.5955	0.4030	0.9624	7340	46723749
USA	-1.7146	0.2610	6.1630	0.5318	-0.8944	0.7079	0.9445	3170	327167434
France	-1.4463	0.1062	4.8529	0.2164	-2.3220	0.2881	0.9805	3024	66987244
Iran	-2.0831	0.0891	0.7660	0.1816	0.4512	0.2417	0.9723	2757	81800269
UK	-0.8831	0.1923	5.8608	0.3919	-1.8835	0.5216	0.9492	1408	66488991
Netherlands	-1.1768	0.1335	5.6064	0.2721	-2.6997	0.3621	0.9720	864	17231017
Germany	-3.2716	0.2609	5.5237	0.5314	-2.1281	0.7074	0.9531	583	82927922
Belgium	-1.4552	0.2056	5.5356	0.4188	-1.5709	0.5575	0.9502	513	11422068
Switzerland	-2.4542	0.4015	3.7730	0.8179	-0.7195	1.0888	0.8318	295	8516543
South Korea	-2.8294	0.1798	0.5210	0.3662	-0.4921	0.4875	0.9170	163	51635256
Brazil	-1.3733	0.1630	4.6957	0.3322	-0.7802	0.4422	0.9623	159	209469333
Sweden	-2.5487	0.2180	3.1632	0.4442	-1.4006	0.5913	0.9289	146	10183175
Portugal	-1.6133	0.1869	4.5786	0.3808	-0.0316	0.5069	0.9589	140	10281762
Indonesia	-0.8911	0.3318	3.1287	0.6760	0.0666	0.8998	0.7592	122	267663435
Austria	-3.3047	0.2346	3.3357	0.4780	-0.6556	0.6363	0.9479	108	8847037
Canada	-2.3077	0.2090	3.0522	0.4258	1.2196	0.5668	0.9501	89	37058856
Phillipines	-0.5942	0.4425	1.8437	0.9014	2.8205	1.2000	0.6445	88	106651922
Denmark	-2.7373	0.2493	2.5178	0.5079	-1.4210	0.6761	0.8994	77	5797446
Ecuador	-0.9999	0.3167	3.3577	0.6452	0.7964	0.8589	0.8287	62	17084357
Japan	-2.6088	0.2075	-0.5276	0.4228	2.1725	0.5628	0.8952	56	126529100
Ireland	-2.3257	0.1862	2.5336	0.3794	0.9921	0.5050	0.9521	54	4853506
Iraq	-0.7828	0.3067	1.6576	0.6248	1.5704	0.8317	0.7175	46	38433600
Romania	-1.9554	0.1589	2.5345	0.3238	0.9654	0.4310	0.9582	44	19473936
Greece	-2.0835	0.1887	1.9768	0.3844	-0.3164	0.5117	0.9157	43	10727668
Czech R	-0.0691	0.3436	2.2303	0.7000	3.4192	0.9319	0.7722	42	10627165
Egypt	-1.1679	0.2703	2.1682	0.5507	0.0131	0.7330	0.7736	40	98423595
Malaysia	-2.7703	0.2157	2.5369	0.4395	-0.4786	0.5850	0.9340	37	31528585
Morocco	-0.8666	0.2229	1.6626	0.4541	2.0524	0.6045	0.8565	33	36029138
India	-1.5713	0.2176	1.7441	0.4434	1.9060	0.5902	0.9025	32	1352617328

Table 5. Cases and Fatality data 20200313-20200402

Figure 5 shows the predicted cases and fatalities per day for the UK.

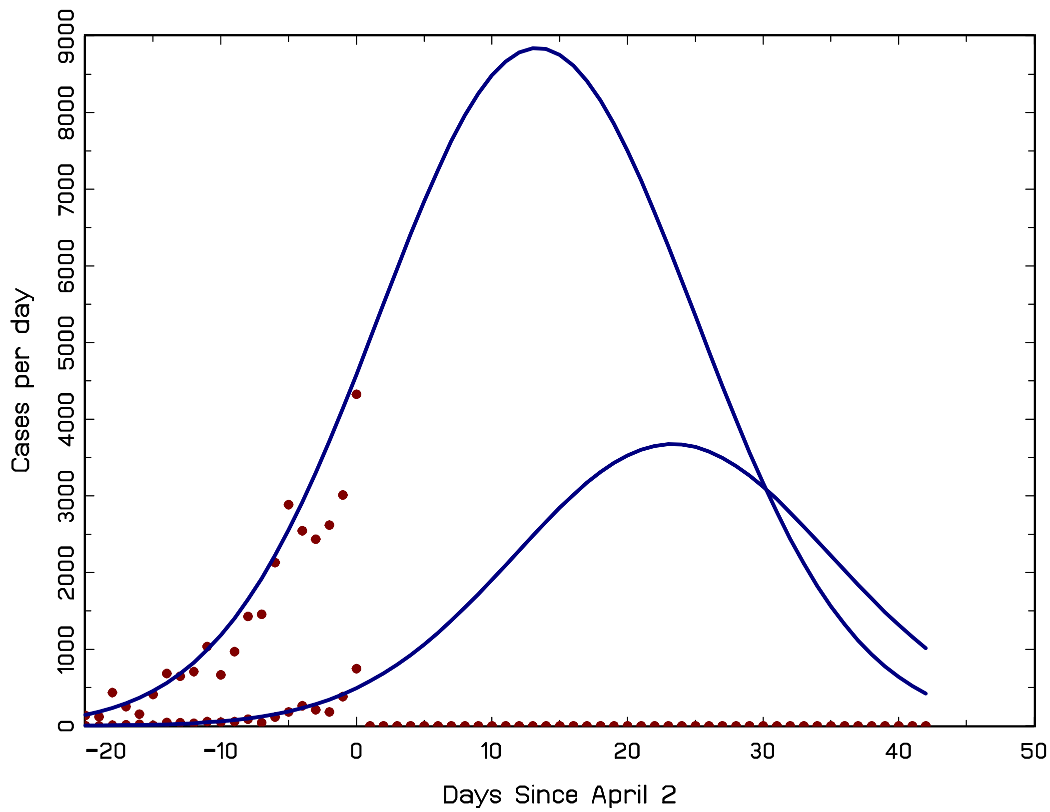

Figure 5. Fitted and extrapolated curves in blue, data points in red.

6 Stock Market Reaction

In Figure 5 we show the level of the FTSE100 stock market index along with the number of epidemic cases in the UK. The market started moving downward before a substantial rise in UK cases had occurred, it responded to earlier news from around the world suggestive of the impending crisis.

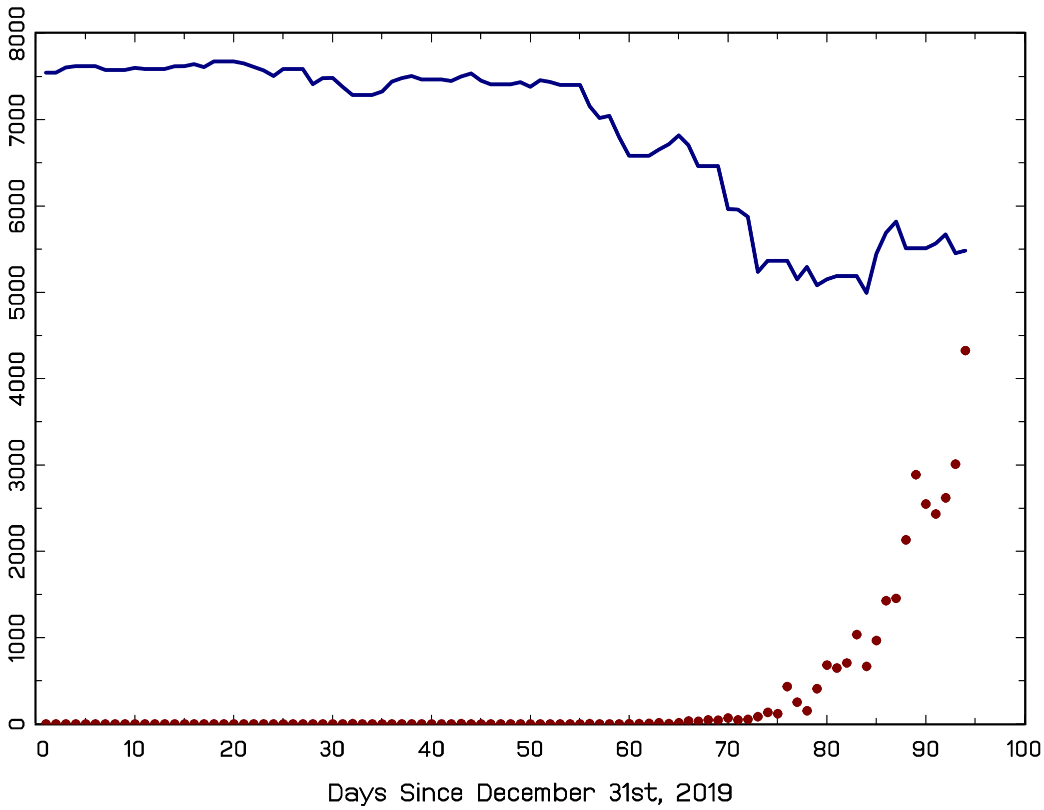

Figure 5. FTSE100 index level in blue (with scale on level) and number of cases in red.

In Figure 6 we report the daily returns over the same period. From this figure one can see the increase in volatility.

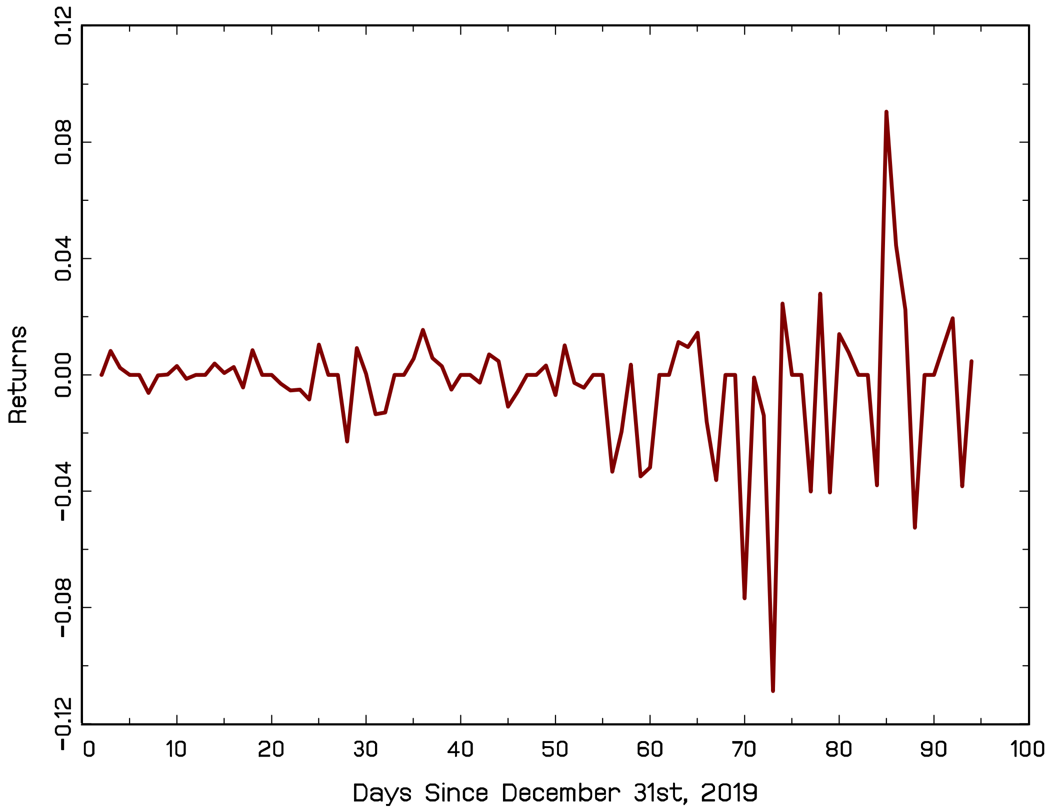

Figure 6. Daily stock returns on the FTSE100

In Figure 7 we show the comparison between the FTSE100 and the FTSE250 indexes; the ratio jumped up rapidly on day 77.

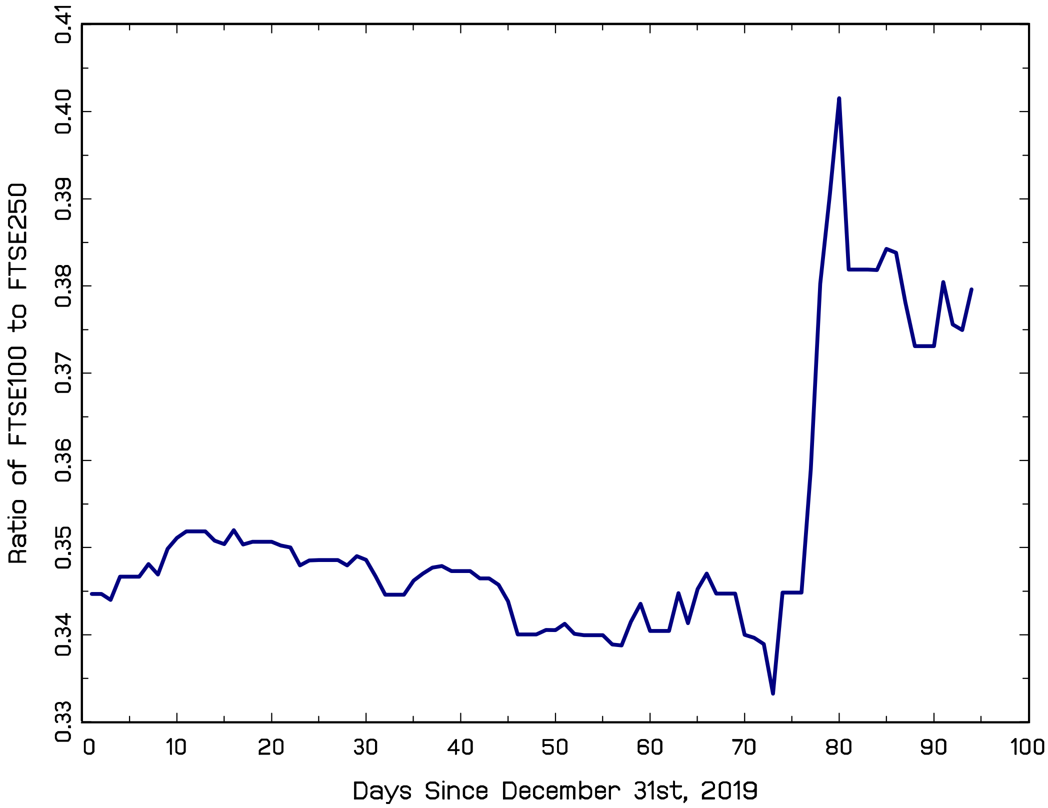

Figure 7. Ratio of FTSE100 price level to FTSE250 level

7 Appendix

Suppose that $m(t) = \alpha + \beta t + \gamma t^2$. A well defined peak of m occurs provided $\gamma < 0$ and occurs at the point $t_{\max} = \frac{\beta}{-2\gamma}$ with maximal value $m(t_{\max}) = \alpha + \beta t_{\max} + \gamma t_{\max}^2 = \alpha - \frac{\beta^2}{4\gamma}$. The value of t when no cases will be achieved is the larger root of $m(t) = 0$, i.e., $t_0 = (-\beta + \sqrt{\beta^2 - 4\alpha\gamma})/2\gamma$. The total number of cases from the beginning of the estimation period is $N_{\max} = \int_0^{t_0} \exp(m(s))ds - t_0$. The lethality of the epidemic would be the ratio of $N_{\max}$ for fatalities to $N_{\max}$ for cases.

With OLS estimators $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})^{\intercal}$, we estimate

$$\widehat{t}_{\max} = \frac{\widehat{\beta}}{-2\widehat{\gamma}}, \quad \widehat{m}(\widehat{t}_{\max}) = \widehat{\alpha} - \frac{\widehat{\beta}^2}{4\widehat{\gamma}}, \quad t_0 = \frac{-\widehat{\beta} + \sqrt{\widehat{\beta}^2 - 4\widehat{\alpha}\widehat{\gamma}}}{2\widehat{\gamma}}, \tag{4}$$

The standard errors are available from the delta method. We have:

$$\mathrm{avar}(\widehat{t}_{\max}) = \left(0 - \frac{1}{2\gamma}\ \frac{\beta}{2\gamma^2}\right) \mathrm{avar}(\widehat{\theta}) \begin{pmatrix} 0 \\ -\frac{1}{2\gamma} \\ \frac{\beta}{2\gamma^2} \end{pmatrix},$$

$$\mathrm{avar}(\widehat{m}(\widehat{t}_{\max})) = \begin{pmatrix} 1 - \frac{\beta}{2\gamma} \, \frac{\beta^2}{4\gamma^2} \end{pmatrix} \mathrm{avar}(\widehat{\theta}) \begin{pmatrix} 1 \\ -\frac{\beta}{2\gamma} \\ \frac{\beta^2}{4\gamma^2} \end{pmatrix},$$

where $\mathrm{avar}(\widehat{\theta})$ is the covariance matrix of the least squares parameter estimates. Results are available from the author upon request.

We predict y_s by $\exp(\widehat{\alpha} + \widehat{\beta}s + \widehat{\gamma}s^2) + 1$; we multiply this by the mean of the exponential of the OLS residuals to correct for the transformation bias. We estimate $N_{\max}$ by $\widehat{N}_{\max} = \int_0^{\widehat{t}_0} \exp(\widehat{\alpha} + \widehat{\beta}s + \widehat{\gamma}s^2)ds - \widehat{t}_0$, which can be approximated by the sum of $\exp(\widehat{m}(t))$ over some long extrapolation into the future; Standard errors for this quantity can also be obtained by the delta method.

We next discuss how to estimate the restricted model of Section 5. For given k, we have with x being cases and y fatalities:

$$y^c = \left(1, t, t^2\right)_{t=1}^{T} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} + \varepsilon^c, \quad y^d = \left(1, t-k, (t-k)^2\right)_{t=1}^{T} \begin{pmatrix} \alpha' \\ \beta \\ \gamma \end{pmatrix} + \varepsilon^d,$$

and stacking together we obtain the larger regression

$$\begin{aligned} y = \begin{pmatrix} y^c \\ y^d \end{pmatrix} &= \begin{pmatrix} 1 & 0 & t & t^2 \\ 0 & 1 & t-k & (t-k)^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \alpha' \\ \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \varepsilon^c \\ \varepsilon^d \end{pmatrix} \\ &= Xb + \varepsilon, \end{aligned}$$

where $y, \varepsilon \in \mathbb{R}^{2T}$, X is $2T \times 4$ and b is 4×1. We have $\theta = \alpha' - \alpha$. Estimation of b for given k is straightforward, pooled OLS. Estimation of k can be done by searching over the integer values and choosing the value that maximizes the likelihood.