UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge-INET Institute

Cambridge-INET Working Paper Series No: 2020/11

Cambridge Working Papers in Economics: 2025

WHEN WILL THE COVID-19 PANDEMIC PEAK?

Shaoran Li
(University of Cambridge)

Oliver Linton
(University of Cambridge)

Updated 14th May 2020

We carry out some analysis of the daily data on the number of new cases and the 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 countries.

Keywords: Epidemic; Nonparametric; Prediction; Trend;
JEL Classification: C10

When will the Covid-19 pandemic peak?*

Shaoran Li† University of Cambridge

Oliver Linton‡ University of Cambridge

May 14, 2020

Abstract

We carry out some analysis of the daily data on the number of new cases and the 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 countries.

KEYWORDS: Epidemic; Nonparametric; Prediction; Trend;

JEL CLASSIFICATION: C10

1 Introduction

We provide a model of daily data on the number of new cases of and new deaths from COVID-19 in countries worldwide. Our model is essentially a regression of log outcomes on a quadratic trend function for each country. Since April 5th 2020 we have been estimating our model every day and providing outlooks for the future development of the pandemic in different countries with a view to estimating turnaround dates (before they actually accrued), peak to trough times, and the total number of cases. The results are updated daily at the website `http://covid.econ.cam.ac.uk/linton-uk-covid-cases-predicted-peak`; R-code is also provided. We provide various robustness tests: we consider quantile estimation in place of mean estimation

*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.

†Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: `sl736@cam.ac.uk`.

‡Cambridge INET and Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: `obl20@cam.ac.uk`.

to limit the effect of large measurement error; we consider a "fat-bottomed" model that allows a slower evolution around the peak; we provide a one-step ahead forecasting exercise that keeps track of the model performance. We evaluate the residuals from our model along several dimensions. In particular, we find little evidence of day of the week effects in most countries. We find some evidence of autocorrelation in the residuals but it varies widely across countries with an overall mean effect negative but close to zero consistent with stale reporting. We find some limited evidence of time varying cross-sectional heteroskedasticity. We find the error distribution pooled across countries is not far from symmetric for cases and slightly less so for deaths. There appears to be substantial cross sectional contemporaneous correlation between the errors from cases and deaths within a country (positive) and between cases in one country and another (here, the mean is small and positive but there is wide dispersion). The SUR model we fit cannot benefit from GLS, but the error properties can affect the standard errors and test statistics. We also develop a joint model for the evolution of new cases and new deaths and test the restrictions it implies and estimate the restricted model. Estimation of this model does exploit the error covariance matrix to achieve full efficiency.

Our model is purely statistical and we do not pretend to model the disease dynamics per se, just the data. However, the epidemiological models themselves do not have perfect forecasting records and our approach is complementary to the large literature produced by professional epidemiologists and biostatisticians.[1] One advantage of our model over dynamic epidemiological models is that the publicly available data from many countries is subject to a wide range of errors that make dynamic models very suspect without some additional model for the measurement error, which typically requires many untestable assumptions.

Literature Review. Many researchers apply or extend the SIR or its variants (SIRD, SEIR) to model the dynamics of the Covid-19 outbreak. For example, Wu, Leung, Leung (2020)[28], Anastassopoulou et.al (2020)[2], and Lin, Hu and Zhou (2020)[15]. A short description of the baseline SIR model is as follows. The fixed population (N) can be split into three nonintersecting classes: susceptibles (S), infectious (I) and recovered (R). The number of individuals in each of these classes changes over time and satisfies

$$N = S(t) + I(t) + R(t).$$

The dynamics of each class can be described using ordinary differential equations (ODEs) as follows:

$$\frac{dS}{dt} = -\beta IS, \quad \frac{dI}{dt} = \beta IS - \alpha I, \quad \frac{dR}{dt} = \alpha I,$$

where β is the transmission rate constant, βI is the force of infection, and βIS is the number of individuals who become infected per unit of time. The susceptible individuals who become infected

[1] They do say that Noah's Ark was built and sailed by amateurs, whereas the Titanic was built and sailed by professionals.

move to the class I and individuals who recover (the constant recovery rate is α) leave the infectious class and move to the recovered class. Equipped with initial conditions $S(0), I(0), R(0)$, the model can be easily solved. Elaborations on the basic SIR model that are used in the study of Covid19 includes SIRD (eg. Anastassopoulou et.al (2020)[2]) and SEIR (eg. Wu, Leung, and Leung (2020)[28] as well as Read et al. (2020)[23]). Since there are daily data available on the number of deaths, an additional class 'Deceased' (D) can be included as well, which corresponds to SIRD model. The other popular variant of baseline SIR is SEIR, which separately considers susceptibles (S) and exposed (E).

Chudik, Pesaran and Rebucci (2020)[7] contrast government-mandated social distancing policies with voluntary self-isolation in an SEIR model. They decompose the population, N, into two categories: the exposed N_E and the rest, N_I, that are isolated. The strength of mitigation policy can be measured by $1 - \lambda$ where $\lambda = \frac{N_E}{N}$, and it integrates social distancing policy in the traditional SIR model. They evaluate the costs and benefits of alternative societal decisions on the degree and the nature of government-mandated containment policies by considering alternative values of λ in conjunction with an employment loss elasticity α. They found that the employment loss can be reduced if the social distancing policy is targeted towards people that are most likely to spread the infection. Besides, other articles also discuss this model, see Wu, Leung, and Leung (2020)[28] , Read et al. (2020)[23] , and Peng (2020) et al. [21] for details.

One feature of Covid-19 is that different sub-populations face different risks. Taking that into account, Acemoglu et.al (2020)[1] develop a heterogeneous agent multi-risk SIR model (MR-SIR) where infection, hospitalization and fatality rates vary between groups. Describing the laws of motion of the susceptible, infectious and recovered populations by group, they found that better social outcomes can be achieved with targeted policy that applies an aggressive lockdown on the oldest group instead of a uniform lockdown policy.

Liu, Moon and Schorfheide (2020) [17] decomposes the growth rates of the Covid-19 infections into a deterministic component which approximates the path predicted by a deterministic SIR model and a stochastic component that could be interpreted as either time-variation in the coefficients of an epidemiological model or deviations from such a model.

However, these kinds of models have a high requirement on the data availability and quality, to which the estimates are very sensitive. One of the limitations of Covid-19 data is that the number of susceptible and recovered people are seldom reported. Apart from this reason, Li et al. (2020)[14] point out the limitations of SIR style models, arguing that the real situation could be much more complicated and changing all the time.

Another strand of time series analysis originated from Generalized Logistic Model (GLM), firstly proposed by Richard (1959)[24], developed for modelling the growth, and allowing for more flexible S-shaped aggregation, whence

$$Y(t) = A + \frac{K - A}{(1 + Qe^{-Bt})^{1/v}}$$

where A and K are lower and upper asymptotes; B is the growth rate; v affects the maximum asymptote growth; and Q is related to the value $Y(0)$. An extension of GML is sub-epidemic modelling studied by Chowell, Tarip and Hyman (2019)[6], where they model each group sub-epidemic by a GLM and then comprise a set of n overlapping sub-epidemics to describe the whole epidemic waves. Roosa et al. (2020)[25] extend the aforementioned GML to fit the cumulative number of confirmed Covid-19 cases and forecast short-run new cases in Hubei and other provinces in China, and they found the S-shape curve fits the initial data well, based on the information until 9th Feb. Additionally, they also compare their model with the Richard curve and sub-epidemic modelling. All models provide good visual fits to the epidemic curves, and their estimates obtained from GLM consistently illustrate that the epidemic growth is nearly exponential in Hubei and sub-exponential in other provinces at early stages.

Apart from GLM, several other time trend modelling are worth reviewing, some of which verifies the parabola time trend of daily new cases and deaths. Deb et al. (2020)[8] showed that a time-dependent quadratic trend successfully captures the log incidence pattern of the disease among Chinese provinces. Li, Feng and Quan (2020)[14]use time series modelling to fit the infections and fatalities in China, an apparent quadratic trend and turning points are found. They also demonstrate that the distribution of daily deaths is similar to that of daily infections with a 5 to 6 days delay. Furthermore, Flaxman et al. (2020) develop a semi-mechanistic Bayesian hierarchical model to accommodate the impact of government interventions on both daily new infections and fatalities among 11 European countries. Especially, most of their online plots based on data from European CDC present second-order polynomial time trend, which are consistent with the evidence from China. Moreover, Hafner (2020) [10] consider both serial correlation and spillover effects of Covid-19 by spatial autoregressive models. Finally, Hu et.al (2020)[12] apply artificial intelligence method for real-time forecasting of Covid-19 cases data. By using cumulative confirmed cases data until the end of February 2020, they predicted that the provinces/cities would enter plateaux eventually, although with varied time points.

2 Trend Modelling

There are basically three types of variables studied in time series econometrics: deterministic trend or seasonal variables, stationary stochastic processes, and nonstationary variables. The stationary parts are well understood and studied in the literature. On the other hand, the class of nonstation-

ary processes is extremely broad, and different types of nonstationarity can generate quite different behavior and require quite different analytical techniques. There are two main approaches to depict the structure of nonstationary data. One is the unit root theory for integrated time series, or similar techniques for fractional integrated time series that covers unit root process as a special case. This theory and associated techniques are studied and developed by Park and Phillips (1999,2001)[19][20], Marinucci and Robinson (2001)[18], Hualde and Robinson (2011)[13], and Wang (2014,2015)[26][27], among others. The other is locally stationary processes and deterministic trends. Usually in economics, data series feature increasing trends that are quite close to linear. For epidemic data, the situation is more complex so that typically there is an upward exponential growth, a plateaux, and a declining trend part, all of which need to be captured.

We use daily data on new cases and fatalities downloaded from the website of the European Centre for Disease Protection and Control (ECDC). 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 today's date, which is 130+ days at the time of writing since day 1. These are count data with some zeros initially but the counts get quite big over time and so we consider regressions of the form

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

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.[2] 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 upto day 90, which is the country with the longest exposure.

[2] 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.

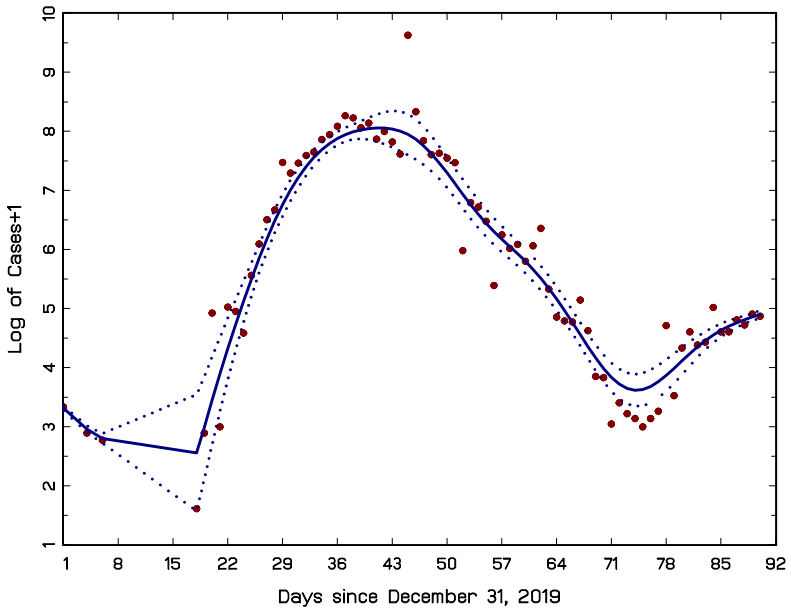


Figure 1. Fitted curve and 95% pointwise confidence band in blue, data points in red.

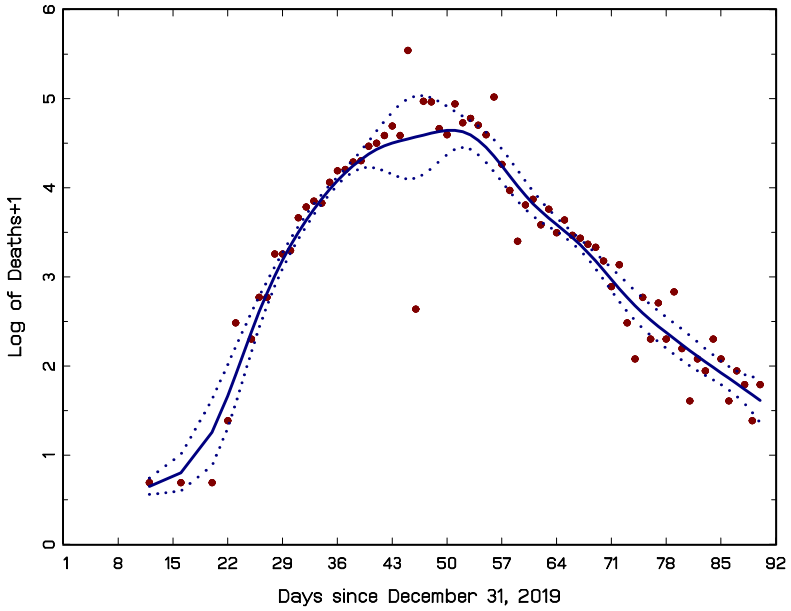


Figure 2. Fitted curve and 95% pointwise confidence band in blue, data points in red.

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 was clearly lagging behind China; the UK had not achieved yet its peak when this project began, which was the motivation for finding methods that could predict when that would occur using just the cases and deaths data.

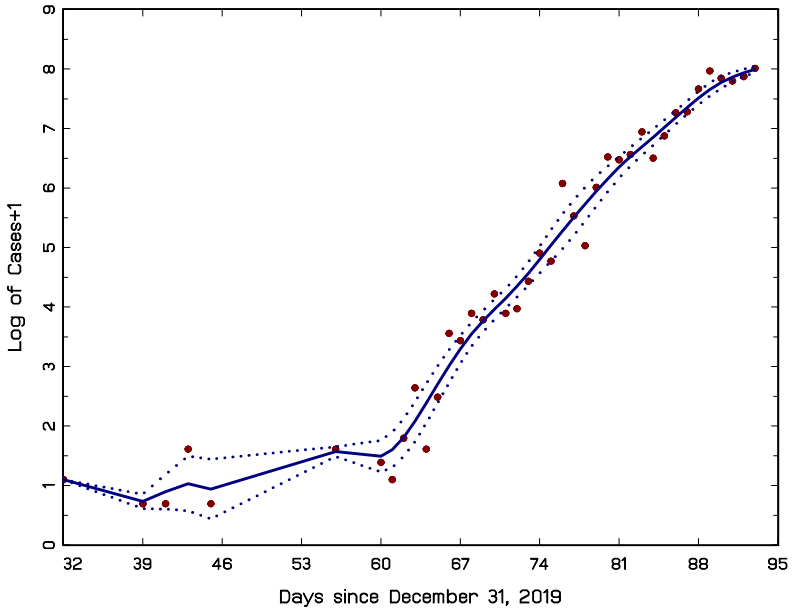


Figure 3. Fitted curve and 95% pointwise confidence band in blue, data points in red.

3 Regression Model

We model the number of new cases and new deaths per day for each country or territory. Suppose that the regression function for the log of cases (and log of deaths) as a function of time can be well approximated by a function m as in (1), where we drop the subscript for convenience. Although the data are counts, i.e., integer values, we do not impose this property; this can matter a bit at the beginning and the end of the trajectory for small countries, but otherwise it doesn't seem to be necessary or helpful to impose some Poisson type process for large countries in the middle of the episode. We focus on the mean regression and do not restrict the error process beyond the zero mean property, it may be autocorrelated and heteroskedastic within some limits. The long run behaviour is fully determined by the trend m. Other models involve dynamic processes for y_t; one problem with this approach is the presence of measurement error in y, which in our framework is simply averaged out, whereas in a dynamic model some additional structure regarding the measurement error would have to be imposed to mitigate its effects on parameter estimation. We discuss both nonparametric and parametric approaches to choosing m. We use natural logarithms and base 10 logarithms interchangeably, where the latter is useful for interpretation of graphs, while the former has a more convenient notation.

3.1 Parametric Forms

Suppose that

$$m(t) = a + bt + ct^2, \tag{2}$$

where a, b, c are parameters to be determined.A quadratic is the simplest function that reflects the possibility of a turning point. A well defined maximum of m occurs if and only if $c < 0$ and occurs at the time $\tau_{\max} = -b/2c$, which results in the maximal value of cases per day of $m(\tau_{\max}) = a - b^2/4c$; finally, the value of t after which no cases would be reported (the end of the epidemic) is the larger root of $m(t) = 0$. A quadratic function can also be rewritten in vertex form so that $m(t) = \alpha + \gamma(t - \mu)^2$, where $\mu = \tau_{\max}$ and $\alpha = m(\tau_{\max})$, and the parameter $\gamma = c$. The vertex form is good for interpretation since it is parameterized in terms of the quantities of interest (but it does require some nonlinear estimation, whereas the parameters of the standard form all have linear parameter effects, which makes it easier to implement). We consider a generalization of the vertex form of the quadratic model

$$m(t) = \alpha + \gamma|t - \mu|^\lambda, \tag{3}$$

where $\alpha, \gamma, \mu, \lambda$ are parameters to be determined. A well defined maximum of m occurs if and only if $\gamma < 0$ and occurs at the time $\tau_{\max} = \mu$,which results in the maximal value of (log) cases per day of $m(\tau_{\max}) = \alpha$. The parameter λ controls the shape of the curve around the peak. When $\lambda > 2$, this is called the *leptobottomed* case, that is, the peak is elongated relative to the $\lambda = 2$ case. When $\lambda = 1$, the *platybottomed* case, the peak is sharper than the $\lambda = 2$ case. The parameter γ measures the speed of approach to the peak and decline from the peak. We look at values $\lambda = 2$ and $\lambda = 4$ primarily. In the case $\lambda = 2$, the equations (2) and (3) are in one to one relation. The case $\lambda = 4$ is a special case of a quartic polynomial. For any λ this model is symmetric about the peak, which enables prediction from before the peak is achieved, although it is restrictive. Only a full dataset including the entire lapse of data would allow discrimination around the symmetry and bottom issue, which is something we plan to do when the full data are in.[3]

We sometimes work with the count data plus one normalized by population $(y_{it} + 1)/n_i$, where n_i is the population of country i and rescaled time. Division by population only affects the constant term, but is done to aid comparability across countries. Rescaling time does not affect fitted values

[3]We have also experimented with the gamma functional form, that is,

$$m(t) = \alpha + \beta \ln(t) + \gamma t, \ t > 0,$$

which is consistent provided $\beta > -1$ and $\gamma < 0$. This functional form can generate asymmetric behavior around the peak located at $-\beta/\gamma$. Empirically, this does not seem to work as well.

and residuals but it does affect raw parameter values and standard errors, which needs to be taken account of.

We take exponentials to convert back to cases, that is, we have

$$E(y_t) = \exp(m(t))E\left[\exp(\varepsilon_t)\right]. \tag{4}$$

When $\lambda = 2$ the model m is quadratic in time and the implied model for $E(y_t)$ is proportional to a Gaussian density as can be seen by completing the square. The model for cumulated counts is then proportional to a Gaussian c.d.f.

3.1.1 Parameters of Interest

The main parameters of interest are the time $\tau_{\max} = \mu$ where the maximum occurs and the maximal value of (log) cases per day of $m(\tau_{\max}) = \alpha$, which are embedded in the vertex form. Transforming back to cases we obtain

$$y_{\max} = \exp(\alpha)E\left[\exp(\varepsilon_t)\right], \tag{5}$$

which is the maximum expected value of y.

Having reached the peak, we may be interested in the expected first passage time to some trough level. For example, suppose that trough is defined as the expected peak number of new cases divided by some number L. Then

$$t_{TP} = \min_{t>\tau_{\max}} \left\{ \frac{y_{\max}}{Ey_t} \geq L \right\}.$$

This is equivalent to finding the first point $t > \tau_{\max}$ for which $(t - \mu) \geq \sqrt{-\log L/\gamma}$.

We are also interested in the total number of cases that would occur, which is $\sum_{t=-\infty}^{\infty} \exp(m(t/K))$. In the case $\lambda = 2$, this is approximately

$$N_{total} = K \times \int_{-\infty}^{\infty} \exp(m(t))dt = K\sqrt{\frac{-\pi}{\gamma}} \exp(\alpha) \tag{6}$$

and in the case $\lambda = 4$

$$N_{total} = K\frac{2\Gamma\left(\frac{5}{4}\right)}{\sqrt[4]{-\gamma}} \exp(\alpha) = \frac{1.812K}{\sqrt[4]{-\gamma}} \exp(\alpha). \tag{7}$$

A bit tongue in cheek, we may define the empirical R_0 as $1 + b + 2ct$ or $1 + 2\gamma(t - \mu)$; this has the property that R_0 starts out above one and declines to one at the peak and then declines further towards zero as the episode ends.

4 Asymptotic Framework

For any epidemic we can think of the first time point $t_{\min}$ and the last time point $t_{\max}$ as finite, so that $y_t = 0$ and $\log(y_t + 1)$ for all $t < t_{\min}$ and $t > t_{\max}$. It follows that there are a finite, albeit

potentially quite large, number of time periods with information about m. We may use the traditional long horizon setting subject to this limitation. Instead, we use an infill asymptotic scheme where we suppose that time is rescaled according to the number of observations being used relative to today. That is, if K is the number of observations being used, time of measurement is t/K, where $t = -K+1, \ldots, 0$, which constitutes a window of length one on the whole horizon. Time is allowed to run from $-\infty$ to $+\infty$ in the scale determined by K. We note that some parameters are relative to the coordinate system $[-1, 0]$ induced by K, where 0 corresponds to today. For example, the parameter $\tau_{\max} = \mu$; to convert back to calendar time (relative to today) we take $\tau_{\max} \to K\tau_{\max}$. As $K \to \infty$ our observations fill up the unit interval. The data generating process is a triangular array indexed by K. If we interpret the model as parametric and appropriate for all data, the model can extrapolate both forward and backward.

Consider the nonparametric point of view. In this case the function $m(.)$ is not specified except that it is a smooth function of (rescaled time). In this case we take the one-sided estimation window of size K to be a small fraction of the total available observations T. In this case, the interpretation is that we are estimating the level of the regression function at the point $u \in [-1, 0]$ using data from previous periods and a local quadratic fit. For the local quadratic (Fan and Gijbels (1996)[9], Gozalo and Linton (2000)**??**), $a = m(u)$, $b = hm'(u)$, and $c = h^2 m''(u)/2$ so that $\gamma = h^2 m''(u)/2$, $\mu = -m'(u)/hm''(u)$, and $\alpha = m(u) - (m'(u)^2/2m''(u))$, where h is a bandwidth parameter $(1/K)$. This suggests that the γ parameter may be hard to estimate since it depends on the second derivative of the regression function and also is scaled by a small quantity. In this case, the model itself has limited extrapolative properties based on the assumption of smoothness.

5 Estimation

Our estimation methodology is very simple. We work with daily data that is available for all countries since December 31st 2019. Each country in the database has a day of first case $t^c_{\min}$ and day of first death $t^d_{\min}$; these vary by country with China having its first recorded cases (as far as this database is concerned) on December 31st, but other countries enter into the fray subsequently. The current time period is denoted 0 and so we have data $\{y^c_{it}, y^d_{it},\ t = -T+1, \ldots, 0,\ i = 1, \ldots, n\}$, which are count data with many zeros at the beginning.

1. We fix λ and estimate the parameters α, γ for given μ by OLS of the log of counts plus one $(\log(y+1))$ for each country using the estimation window data, which contains the *most recent* K *datapoints* with rescaled time (t/K). We then search for the minimum squared error across different μ. We allow country specific case and country specific death values for all parameters.

We provide standard errors for the parameter estimates based on NLLS theory, which is detailed in the appendix. We use the LS standard errors for simplicity, given the small sample size we have for each country the HAC-based standard errors are greatly inflated. We comment on the residual properties below.

2. We then extrapolate the estimated regression curves outside the estimation window and take exponentials to deliver predictions of the number of new cases and new deaths per day. That is, we calculate $\widehat{m}(t)$ and $\widehat{E}y_t = \exp(\widehat{m}(t))\widehat{\kappa}_0$ for all t from $-\infty$ to $+\infty$, where $\widehat{\kappa}_0 = \sum_{i-K}^{1} \exp(\widehat{\varepsilon}_i)/K$ and $\widehat{\varepsilon}_i$ are the residuals.

3. We provide FREQUENTIST prediction intervals for Ey_t

$$\mathcal{I}(t) = \left[\exp(\widehat{m}(t))\widehat{\kappa}_0 \exp\left(\widehat{q}_{\alpha/2}\right), \exp(\widehat{m}(t))\widehat{\kappa}_0 \exp\left(\widehat{q}_{1-\alpha/2}\right)\right], \tag{8}$$

where $\widehat{q}_\alpha$ is the α-quantile of the residuals.

4. We provide a forecast of the total number of cases: the total number of cases so far plus $\sum_{t>today} \exp(\widehat{m}(t))$. Alternatively, can use the formula for the total number of cases from the model m, which is approximately given by (6) and (7)

5. Selection of K. Our choice of K is often limited by data availability. We investigate in Section 9 a specific algorithm for choosing K.

Regarding the prediction intervals, a full disclosure. These are intended to assist the reader in gauging the fundamental uncertainty that would be present were the parameter values known and the model were true. The reality is that neither of these conditions are met. If one takes account of parameter uncertainty, then the prediction intervals expand rapidly with horizon. Technically, the parameter uncertainty overwhelms the prediction uncertainty when the horizon is greater than the square root of estimation sample size. We would argue that this is true of any model in this setting unless one believes in Bayesian magic. We discuss short term prediction intervals that take account of parameter uncertainty in more detail in the appendix.

6 Data

We consider the 191+ countries and entities in the ECDC dataset but report separately only the thirty countries with the largest number of cases (excluding China) and with at least K days of data. For the "World" calculation we use all data from all entities that contain a population number. Specifically, cases are the reported daily count of detected and laboratory (and sometimes, depending on the

country reporting them and the criteria adopted at the time, also clinically) confirmed positive and sometimes - depending on the country reporting standards - also presumptive, suspect, or probable cases of detected infection. The size of the gap between detected (whether confirmed, suspect or probable) and reported cases versus actual cases will depend on the number of tests performed and on the country's transparency in reporting. Most estimates have put the number of undetected cases at several multiples of detected cases. There are a number of reporting issues. Some of these include official governmental channels changing or retracting figures, or publishing contradictory data on different official outlets. National or State figures with old or incomplete data compared to regional, local (counties, in the US) government's reports is the norm.

7 Results

We consider the thirty countries with the largest number of cases (excluding China) and with at least K days of data. We re-estimate every day as new data comes in. We fully expect the parameters to change over time and to vary across country, and they do.

7.1 Estimation of Vertex Models

We fit the models (3) on each country's case and death data separately with the *most recent* K *datapoints* (and using rescaled time with estimation window $[-1, 0]$) and report the most recent results below in Tables 1a (cases, $\lambda = 2$), Tables 1b (cases, $\lambda = 4$), Table 2a (fatalities, $\lambda = 2$) and Tables 2b (fatalities, $\lambda = 4$). We use $K = 60$ here. The regressions generally have high R^2 (low R^2 may partly reflects data quality and country size and phase of the cycle). The α and μ parameters are all significantly different from zero. We report $y_{\max}$ in the table below, which is a transformation of α. There is quite a bit of heterogeneity across the parameters consistent with different countries being at different stages of the cycle and having taken different approaches to managing the epidemic and having different populations. By now most countries have μ significantly less than zero, which means their peak has passed, but a number of countries still have yet to reach their peak such as Brazil, Peru, and Saudi Arabia. The γ parameters are all negative but the standard errors are quite wide, and in some cases the confidence intervals around these estimates includes zero. The $y_{\max}$ parameter is quite well estimated for most countries with some exceptions such as Pakistan.

We provide likelihood ratio test statistics ℓ that can be used to test the difference between $\lambda = 2$ and $\lambda = 4$; the critical value for the likelihood ratio test $\ell_2 \pm \ell_4$ is $\chi^2_{0.95} = 3.84$. In the past there was not much to choose between the two models, for most countries. Currently, the US, UK, and Spain strongly prefers the fat-bottomed model.

Country	ymax	se(ymax)	γ	se(γ)	μ	se(μ)	ℓ	Total Cases
USA	38809	8580	-7.7256	2.9834	-0.3871	0.1148	58.0340	1390746
Russia	10353	2632	-8.5914	1.6030	-0.0912	0.0829	28.8474	242271
Spain	5345	1176	-4.9125	2.2070	-0.7344	0.1612	50.7048	230117
UK	6340	1207	-7.3310	2.3858	-0.3621	0.1012	74.1673	229705
Italy	4819	644	-2.7050	1.0933	-0.8225	0.1610	104.2719	222104
Brazil	9475	3772	-3.6545	1.0545	0.1429	0.1891	63.6008	188974
Germany	3652	1261	-4.9252	3.6748	-0.7065	0.2580	-1.4682	172239
Turkey	5137	1234	-13.4068	3.4786	-0.4080	0.0740	49.3770	143114
France	3035	789	-5.7580	3.2089	-0.6435	0.1750	36.5847	140734
Iran	1438	541	-0.1279	1.8171	-0.9810	7.0726	-1.4086	112725
India	3036	602	-6.4238	1.1918	-0.0775	0.0840	60.4317	78003
Peru	5385	5530	-3.5841	1.7864	0.2843	0.3960	0.3417	76306
Canada	1996	403	-7.2755	2.4623	-0.3519	0.1071	66.7055	72278
Belgium	1367	319	-7.7612	4.2842	-0.5129	0.1406	60.6694	53981
Saudi Arabia	2683	2133	-3.5194	1.6912	0.2129	0.3480	6.9111	44830
Netherlands	1059	210	-6.7025	3.1790	-0.5640	0.1285	75.0737	43211
Mexico	1836	475	-4.0458	0.8554	0.0785	0.1252	88.7098	40186
Ecuador	336	146	-3.6784	4.3497	-0.2655	0.4243	-30.7739	38468
Pakistan	167065	2912100	-0.8290	2.2408	2.2976	7.5697	-26.8519	35788
Chile	2002	2006	-1.6285	1.1267	0.4657	0.6739	55.6490	34381
Switzerland	673	218	-7.5093	3.5314	-0.6951	0.1601	7.2165	30330
Portugal	641	226	-7.1334	6.0097	-0.5448	0.2214	8.4495	28454
Sweden	632	123	-4.4370	2.2411	-0.3289	0.1659	69.5536	27909
Qatar	1786	1423	-3.7725	1.6019	0.2322	0.3156	13.4277	26539
Belarus	1419	1122	-5.7032	2.2172	0.1258	0.2482	-25.5806	25825
Singapore	1474	2330	-2.4358	2.0496	0.4020	0.7664	-16.1524	25346
Ireland	634	135	-9.0122	3.7264	-0.4630	0.1075	69.4797	23401
United Arab Emirates	640	248	-7.3384	2.6818	-0.1196	0.1563	-23.8342	20386
Bangladesh	3309	4466	-4.2605	1.6764	0.4206	0.3657	7.9669	17822
Poland	419	80	-5.1950	2.3313	-0.3518	0.1421	73.2037	17204

Table 1a. Estimation using daily cases data upto 20200514. $\log(newcases + 1) = \alpha + \gamma|t - \mu|^2 + err$

Country	ymax	se(ymax)	γ	se(γ)	μ	se(μ)	ℓ	Total Cases
USA	30906	2706	-29.7090	15.0248	-0.4001	0.0691	99.6567	1390746
Russia	9643	2580	-2.2186	1.0341	0.3632	0.1558	26.2471	242271
Spain	4379	612	-4.9067	5.1871	-0.8672	0.2109	39.9047	230117
UK	4900	426	-21.4541	11.9822	-0.3567	0.0821	96.5777	229705
Italy	4191	404	-2.9465	2.8118	-0.9112	0.2009	88.2666	222104
Brazil	8455	2679	-0.5834	0.3753	0.7143	0.2829	62.5448	188974
Germany	2950	592	-4.4908	8.0594	-0.8532	0.3518	-4.6171	172239
Turkey	3351	366	-49.6875	20.0154	-0.4127	0.0538	74.1200	143114
France	2443	362	-5.0474	7.1406	-0.8217	0.2664	29.5057	140734
Iran	1396	304	-0.0543	4.2156	-0.9905	17.9292	-1.4413	112725
India	2897	587	-1.6195	0.7254	0.3836	0.1524	62.9319	78003
Peru	4280	2991	-0.4407	0.4561	0.9265	0.5236	-0.6939	76306
Canada	1534	154	-19.3063	12.5179	-0.3383	0.0980	77.8218	72278
Belgium	1093	149	-34.5242	41.8431	-0.4958	0.1402	55.1727	53981
Saudi Arabia	2272	1319	-0.4909	0.5070	0.8194	0.4879	5.9026	44830
Netherlands	856	108	-29.0480	33.4403	-0.5323	0.1375	61.1348	43211
Mexico	1678	398	-0.7303	0.3783	0.6178	0.2125	82.1965	40186
Ecuador	290	73	-16.1789	34.8028	-0.3584	0.3153	-30.1975	38468
Pakistan	-	-	-0.0131	0.0569	3.9461	6.5018	-26.8825	35788
Chile	1671	984	-0.1511	0.1980	1.1983	0.7759	56.5933	34381
Switzerland	506	96	-7.1573	7.8718	-0.8475	0.2140	1.6660	30330
Portugal	513	105	-31.2333	62.5322	-0.5069	0.2318	6.1666	28454
Sweden	550	56	-5.7980	5.8270	-0.2083	0.1819	73.7050	27909
Qatar	1464	841	-0.5071	0.4631	0.8483	0.4396	11.3978	26539
Belarus	1179	760	-0.9318	0.8253	0.6887	0.3826	-26.7177	25825
Singapore	1157	1131	-0.2463	0.4101	1.1032	0.9351	-16.6248	25346
Ireland	479	63	-37.9699	35.7048	-0.4709	0.1117	56.5365	23401
United Arab Emirates	603	232	-2.0164	1.7528	0.3205	0.2797	-24.2894	20386
Bangladesh	2111	1780	-0.4186	0.3314	1.1310	0.4517	4.7352	17822
Poland	352	31	-13.0042	10.1663	-0.3246	0.1205	92.5052	17204

Table 1b. Estimation using daily cases data upto 20200514. $\log(newcases + 1) = \alpha + \gamma|t - \mu|^4 + err$

Country	ymax	se(ymax)	γ	se(γ)	μ	se(μ)	ℓ	Total Deaths
USA	2984	841	-12.9536	3.4033	-0.3486	0.0836	26.5783	84133
UK	1053	243	-11.3028	3.1697	-0.3928	0.0824	53.1144	33186
Italy	620	129	-2.0896	2.0947	-0.7342	0.3595	57.0413	33106
Spain	642	237	-8.2591	6.6002	-0.5288	0.2059	4.2269	27104
France	772	204	-11.1299	4.7868	-0.4771	0.1102	44.6179	27074
Brazil	523	103	-6.6682	1.1347	-0.0654	0.0784	63.0187	13149
Belgium	315	84	-14.6624	4.0282	-0.4197	0.0767	37.9850	8843
Germany	239	84	-13.0416	5.2702	-0.4177	0.1133	4.7062	7723
Iran	139	108	-0.6435	1.4373	-1.2602	1.7215	26.4329	6783
Netherlands	153	39	-8.9751	4.6136	-0.4731	0.1322	47.7575	5562
Canada	176	37	-7.8702	1.7883	-0.1911	0.0891	50.2206	5304
Mexico	208	71	-5.3687	1.2092	0.0612	0.1296	47.1672	4220
Turkey	140	28	-12.3428	2.5852	-0.3771	0.0634	70.5343	3952
Sweden	81	27	-8.8844	3.9076	-0.3428	0.1413	7.7341	3460
India	112	32	-4.8348	1.1453	0.0256	0.1281	53.6900	2549
Ecuador	31	11	-4.1746	2.6258	-0.1393	0.2624	-14.4707	2334
Russia	137	60	-4.3327	1.1027	0.1599	0.1710	58.2364	2212
Peru	128	74	-3.6467	1.3049	0.1950	0.2528	38.0287	2169
Switzerland	37	13	-8.8414	6.3073	-0.5039	0.1812	15.3460	1563
Ireland	42	12	-11.4195	3.6630	-0.3769	0.0971	28.6433	1497
Portugal	34	10	-10.0752	5.1152	-0.4481	0.1349	24.2923	1175
Indonesia	25	7	-6.1977	3.5437	-0.3684	0.1758	29.4365	1028
Romania	30	7	-7.7107	2.5310	-0.3088	0.1111	48.3342	1016
Poland	25	6	-7.5273	2.5797	-0.3066	0.1164	45.4032	861
Philippines	17	5	-4.1016	2.9907	-0.2789	0.2571	18.4014	772
Pakistan	91	185	-1.5361	1.3746	0.7349	1.1108	31.7891	770
Japan	-	-	-0.1097	1.5373	11.7729	171.9905	18.3608	687
Austria	18	6	-10.3478	5.7146	-0.5041	0.1402	27.1560	624
Egypt	13	4	-3.9926	2.5025	-0.2271	0.2356	22.5528	556
Denmark	15	4	-6.9342	4.4063	-0.4524	0.1677	44.4319	545

Table 2a. Estimation using daily fatality data upto 20200514.

$$\log(newdeaths + 1) = \alpha + \gamma|t - \mu|^2 + err$$

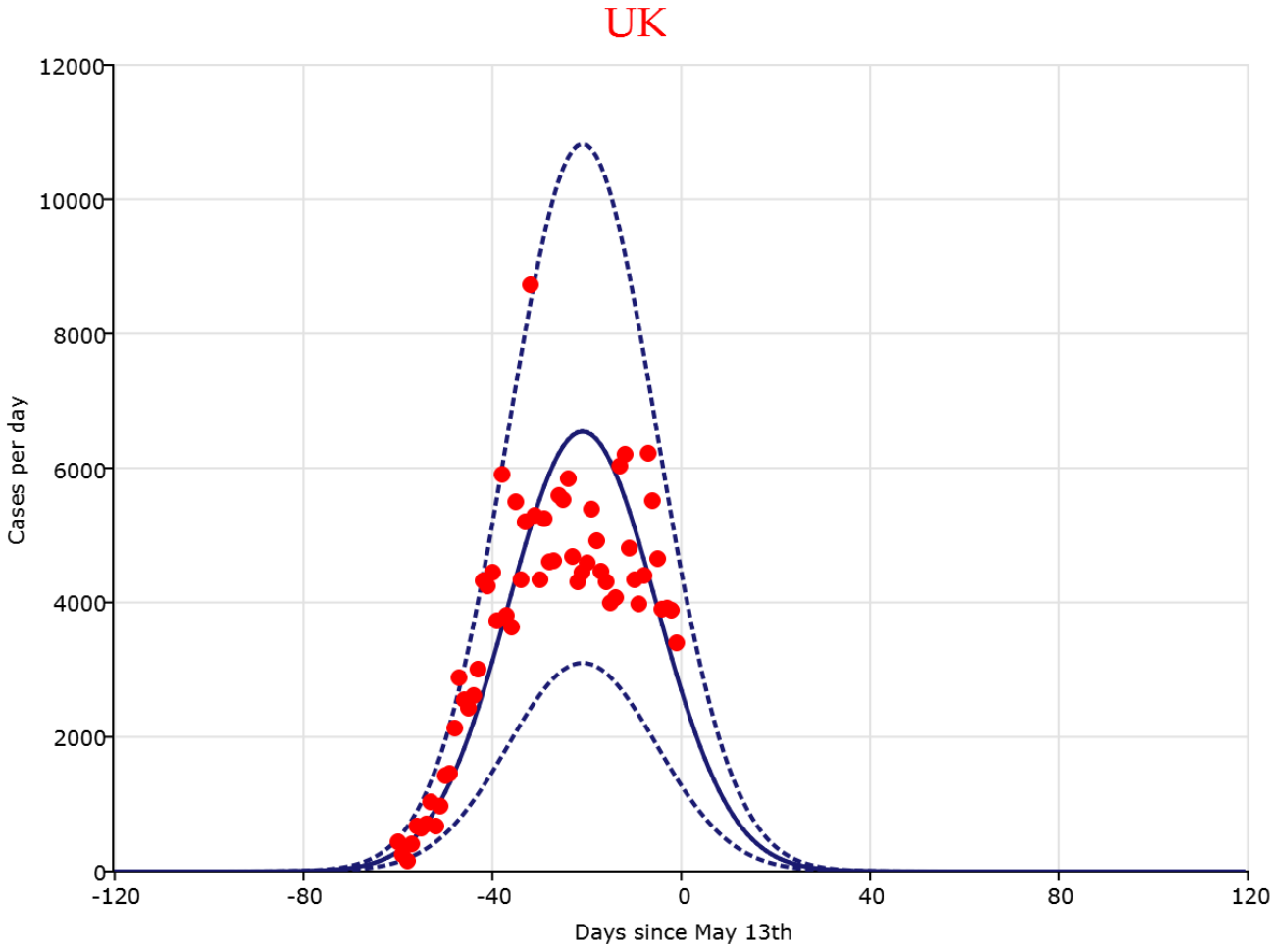
Country	ymax	se(ymax)	γ	se(γ)	μ	se(μ)	ℓ	Total Deaths
USA	1922	273	-20.7378	11.0254	-0.2580	0.0900	32.3249	84133
UK	706	88	-35.6409	20.1017	-0.3875	0.0788	56.0176	33186
Italy	562	71	-1.9356	4.6870	-0.8671	0.4830	52.1219	33106
Spain	514	107	-38.3333	63.0842	-0.4902	0.1908	3.8790	27104
France	557	83	-47.5986	40.3728	-0.4705	0.1008	41.4068	27074
Brazil	497	99	-1.6405	0.6716	0.4012	0.1414	67.2217	13149
Belgium	198	28	-56.2363	27.4759	-0.4251	0.0636	45.4035	8843
Germany	157	30	-48.0119	35.8133	-0.4173	0.0987	7.6423	7723
Iran	121	33	-0.2744	0.9690	-1.3902	1.2081	26.3058	6783
Netherlands	121	16	-40.4200	36.2020	-0.4689	0.1067	52.8471	5562
Canada	164	31	-2.5417	1.3277	0.2134	0.1521	44.3345	5304
Mexico	186	59	-1.0013	0.5525	0.5918	0.2220	42.6084	4220
Turkey	89	9	-35.5248	15.3888	-0.3671	0.0626	73.9161	3952
Sweden	62	11	-8.8200	8.1674	-0.1700	0.1762	8.1318	3460
India	103	29	-0.9708	0.5744	0.5384	0.2285	50.4410	2549
Ecuador	28	7	-2.3455	3.0417	0.1051	0.3390	-13.8379	2334
Russia	117	43	-0.6646	0.4002	0.7400	0.2696	49.6604	2212
Peru	109	48	-0.5265	0.4137	0.7918	0.3645	35.5024	2169
Switzerland	29	6	-40.9506	57.2700	-0.4906	0.1621	15.6923	1563
Ireland	29	5	-10.2698	7.9005	-0.1849	0.1436	21.6467	1497
Portugal	25	4	-40.1872	36.5908	-0.4429	0.1143	25.9474	1175
Indonesia	20	3	-16.2756	18.9511	-0.3441	0.1744	32.3387	1028
Romania	24	3	-7.6967	4.8800	-0.1440	0.1247	53.9105	1016
Poland	22	3	-3.2929	2.4639	0.0334	0.1812	42.6514	861
Philippines	16	3	-2.0017	3.0042	0.0341	0.3637	18.5626	772
Pakistan	64	69	-0.0975	0.1587	1.6022	1.1744	31.1312	770
Japan	-	-	-0.0001	0.0022	18.1606	132.7044	18.3540	687
Austria	13	2	-45.7686	53.7189	-0.4922	0.1359	24.0641	624
Egypt	13	3	-1.4218	1.8753	0.1593	0.3643	22.7437	556
Denmark	12	2	-29.4153	32.6466	-0.4512	0.1369	47.3756	545

Table 2b. Estimation using daily fatality data upto 20200514.

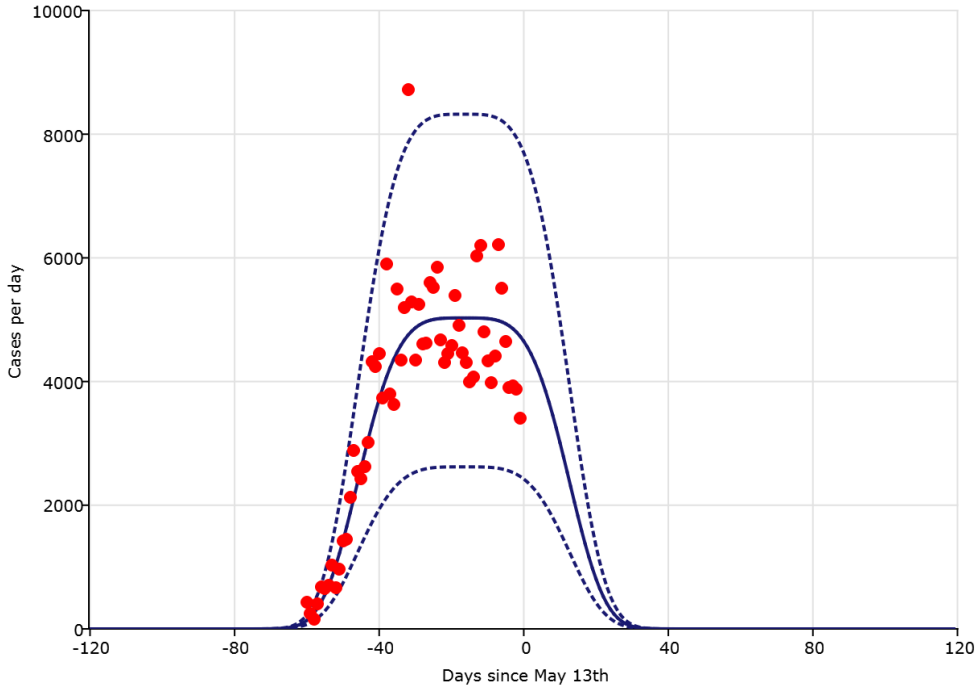
$$\log(newdeaths + 1) = \alpha + \gamma|t - \mu|^4 + err$$

7.2 Prediction of the Future

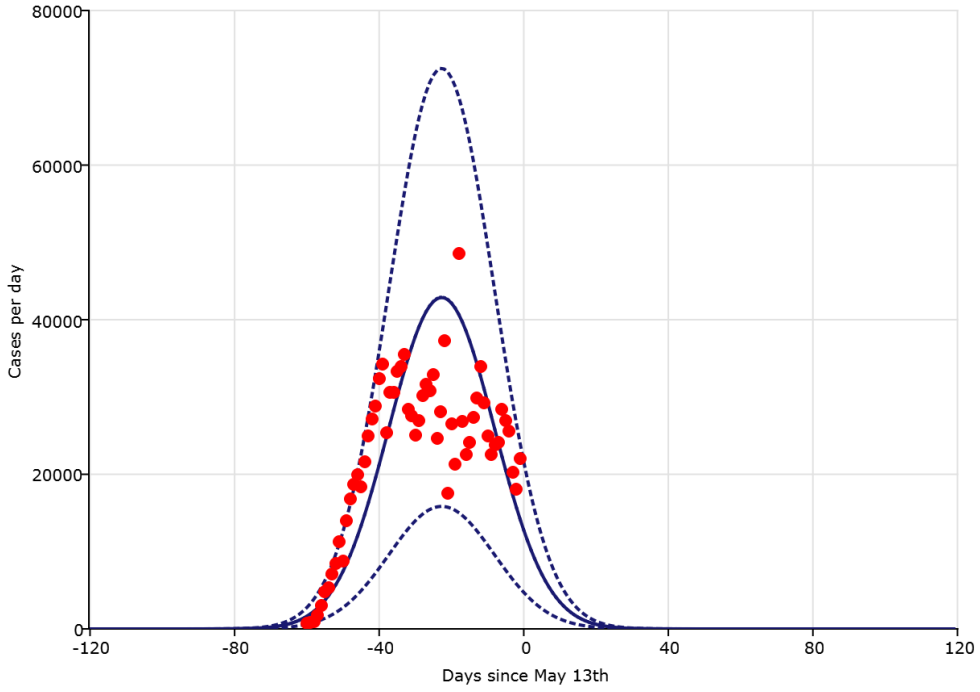
We present the estimated curves along with their 95% prediction intervals for selected countries in the graphs below. The estimated curve in solid blue, confidence intervals in dotted blue, and data points in red. The extrapolation curve is a scaled Gaussian density function as mandated by our curve model.

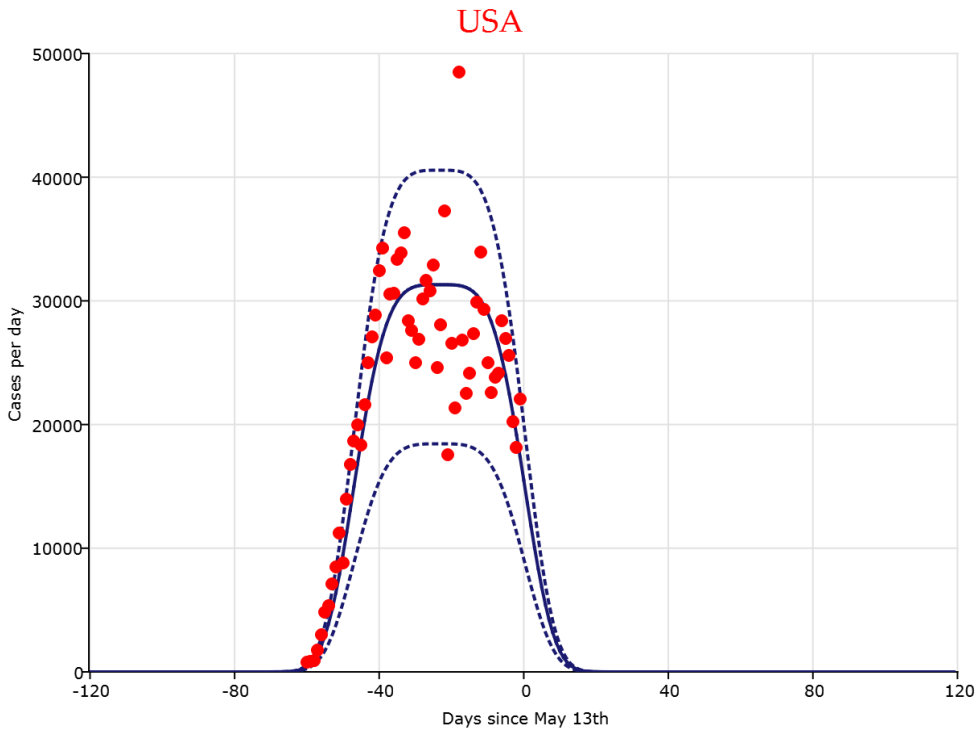


UK


10000
8000
6000
4000
2000
0
Cases per day
-120
-80
-40
0
40
80
120
Days since May 13th

USA

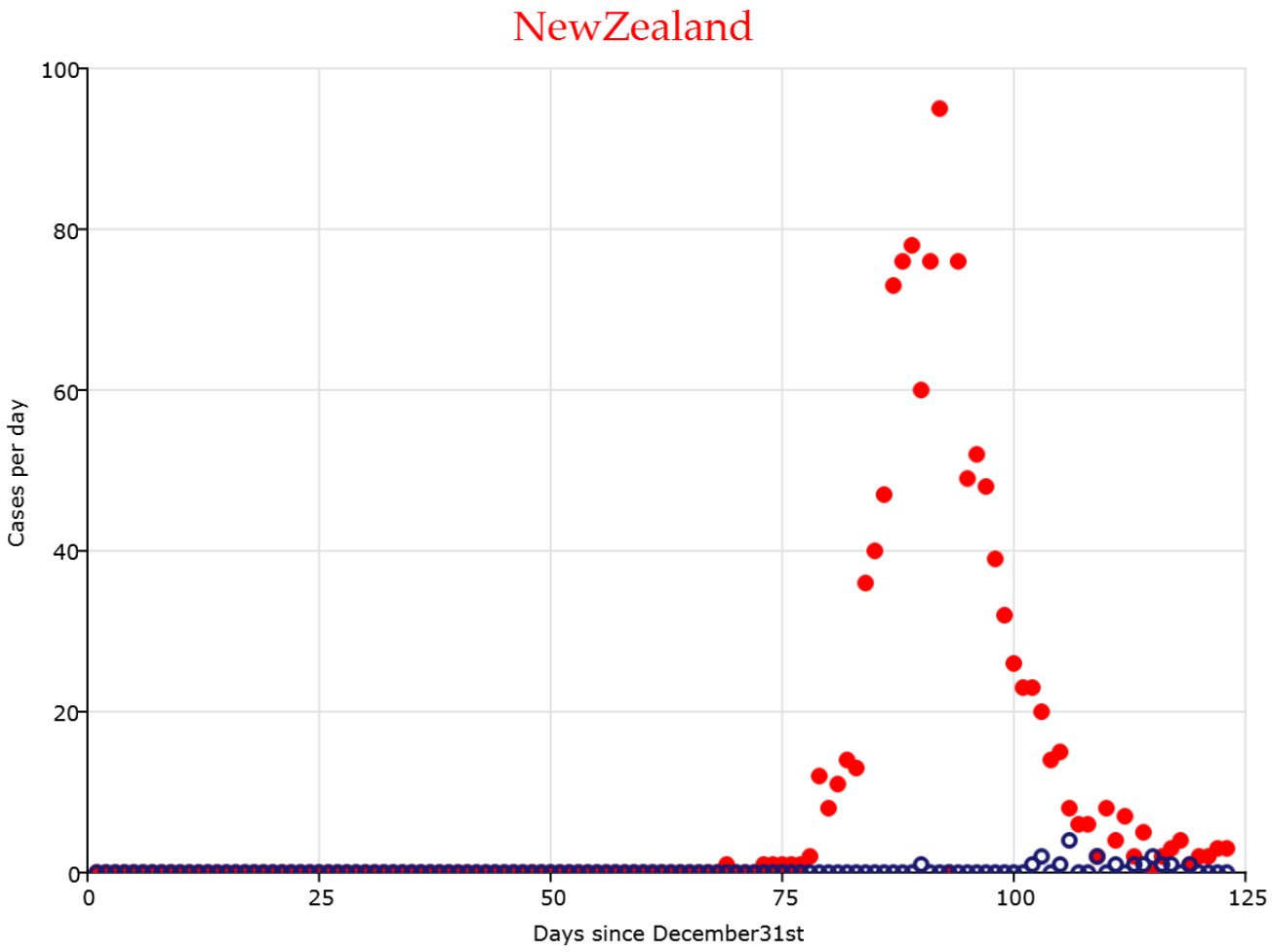

80000
60000
40000
20000
0
Cases per day
-120
-80
-40
0
40
80
120
Days since May 13th



We try to estimate the time that it has taken for selected countries to pass from peak to trough, where trough is defined as the estimated peak number of new cases divided by 10. Admittedly, this is not a very exacting standard, but the advantage is that there are several countries that have effectively satisfied this. An empirical estimator is just based on the first passage time

$$\widehat{t}_{TP} = \min_{t>\tau_{\max}} \left\{ \log_{10}(y_{\tau_{\max}}) - \log_{10}(y_t) \geq 1 \right\},$$

where $\log_{10}(10) \simeq 1$ and $\tau_{\max} = \arg\max_s \log_{10}(y_s)$. For New Zealand, Australia, and Austria, $\widehat{t}_{TP} = 14$, for South Korea $\widehat{t}_{TP} = 15$, for China $\widehat{t}_{TP} = 19$, for Norway $\widehat{t}_{TP} = 21$, for Israel $\widehat{t}_{TP} = 21$, while for Switzerland $\widehat{t}_{TP} = 31$. These estimates are rather rough and in each case there was some bounce back. Incidentally, looking at these successful countries' trajectories, there seems to be a fairly symmetric curve around the peak, see below the case of New Zealand.



We can also estimate the passage time from the model for countries that have not yet passed the threshold by computing

$$\min\left\{t : m(\tau_{\max}) - m(\tau_{\max} + t) \geq 1\right\},$$

This is equivalent to $(t - \mu) \geq \sqrt{-1/\gamma}$. For the UK, this gives around 28 days based on the latest estimates, although the confidence interval is rather wide.

8 Model Tests and Robustness

8.1 Quantile Estimation

There are some concerns about the data quality regarding both cases and deaths. In some cases there is stale reporting where countries report deaths essentially on a less than daily basis and so one observes large values of the dependent variable occasionally. The UK revised its fatality figures

on 29th April to include care home deaths, which led to a jump of 4000 or more additional deaths on that day. The deaths have now been distributed temporally roughly according to when they occurred, but apparently they have not yet revised the number of cases accordingly (they have not even added the number of care home deaths as additional cases). This made us concerned about robustness issues. For this reason we consider quantile regression methods to fit the trend curves in this section.

We suppose that

$$\log y_t = m(t) + \varepsilon_t, \quad m(t) = \alpha + \gamma|t - \mu|^\lambda, \tag{9}$$

where ε_t is median zero. In this case we have

$$\text{med}(y_t) = \exp(m(t))$$

because of the equivariance of quantiles to any monotonic transformation. The results do not change very much in most cases.

8.2 One Step Ahead Forecasting

In this section we report the results of a one step ahead prediction exercise using yesterday's data to predict today's (20200513) outcome ($K = 40$ and $\lambda = 2$). We here take account of both estimation error and prediction error, which widens the bands shown in the graphs. We report the t-statistics (computed on the log scale) for one-step ahead prediction and a joint test based on Pesaran and Yamagata (2012)[22]. The results are shown below in two tables. Spain reported significantly more new cases than predicted yesterday, while Brazil and Chile reported significantly less than predicted; otherwise the one step ahead predictions were individually and collectively within the model bounds. France and Turkey reported significantly more deaths than predicted, while Netherlands reported significantly less deaths than predicted; otherwise the one-step ahead predictions were individually and collectively within the model bounds. The UK reported less cases and less deaths than predicted but both within model bounds.

Country	Predicted	Actual	t-stat	Total Cases	Population
USA	24189	18118	1.4639	1347916	327167434
Spain	713	3047	-4.2005	230296	46723749
UK	4324	3878	0.4979	223060	66488991
Russia	10584	11657	-0.6578	221344	144478050
Italy	928	745	1.3317	219814	60431283
Germany	605	934	-0.5050	170508	82927922
Brazil	11654	5633	2.5836	168331	209469333
Turkey	1177	1115	0.2088	139771	82319724
France	532	457	-0.0325	139519	66987244
Iran	1405	1684	-0.3336	109286	81800269
India	4125	3605	0.5800	70756	1352617328
Canada	1376	1134	0.9904	69981	37058856
Peru	4968	1516	0.7345	68822	31989256
Belgium	331	369	-0.4372	53449	11422068
Netherlands	202	162	0.6110	42788	17231017
Saudi Arabia	1614	1967	-0.4029	41014	33699947
Ecuador	2185	51	1.2653	37491	17084357
Mexico	1913	1306	1.9296	36327	126190788
Pakistan	2427	1141	0.1368	32081	212215030
Switzerland	45	40	0.2862	30261	8516543
Chile	1984	1198	2.0950	30063	18729160
Portugal	162	99	0.2495	28001	10281762
Sweden	535	349	1.2507	26670	10183175
Belarus	1525	934	0.0424	23906	9485386
Singapore	924	452	0.2440	23787	5638676
Qatar	1052	1104	-0.2628	23623	2781677
Ireland	124	140	-0.4686	23135	4853506
United Arab Emirates	676	681	-0.3121	18878	9630959
Israel	22	30	-0.7426	16506	8883800
Poland	303	331	-0.4841	16326	37978548

Table 3. One step ahead prediction using cases data. $\tau_{PY} = 1.17$

Country	Predicted	Actual	t-stat	Total Deaths	Population
USA	1177	1157	-0.0884	80684	327167434
UK	352	211	1.4294	32065	66488991
Italy	190	180	0.2002	30739	60431283
Spain	173	124	0.1852	26744	46723749
France	106	264	-2.3430	26643	66987244
Brazil	705	397	1.6085	11519	209469333
Belgium	57	52	0.0461	8707	11422068
Germany	45	117	-1.1654	7533	82927922
Iran	51	46	-0.0400	6685	81800269
Netherlands	41	17	2.0272	5456	17231017
Canada	143	123	0.5059	4993	37058856
Turkey	40	56	-4.0450	3841	82319724
Mexico	191	109	1.1784	3573	126190788
Sweden	48	32	0.1231	3256	10183175
India	127	88	0.8018	2293	1352617328
Ecuador	116	19	0.6989	2145	17084357
Russia	95	95	-0.1228	2009	144478050
Peru	130	73	0.7571	1961	31989256
Switzerland	7	6	0.0539	1542	8516543
Ireland	14	10	0.2998	1467	4853506
Portugal	7	10	-0.5622	1144	10281762
Indonesia	14	19	-0.6933	991	267663435
Romania	25	21	0.3088	972	19473936
Poland	14	12	0.3495	811	37978548
Philippines	16	8	0.7428	726	106651922
Pakistan	36	40	-0.3429	706	212215030
Japan	19	23	-0.5140	643	126529100
Austria	2	3	-0.4816	620	8847037
Denmark	6	5	0.6644	533	5797446
Egypt	15	9	0.4671	533	98423595

Table 4. One step ahead prediction using deaths data. $\tau_{PY} = 0.63$

8.2.1 Choice of K

This section illustrates the selection of the window width K. Let today is the day 0, K ranges from 14 to 60 and L is the length of the test data. We choose K=60 here because the date of WHO announcing COVID-19 as pandemic is 2020-03-12, approximately 60 days ago. We compute the one step ahead forecast $\hat{m}(1-l;K)$ of $\log y_{1-l}$ based on the data from $\{-l,-l-1,\ldots,-l-K\}$ for $l=1,2,\ldots,L$ and calculate:

$$\hat{Q}(K;L)=\frac{1}{L}\sum_{1}^{L}|(\log y_{i-l})-\hat{m}(1-l;K)|.$$

We show the empirical results for both daily infections and fatalities below. The choice $K=40$ works quite well.

Table 1: One Step Ahead Prediction Loss of Cases

	Name	$\hat{Q}(K;L)$	Minimum Loss	Loss K=50	Loss K=40
1	USA	5148.37	2330	9237	3310
2	Spain	362.83	271	474	292
3	UK	1227.93	847	1717	909
4	Russia	969.67	235	268	830
5	Italy	194.22	145	163	151
6	Germany	611.83	490	583	515
7	Brazil	2408.65	1783	2904	2162
8	France	1017.48	836	838	945
9	Iran	277.61	161	334	162
10	India	637.96	318	838	812
11	Canada	191.11	78	222	119
12	Peru	1583.76	548	2049	1088
13	Belgium	240.11	174	273	226
14	Netherlands	94.24	67	112	86
15	Saudi_Arabia	548.33	79	863	526
16	Ecuador	1397.80	786	787	937
17	Mexico	266.07	174	257	174
18	Pakistan	996.91	237	1382	753
19	Switzerland	17.61	14	28	14
20	Chile	372.83	160	343	166
21	Portugal	282.09	261	283	262
22	Sweden	195.28	173	226	193
23	Belarus	1216.78	22	1498	82
24	Singapore	449.00	41	815	271
25	Qatar	421.96	126	269	168
26	Ireland	61.70	56	59	61
27	United Arab Emirates	304.61	86	423	115
28	Israel	19.93	19	20	19
29	Poland	54.04	17	90	48
30	Japan	44.04	16	41	29

This table summarises the average loss of one step prediction for new daily cases when L=5. And K ranges from 14 to 60 days. The third column presents the minimum average loss of 5 days within that range. And the corresponding results of K=50 and K=40 are provided for comparisions.

Table 2: One Step Ahead Prediction Loss of Deaths

	Name	$\hat{Q}(K;L)$	Minimum Error	Error K=50	Error K=40
1	USA	664.20	525	831	655
2	UK	178.04	145	211	166
3	Italy	62.54	53	61	54
4	Spain	95.59	39	111	59
5	France	73.48	66	79	71
6	Brazil	160.54	98	211	163
7	Belgium	74.02	66	84	71
8	Germany	56.61	49	56	52
9	Iran	10.37	8	12	11
10	Netherlands	25.61	24	27	24
11	Canada	28.00	18	21	36
12	Mexico	58.52	45	46	55
13	Sweden	52.37	47	49	50
14	India	25.78	11	13	12
15	Ecuador	90.74	85	87	85
16	Russia	22.74	7	31	13
17	Peru	27.09	8	28	27
18	Switzerland	4.78	4	4	5
19	Ireland	9.70	6	10	9
20	Portugal	6.28	5	8	6
21	Indonesia	8.67	7	8	8
22	Romania	5.96	4	9	5
23	Poland	4.96	4	5	6
24	Philippines	7.30	7	7	7
25	Pakistan	13.78	7	13	12
26	Japan	20.52	14	25	21
27	Austria	1.87	1	1	2
28	Egypt	5.02	4	4	4
29	Denmark	2.61	2	2	3
30	Algeria	2.93	1	4	4

This table summarises the average loss of one step prediction for new daily deaths when L=5. And K ranges from 14 to 60 days. The third column presents the minimum average loss of 5 days within that range. And the corresponding results of K=50 and K=40 are provided for comparisions.

8.3 Residual Properties

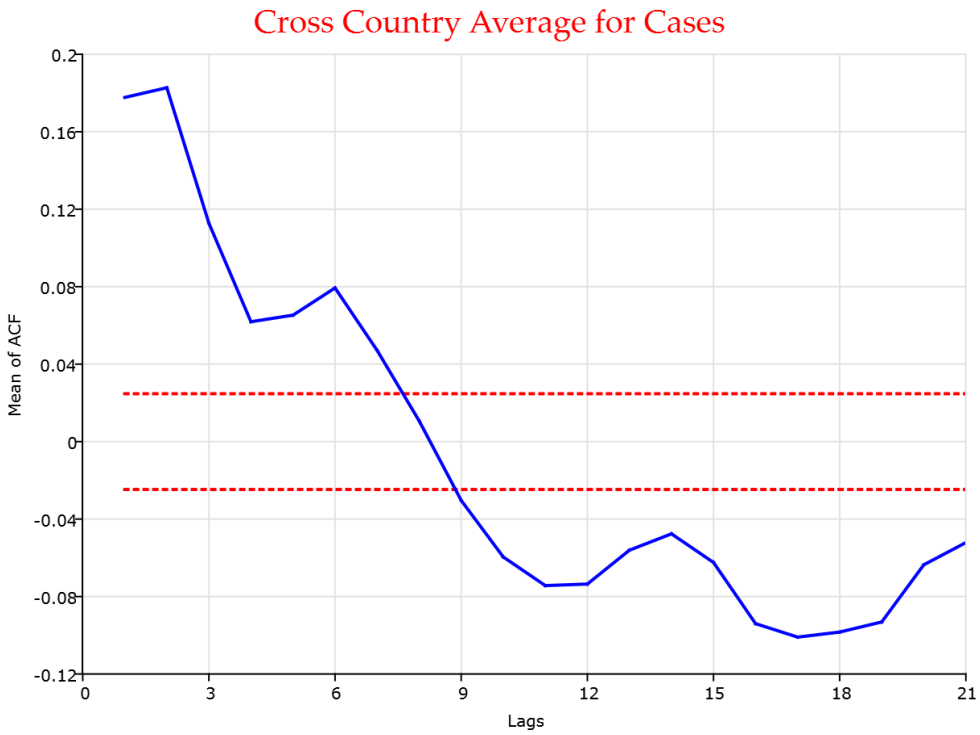
In this section we look at the properties of the residuals from the trend fitting. The model assumptions do not exclude autocorrelation or heteroskedasticity but there is limited scope to improve efficiency by exploiting these properties. They would however affect standard errors and might suggest alternative short term predictors.

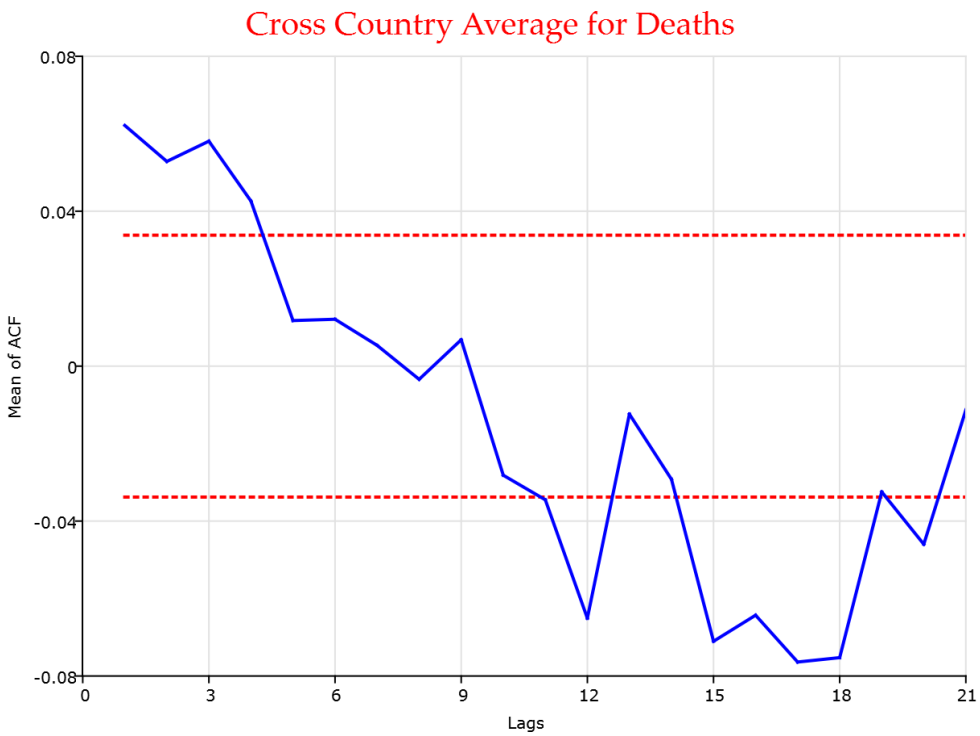
8.3.1 Seasonal Properties

We look at whether the daily data have day of the week effects, which are frequently cited in the media. We consider two models: (1) six daily dummy variables or (2) include the two periodic functions $\sin(\pi t/7), \cos(\pi t/7)$. We find limited evidence for seasonal effects in some countries' data for some choice of estimation windows, but mostly it is not statistically significant. We calculate the classical F-test for the exclusion of variables for the cases and deaths response for each country. The results do depend a bit on the choice of estimation window, but generally there is little evidence for day of the week effect once trend is allowed for. For the UK with the 41 most recent observations, the t-statistics are not individually significant. Furthermore, $\overline{R}^2 = 0.559$ for both cases and deaths quadratic regression with the two seasonal effects, while without the seasonal effects $\overline{R}^2 = 0.560$ for cases and $\overline{R}^2 = 0.567$ for deaths.

8.3.2 Autocorrelation

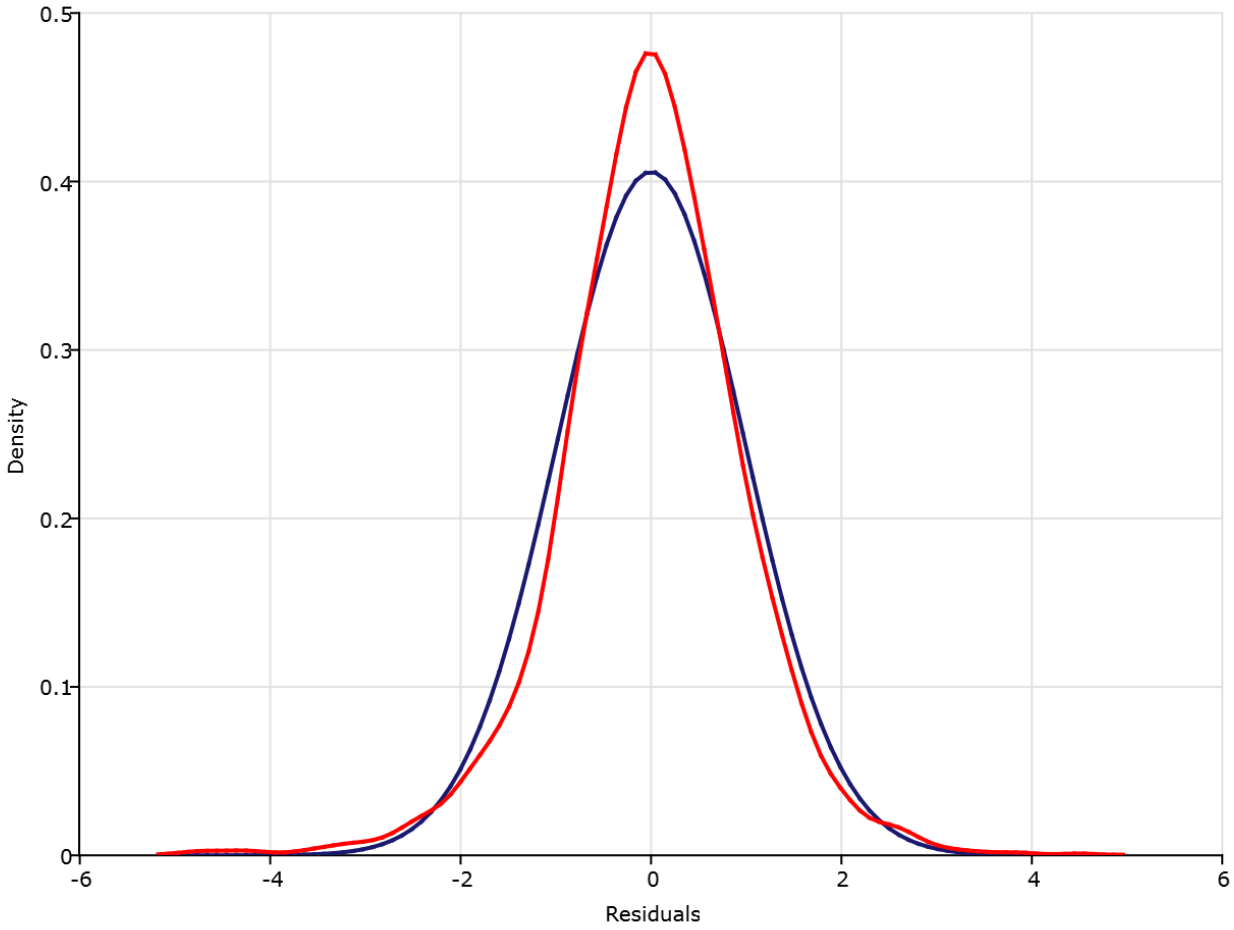
We first estimate the autocorrelation function of $\{\widehat{\varepsilon}_{it},\ t = 1, \ldots, K\}$ denoted $\widehat{\rho}_i(j),\ j = 1, \ldots, J$ for all the available countries that have at least 60 observations, currently $n = 75$ countries. We take $J = 21$ to allow for long lagged effects, which might be predicted for deaths by epidemiological models. We calculate the proportion of rejections against the standard 5% Bartlett intervals across all lags: for cases we find 13.8% (which is reduced to 13.6% when using the bias corrected $\widehat{\rho}_i(j)$) and for deaths we find 6.5% (which increased to 6.6% when using the bias corrected estimator). We also report the estimated mean value of $\widehat{\rho}_i(j)$ across countries along with standard errors that take account of the cross-sectional averaging but allow for cross sectional correlation, Linton (2020, Theorem 16)[16]. These confidence intervals are shrunk quite a lot relative to the Bartlett intervals as the average squared pairwise correlations are quite small. The pattern of autocorrelation is similar in both cases and deaths, weak positive autocorrelation at low lags that declines across horizon to negative autocorrelation after about 8 days, which declines until around 18 days and then returns to zero. The autocorrelation in cases is larger in magnitude than the autocorrelation in deaths.


Cross Country Average for Cases
0.2
0.16
0.12
0.08
0.04
0
-0.04
-0.08
-0.12
Mean of ACF
0
3
6
9
12
15
18
21
Lags


Cross Country Average for Deaths
0.08
0.04
0
-0.04
-0.08
Mean of ACF
0
3
6
9
12
15
18
21
Lags

8.3.3 Distributional Properties

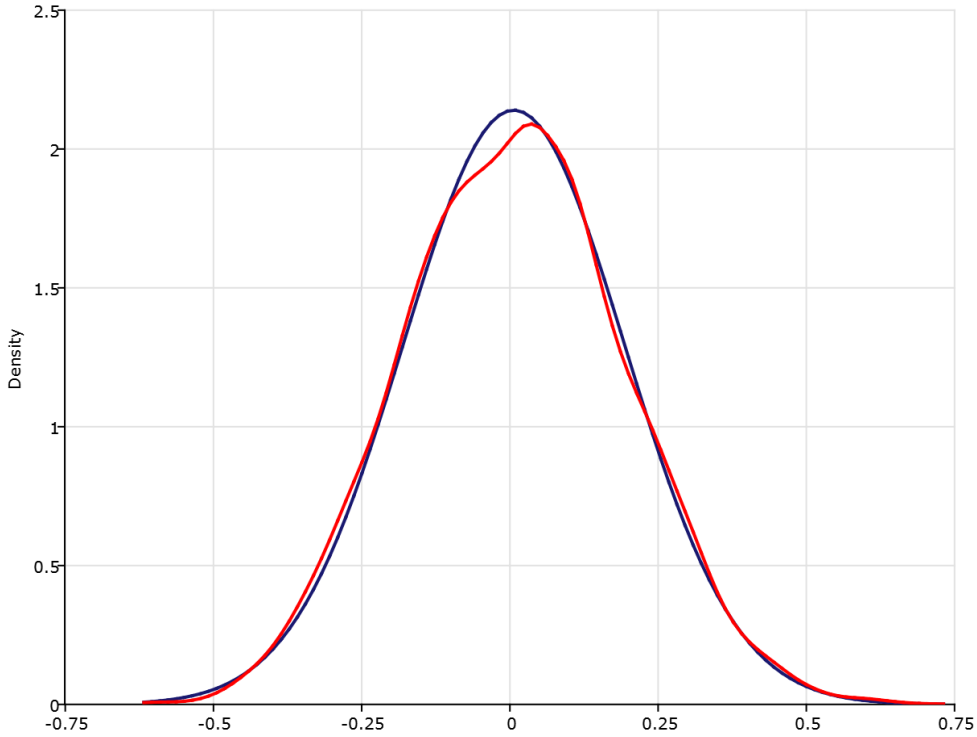
We next show the kernel density estimate of the pooled standardized residuals, which appears not far from a Gaussian shape, at least, roughly symmetrical.



For the deaths data, the results are similar with a mean value of t_i equal to -0.4 based on 133 countries data and 23 violations.

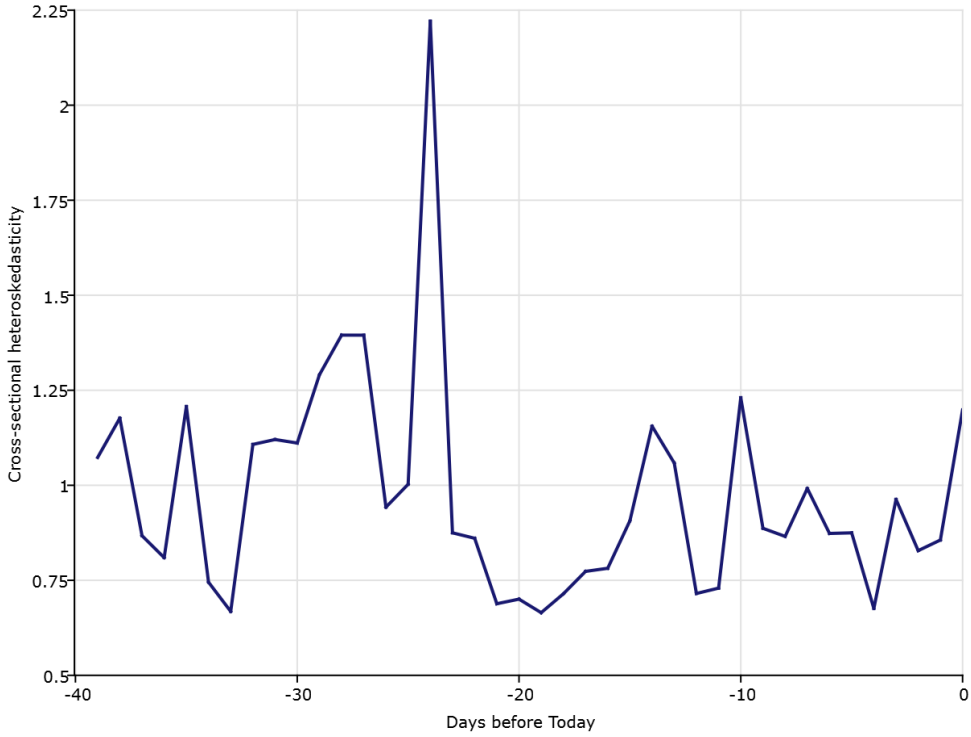
8.3.4 Cross-sectional Correlation

We also looked at the cross-sectional correlation by computing the $n \times n$ covariance matrix of the standardized residuals using the time series data. The ratio of the largest eigenvalue to the smallest non-zero eigenvalue is 17.75/2.39 for the cases data and 10.18/1.61 for the deaths data, revealing a lot of cross-sectional dependence. Below we show the cross-sectional dependence (pairwise correlations) density plot compared with normal distribution. There is a wide range of contemporaneous correlations both positive and negative reflecting calendar effects perhaps; mean value is close to zero.



8.3.5 Heteroskedasticity

We look at time varying heteroskedasticity in the residuals, specifically we graph the time series of mean squared residuals $\sum_{i=1}^{n} \widehat{\varepsilon}_{it}^2/n$. It seems that 24 days ago at the time of writing (April 15th) there was quite a spike in the cross country variability of the error term but otherwise it trades in the range of 0.75 to 1.25, a modest variation.



9 Combining Case and Fatality Models

Total fatalities should be a fraction of the total cases reported, and fatalities should follow cases, at least individually. 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:

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

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

$$m_i^d(t) = \theta_i + m_i^c(t-k), \tag{10}$$

which is a special case of the model considered by Härdle and Marron (1990)[11]. This imposes restrictions across the coefficients of the two quadratic equations. The turning point for m^d occurs k periods after the turning point for m^c, that is, $\alpha^d = \alpha^c + \theta$ and $\mu^d = \mu^c + k$. The only equality restriction is that the γ parameter is the same across both cases and deaths. We can test this by comparing the statistic

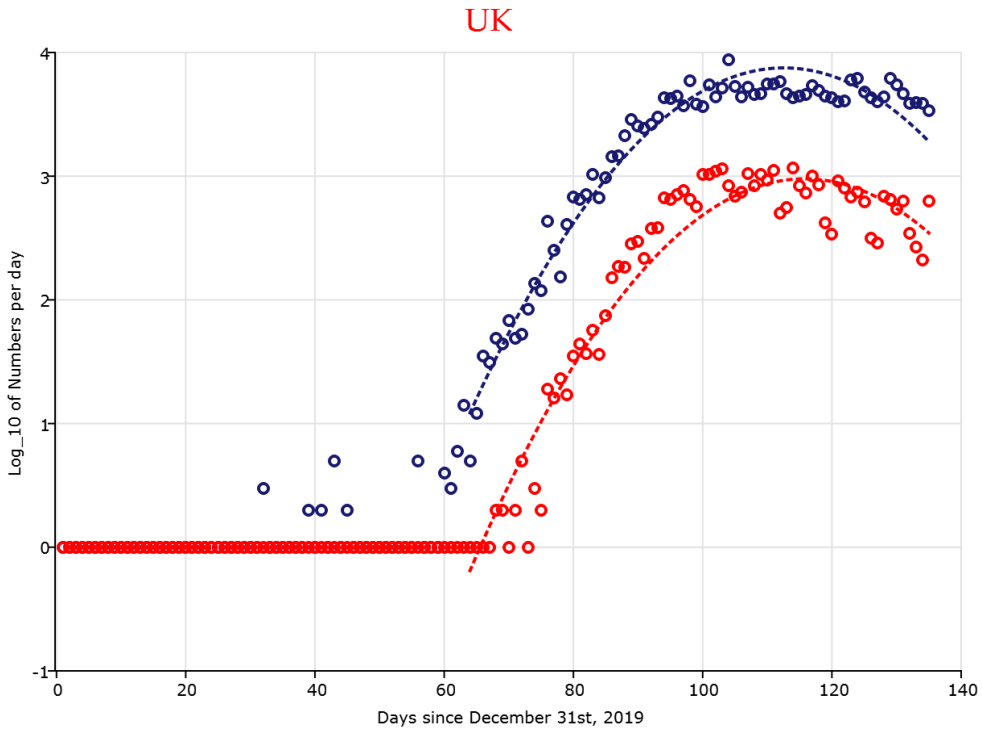
$$\frac{\widehat{\gamma}^d - \widehat{\gamma}^c}{\sqrt{\widehat{\text{var}}(\widehat{\gamma}^d - \widehat{\gamma}^c)}} \tag{11}$$

with the standard normal critical values. Regarding the inequality restrictions $\alpha^d - \alpha^c < 0$ and $\mu^d - \mu^c > 0$, these can also be tested separately by similar t-statistics with one-sided critical values. Specifically, we consider

$$\frac{\hat{\mu}^c - \hat{\mu}^d}{\sqrt{\text{var}(\hat{\mu}^c - \hat{\mu}^d)}}.$$

The tests results are presented below in two tables. The model does not fare well on either count (although the significance mostly disappears when HAC standard errors are used)

We estimate the constrained model as described in the appendix, the results are shown below graphically. For the UK the results look quite plausible.



We calculate the ratio of expected total deaths to expected total cases (the fatality ratio) by country as

$$Fr = \exp(\alpha^d - \alpha^c) = \exp\left(a^d - a^c - \frac{(b^d)^2}{c} + \frac{(b^c)^2}{c}\right).$$

At the time of writing, UK has 0.136, France has 0.170, Italy has 0.132, Spain has 0.101, Germany has 0.045, Australia has 0.052, and USA has 0.051. This may be saying more about the disparity between countries in testing rather than the quality of treatment.

	countryname	$\hat{\gamma}_{case}$	$\hat{\gamma}_{death}$	s.e($\Delta\hat{\gamma}$)	t-stat
1	USA	-7.2226	-12.118	1.026	4.771
2	Spain	-4.5958	-7.7264	1.512	2.071
3	UK	-6.8588	-10.5715	0.593	6.261
4	Russia	-8.0363	-4.0529	1.131	-3.522
5	Italy	-2.5306	-3.0668	0.346	1.55
6	Germany	-4.6078	-12.2052	0.854	8.896
7	Brazil	-3.4188	-6.2389	0.414	6.812
8	France	-5.3866	-10.4125	0.807	6.228
9	Iran	-0.1196	-0.602	0.77	0.626
10	India	-6.0106	-4.5226	0.75	-1.984
11	Canada	-6.8034	-7.361	0.9	0.62
12	Peru	-3.3526	-3.4119	1.184	0.05
13	Belgium	-7.2607	-13.7131	0.879	7.341
14	Netherlands	-6.271	-8.3964	0.647	3.285
15	Saudi_Arabia	-3.2925	-2.7128	1.504	-0.385
16	Ecuador	-3.4408	-3.905	2.661	0.174
17	Mexico	-3.7846	-5.0223	0.808	1.532
18	Pakistan	-0.7756	-1.4371	2.003	0.33
19	Switzerland	-7.0251	-8.2712	1.655	0.753
20	Chile	-1.5235	-4.2449	0.859	3.168
21	Portugal	-6.6726	-9.4248	1.158	2.377
22	Sweden	-4.1509	-8.3115	1.285	3.238
23	Belarus	-5.3347	-1.7972	2.429	-1.456
24	Singapore	-2.2783	-0.5764	2.191	-0.777
25	Qatar	-3.529	-0.6986	1.37	-2.066
26	Ireland	-8.4304	-10.6836	1.24	1.817
27	United Arab Emirates	-6.8651	-0.5727	2.484	-2.533
28	Israel	-11.2753	-6.4296	1.333	-3.635
29	Poland	-4.8589	-7.0389	0.828	2.633
30	Japan	-10.584	-0.1026	1.497	-7.002

Table 3: The Hypotheses Test On $\hat{\gamma}_{case} = \hat{\gamma}_{death}$ of Top 30 Countries

	countryname	$\hat{\mu}_{case}$	$\hat{\mu}_{death}$	s.e($\Delta\hat{\mu}$)	t-stat
1	USA	-0.3831	-0.3435	0.019	-2.084
2	Spain	-0.7423	-0.5298	0.046	-4.62
3	UK	-0.3575	-0.3891	0.012	2.633
4	Russia	-0.0772	0.1824	0.115	-2.257
5	Italy	-0.8334	-0.6924	0.039	-3.615
6	Germany	-0.7135	-0.4151	0.096	-3.108
7	Brazil	0.1647	-0.0508	0.087	2.477
8	France	-0.6483	-0.4764	0.032	-5.372
9	Iran	-0.9973	-1.286	6.049	0.048
10	India	-0.0632	0.0434	0.082	-1.3
11	Canada	-0.3468	-0.1806	0.037	-4.492
12	Peru	0.311	0.2184	0.299	0.31
13	Belgium	-0.5133	-0.4169	0.013	-7.415
14	Netherlands	-0.5661	-0.4722	0.013	-7.223
15	Saudi_Arabia	0.2371	-0.0481	0.335	0.851
16	Ecuador	-0.2575	-0.1271	0.233	-0.56
17	Mexico	0.0981	0.0802	0.103	0.174
18	Pakistan	2.3923	0.7767	9.155	0.176
19	Switzerland	-0.7017	-0.504	0.052	-3.802
20	Chile	0.4984	-0.2086	0.439	1.61
21	Portugal	-0.5464	-0.4463	0.027	-3.707
22	Sweden	-0.3231	-0.3375	0.033	0.436
23	Belarus	0.147	0.1089	0.32	0.119
24	Singapore	0.4327	-0.3475	0.906	0.861
25	Qatar	0.2571	-0.4391	0.303	2.298
26	Ireland	-0.4618	-0.3727	0.021	-4.243
27	United Arab Emirates	-0.1067	1.71	3.397	-0.535
28	Israel	-0.5716	-0.4177	0.025	-6.156
29	Poland	-0.3467	-0.2999	0.03	-1.56
30	Japan	-0.4339	12.1906	152.684	-0.083

Table 4: The Hypotheses Test On $\hat{\mu}_{case} < \hat{\mu}_{death}$ of Top 30 Countries

Epidemiological models often build in the effect of lagged cases on deaths, reflecting the natural causal ordering of cases on deaths. We consider an alternative model that combines that feature with our quadratic regression. We suppose that the log death regression is a weighted delay of log cases regression

$$m^d(t) = \int_0^\infty w_s m^c(t-s)ds. \tag{12}$$

In principle, if m^c, m^d were both nonparametric and estimable one might be able to estimate the function w by sieve expansion, using the methods described in inter alia Chen (2007)[4] and Chen and Christensen (2015)[5]. We take a parametric approach where both m functions are quadratic. This is possible provided

$$\begin{aligned} a^d + b^d t + c^d t^2 &= \int_0^\infty w_s \left(a^c + b^c(t-s) + c^c(t-s)^2\right) ds \\ &= \int_0^\infty w_s \left(a^c + b^c t + c^c t^2\right) ds - \int_0^\infty w_s s \left(b^c + 2c^c t\right) ds + c^c \int_0^\infty w_s s^2 ds \\ &= a^c \int_0^\infty w_s ds - b^c \int_0^\infty w_s s ds + c^c \int_0^\infty w_s s^2 ds + t\left(b^c \int_0^\infty w_s ds - 2c^c \int_0^\infty w_s s ds\right) \\ &\quad + t^2 c^c \int_0^\infty w_s ds. \end{aligned}$$

We suppose that w is a constant ψ times the density of a normal random variable whose mean is ϑ and whose variance is one. This shape seems plausible with a peak loading some days previous. We need the constants $\varphi_j = \int_0^\infty s^j w_s ds$. Suppose that $\varphi_j(\theta)$, where $\theta = (\psi, \vartheta) \in \mathbb{R}^2$; these constants are obtainable in closed form as integrals of normal densities, Barr and Sherrill (1999)[3]:

$$\int_0^\infty w_s ds = \psi \int_{-\vartheta}^\infty \phi(s)ds = \psi\left(1 - \Phi(-\vartheta)\right) = \psi\Phi(\vartheta)$$

$$\int_0^\infty w_s s ds = \psi \int_{-\vartheta}^\infty \phi(s)s ds = \psi E(Z|Z > -\vartheta)\Pr(Z > -\vartheta) = \psi\phi(\vartheta)$$

$$\begin{aligned} \int_0^\infty w_s s^2 ds &= \psi \int_{-\vartheta}^\infty \phi(s)s^2 ds = \psi E(Z^2|Z > -\vartheta)\Pr(Z > -\vartheta) = \psi\frac{1}{2}\int_{\vartheta^2}^\infty \frac{1}{2^{3/2}\Gamma(3/2)} u^{1/2} e^{-u/2} du \\ &= \psi\frac{1}{2}\left(1 - F_{\chi^2(3)}(\vartheta^2)\right), \end{aligned}$$

where ϕ, Φ are the density function and c.d.f. of a standard normal random variable. This expresses the parameters of the death curve uniquely as a function of the parameters of the case curve and the two parameters ψ, ϑ. In particular, the location of the death peak is related to the case peak as follows

$$\mu^d = \frac{-b^d}{2c^d} = -\frac{b^c \int_0^\infty w_s ds - 2c^c \int_0^\infty w_s s ds}{2c^c \int_0^\infty w_s ds} = \frac{-b^c}{2c^c} + \frac{\int_0^\infty w_s s ds}{\int_0^\infty w_s ds} = \frac{-b^c}{2c^c} + \frac{\phi(\vartheta)}{\Phi(\vartheta)} = \mu^c + \lambda(\vartheta).$$

Since the Heckman correction $\lambda(\vartheta) \geq 0$, this implies that $\mu^d \geq \mu^c$ as for the previous model. This equation allows the determination of ϑ because the function $\lambda(\vartheta) = \phi(\vartheta)/\Phi(\vartheta) > 0$ is monotonic decreasing. That is, $\vartheta = \lambda^{-1}(\mu^d - \mu^c)$. Whence, $\psi = c^d/\varphi_0(\vartheta)c^c$. It follows that the quadratic model for both cases and deaths is compatible with the posited relation between deaths and cases.

Regarding the peak of the two curves, we have

$$m^d_{\max} = a^d - \frac{(b^d)^2}{4c^d} = \psi \left[\varphi_0 m^c_{\max} + c^c \left(\varphi_2 - \frac{\varphi_1^2}{\varphi_0} \right) \right].$$

By the Cauchy-Schwarz inequality, $\varphi_2 - \frac{\varphi_1^2}{\varphi_0} \geq 0$ so that with negative c^c we have $m^d_{\max} \leq \psi\varphi_0 m^c_{\max}$. This model imposes exactly one restriction like the shape invariant model but the nature of the restrictions are quite different. We plan to investigate this model in the sequel.

10 Conclusion

There are many challenges in modelling the COVID data. Countries differ widely in their reporting methods and standards, which makes the data noisy and sometimes very unreliable. Our model does not impose any restrictions across countries for the evolution of the epidemic and we find there are very large differences across countries in all the key parameters. Another complication is due to different interventions implemented in different countries at different times. Actually, most European countries introduced lockdown measures within the month of March 2020 so that whatever effects these measures have had will be fully reflected in the subsequent data. Our model allows us to estimate future turning points of the curves without imposing much structure. It also allows the forecasting of total number of cases, although the confidence intervals around such estimates are extremely large when proper attention is given to parameter uncertainty.

11 Appendix

11.1 Forecasting Test

We discuss here our approach to obtaining prediction intervals. We have a classical linear regression with K observations:

$$y_t = \theta^\intercal x_t + \varepsilon_t, \quad t = 1, \ldots, K$$

$$\widehat{\theta} = (X^\intercal X)^{-1} X^\intercal y$$

$$V^{-1/2}(\widehat{\theta} - \theta) \Longrightarrow N(0, I), \quad V = \sigma^2_\varepsilon (X^\intercal X)^{-1}$$

$$y_{K+s|K} = \widehat{\theta}^\intercal x_{K+s}.$$

We have

$$y_{K+s} - y_{K+s|K} = (\widehat{\theta} - \theta)^{\intercal} x_{K+s} + \varepsilon_{K+s} \simeq N\left(0, \sigma_{\varepsilon}^2 x_{K+s}^{\intercal}(X^{\intercal}X)^{-1}x_{K+s}\right) + \varepsilon_{K+s},$$

where the two random variables are independent by assumption.

As $K \to \infty$, the parameter uncertainty is small for given s, but as $s \to \infty$, the parameter uncertainty grows without bound (in our case since $x_{K+s} = (1, 1 + s/K, (1 + s/K)^2)^{\intercal}$). In fact the estimator of m is only consistent in the range where $s^2/K \to 0$. Our intervals, or any intervals, are only valid over short horizons without strong additional assumptions or Bayesian magic. In our graphs above we have ignored the contribution from estimation uncertainty, which mostly affects long term prediction intervals, and affects them dramatically.

A simple approach is to assume normality for ε_{K+s}, in which case $\varepsilon_{K+s|K} = y_{K+s} - y_{K+s|K} \sim N\left(0, \sigma_{\varepsilon}^2(1 + x_{K+s}^{\intercal}(X^{\intercal}X)^{-1}x_{K+s})\right)$. We provide a test of the vector of one step ahead predictions for $n = 30$ countries. We have for the vector of one step ahead predictions

$$\varepsilon_{K+1|K} \sim N\left(0, (1 + x_{K+1}^{\intercal}(X^{\intercal}X)^{-1}x_{K+1})\Omega_{\varepsilon}\right),$$

where Ω_{ε} is the error covariance matrix, so that

$$t_i = \frac{y_{K+1} - y_{K+1|K}}{\sqrt{(1 + x_{K+1}^{\intercal}(X^{\intercal}X)^{-1}x_{K+1})\widehat{\sigma}_{ii}}},$$

is approximately standard normal for each i. Since $n > K$, we use a version of the Pesaran and Yamagata (2012) statistic to aggregate across countries

$$\tau_{PY} = \frac{\sum_{i=1}^{n} t_i^2 - n}{\sqrt{2i_n^{\intercal}\left(\widehat{R}_{\varepsilon} \odot \widehat{R}_{\varepsilon}\right) i_n}}, \quad \widehat{R}_{\varepsilon} = \mathrm{diag}(\widehat{\Omega}_{\varepsilon})^{-1/2}\widehat{\Omega}_{\varepsilon}\mathrm{diag}(\widehat{\Omega}_{\varepsilon})^{-1/2},$$

which is asymptotically standard normal under the null (provided n is large and the cross-sectional dependence is weak); the test is rejected if $\tau_{PY} > z_{1-\alpha}$, where $\Phi(z_a) = a$. Here, $\odot$ denotes Hadamard product and $\widehat{\Omega}_{\varepsilon}$ is the residual covariance matrix estimate

$$\widehat{\Omega}_{\varepsilon} = \frac{1}{K}\sum_{t=1}^{K} \widehat{\varepsilon}_t \widehat{\varepsilon}_t^{\intercal}.$$

These are the statistics reported in Tables 3 and 4.

11.2 Quantile Regression

We discuss next how to estimate this model. Let $\widehat{\theta} = (\widehat{\alpha}, \widehat{\gamma}, \widehat{\mu})^{\intercal}$ minimize

$$Q_K(\theta) = \frac{1}{K}\sum_{t=1}^{K} \left| y_t - \alpha - \gamma|t - \mu|^{\lambda} \right|.$$

In practice we solve this problem in two steps. Let

$$X(\mu,\lambda)=\left(\begin{array}{cc}1 & |1/K-\mu|^{\lambda}\\ & \vdots \\ 1 & |1-\mu|^{\lambda}\end{array}\right).$$

We define the profiled linear quantile regression estimators

$$\left(\begin{array}{c}\widehat{\alpha}(\mu)\\ \widehat{\gamma}(\mu)\end{array}\right)=\arg\min\sum_{t=1}^{T}\left|y_t-x_t^{\intercal}(\mu)\left(\begin{array}{c}\alpha\\ \gamma\end{array}\right)\right|$$

for each value of μ, and then let

$$Q_K^P(\mu)=\sum_{t=1}^{K}\left|y_t-x_t^{\intercal}(\mu)\left(\begin{array}{c}\widehat{\alpha}(\mu)\\ \widehat{\gamma}(\mu)\end{array}\right)\right|,$$

and then let $\widehat{\mu}$ minimize $Q_K^P(\mu)$ with respect to μ. This is done by grid search over $\mu\in[0.5\widehat{\mu}_{\lambda=2},1.5\widehat{\mu}_{\lambda=2}]$. We have:

$$\frac{\partial\ell}{\partial\alpha}=\sum_{t=1}^{K}\text{sign}(y_t-\alpha-\gamma|t-\mu|^{\lambda})$$

$$\frac{\partial\ell}{\partial\gamma}=\sum_{t=1}^{K}\text{sign}(y_t-\alpha-\gamma|t-\mu|^{\lambda})|t-\mu|^{\lambda}$$

$$\frac{\partial\ell}{\partial\mu}=\sum_{t=1}^{K}\text{sign}(y_t-\alpha-\gamma|t-\mu|^{\lambda})\gamma\lambda|t-\mu|^{\lambda-1}$$

In fact, let

$$\mathcal{M}=\left[\begin{array}{ccc}1 & M_{\lambda,0} & \gamma\lambda M_{\lambda-1,0}\\ M_{\lambda,0} & M_{2\lambda,0} & \gamma\lambda M_{2\lambda-1,0}\\ \gamma\lambda M_{\lambda-1,0} & \gamma\lambda M_{2\lambda-1,0} & \gamma^2\lambda^2 M_{2\lambda-2,0}\end{array}\right]$$

$$M_{r,s}(\mu)=\frac{1}{K}\sum_{t=1}^{K}|t-\mu|^{r}\ln^{s}(|t-\mu|).$$

It follows that (under iid error assumptions) as $K\to\infty$

$$\sqrt{K}\left(\widehat{\theta}-\theta\right)\Longrightarrow N(0,\frac{1}{4f_{\varepsilon}^2(0)}\mathcal{M}^{-1}).$$

Standard errors are computed as

$$\frac{1}{\sqrt{K4\widehat{f}_{\varepsilon}^2(0)}}\text{sqrt}(\text{diag}(\widehat{\mathcal{M}}^{-1})),$$

In practice we define the $K \times 3$ data matrix

$$X = \left(\begin{array}{ccc} 1 & |t_i - \widehat{\mu}|^{\widehat{\lambda}} & \widehat{\gamma}\widehat{\lambda}|t_i - \widehat{\mu}|^{\widehat{\lambda}-1} \end{array} \right)_{i=1}^{K}$$

$$\widehat{\mathcal{M}} = X^{\intercal} X / K.$$

In an earlier update we observed that the density of the pooled errors was close to symmetric and higher peaked than a corresponding normal. For this reason, we replace $\widehat{f}_{\varepsilon}^{2}(0)$ by $1/2\pi\widehat{\sigma}_{R}^{2}$, where $\widehat{\sigma}_{R}^{2} = 1.349 \times (q_{\varepsilon}(0.75) - q_{\varepsilon}(0.25))$, where $q_{\varepsilon}(\alpha)$ is the $\alpha-$ quantile of the residuals.

The theory for estimation of λ is also available, but in practice there is a lot of colinearity such that the standard errors for λ are quite wide and this also affects the standard errors for μ.

11.3 Restricted Model Estimation

We next discuss how to estimate the restricted model of Section 5. In this case we have a pair of quadratic equations with an equality restriction,

$$\log\left(y_t^c + 1\right) = \alpha^c + \beta^c t + \gamma^c t^2 + \varepsilon_t^c,$$

$$\log\left(y_t^d + 1\right) = \alpha^d + \beta^d t + \gamma^d t^2 + \varepsilon_t^d,$$

where

$$E\left[\left(\begin{array}{c} \varepsilon_t^c \\ \varepsilon_t^d \end{array} \right) \left(\begin{array}{cc} \varepsilon_t^c & \varepsilon_t^d \end{array} \right) \right] = \Omega.$$

This is a SUR with a cross-equation restriction and the optimal estimator is a GLS. The unrestricted estimator vector $\widehat{\theta} = (\widehat{\alpha}^c, \widehat{\beta}^c, \widehat{\gamma}^c, \widehat{\alpha}^d, \widehat{\beta}^d, \widehat{\gamma}^d)^{\intercal}$ has variance $\Omega \otimes (X^{\intercal}X)^{-1}$. We define $\theta = (\alpha^c, \beta^c, \alpha^d, \beta^d, \gamma)^{\intercal}$ and L the 6×5 matrix of zeros and ones that picks out the right element

$$\widetilde{\theta} = \left(L^{\intercal} \left(\widehat{\Omega}^{-1} \otimes (X^{\intercal}X) \right) L \right)^{-1} L^{\intercal} \left(\widehat{\Omega}^{-1} \otimes (X^{\intercal}X) \right) \widehat{\theta}$$

with asymptotic variance

$$\left(L^{\intercal} \left(\Omega^{-1} \otimes (X^{\intercal}X) \right) L \right)^{-1}.$$

In our second model, the quasi likelihood can be used with one step taken from initial consistent estimators of $\theta = (a^c, b^c, c^c, \psi, \vartheta)^{\intercal}$. The quasi-likelihood is

$$\ell(\theta, \Omega) = -\frac{K}{2} \log\det\Omega - \frac{1}{2} \sum_{t=1}^{K} \left(y_t - m_t(\theta)\right)^{\intercal} \Omega^{-1} \left(y_t - m_t(\theta)\right),$$

where $y_t = (y_t^c, y_t^d)^{\intercal}$ and $m_t = (m_t^c, m_t^d)^{\intercal}$, where, with $\varphi_0(\vartheta) = \Phi(\vartheta)$, $\varphi_1(\vartheta) = \phi(\vartheta)$, and $\varphi_2(\vartheta) = \frac{1}{2}(1 - F_{\chi^2(3)}(\vartheta^2))$, we have

$$m_t^d = \psi\left(a^c\varphi_0(\vartheta) - b^c\varphi_1(\vartheta) + c^c\varphi_2(\vartheta)\right) + t\psi\left(b^c\varphi_0(\vartheta) - 2c^c\varphi_1(\vartheta)\right) + t^2\psi c^c\varphi_0(\vartheta).$$

This is an SUR system with a nonlinear cross equation restriction. The efficient estimator should use the error covariance matrix.

We can alternatively estimate the unrestricted model $(a^c, b^c, c^c, a^d, b^d, c^d)^\intercal$ and then impose the restrictions afterwards by minimum distance where

$$\begin{pmatrix} a^d \\ b^d \\ c^d \end{pmatrix} = \psi \begin{pmatrix} \varphi_0(\vartheta) & -\varphi_1(\vartheta) & \varphi_2(\vartheta) \\ 0 & \varphi_0(\vartheta) & -2\varphi_1(\vartheta) \\ 0 & 0 & \varphi_0(\vartheta) \end{pmatrix} \begin{pmatrix} a^c \\ b^c \\ c^c \end{pmatrix}.$$

for some norm. This is similar to the case we have already studied because there are five unrestricted parameters and 6 estimable quantities.

References

[1] Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. A multi-risk sir model with optimally targeted lockdown. Tech. rep., National Bureau of Economic Research, 2020.

[2] Anastassopoulou, C., Russo, L., Tsakris, A., and Siettos, C. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PloS one 15*, 3 (2020), e0230405.

[3] Barr, D. R., and Sherrill, E. T. Mean and variance of truncated normal distributions. *The American Statistician 53*, 4 (1999), 357–361.

[4] CHEN, X. Vlarge% sample sieve estimation of semi% nonparameteric modelsv, chapter 76 in j. heckman and e. leamer (eds), handbook of econometrics. volume 6b, 2007.

[5] Chen, X., and Christensen, T. M. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics 188*, 2 (2015), 447–465.

[6] Chowell, G., Tariq, A., and Hyman, J. M. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. *BMC medicine 17*, 1 (2019), 164.

[7] Chudik, A., Pesaran, M. H., and Rebucci, A. Voluntary and mandatory social distancing: Evidence on covid-19 exposure rates from chinese provinces and selected countries. Tech. rep., National Bureau of Economic Research, 2020.

[8] Deb, S., and Majumdar, M. A time series method to analyze incidence pattern and estimate reproduction number of covid-19. *arXiv preprint arXiv:2003.10655* (2020).

[9] Fan, J., and Gijbels, I. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, vol. 66. CRC Press, 1996.

[10] Hafner, C. M. The spread of the covid-19 pandemic in time and space. *working paper* (2020).

[11] Hardle, W., and Marron, J. S. Semiparametric comparison of regression curves. *The Annals of Statistics* (1990), 63–89.

[12] Hu, Z., Ge, Q., Jin, L., and Xiong, M. Artificial intelligence forecasting of covid-19 in china. *arXiv preprint arXiv:2002.07112* (2020).

[13] Hualde, J., Robinson, P. M., et al. Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics 39*, 6 (2011), 3152–3181.

[14] Li, Q., Feng, W., and Quan, Y.-H. Trend and forecasting of the covid-19 outbreak in china. *Journal of Infection 80*, 4 (2020), 469–496.

[15] Lin, Q., Hu, T., and Zhou, X.-H. Estimating the daily trend in the size of covid-19 infected population in wuhan. *medRxiv* (2020).

[16] Linton, O. B. *The Models and Methods of Financial Econometrics.* Cambridge University Press, 2020.

[17] Liu, L., Moon, H. R., and Schorfheide, F. Panel forecasts of country-level covid-19 infections. *working paper* (2020).

[18] Marinucci, D., and Robinson, P. M. Semiparametric fractional cointegration analysis. *Journal of Econometrics 105*, 1 (2001), 225–247.

[19] Park, J. Y., and Phillips, P. C. Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory 15*, 3 (1999), 269–298.

[20] Park, J. Y., and Phillips, P. C. Nonlinear regressions with integrated time series. *Econometrica 69*, 1 (2001), 117–161.

[21] Peng, L., Yang, W., Zhang, D., Zhuge, C., and Hong, L. Epidemic analysis of covid-19 in china by dynamical modeling. *arXiv preprint arXiv:2002.06563* (2020).

[22] Pesaran, M. H., and Yamagata, T. Testing capm with a large number of assets. In *AFA 2013 San Diego Meetings Paper* (2012).

[23] Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A., and Jewell, C. P. Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions. *MedRxiv* (2020).

[24] Richards, F. A flexible growth function for empirical use. *Journal of experimental Botany 10*, 2 (1959), 290–301.

[25] Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., Yan, P., and Chowell, G. Real-time forecasts of the covid-19 epidemic in china from february 5th to february 24th, 2020. *Infectious Disease Modelling 5* (2020), 256–263.

[26] Wang, Q. Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory 30*, 3 (2014), 509–535.

[27] Wang, Q. *Limit theorems for nonlinear cointegrating regression*. World Scientific, 2015.

[28] Wu, J. T., Leung, K., and Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet 395*, 10225 (2020), 689–697.