Global Empirical Forecasts of COVID-19 Trajectories Under Limited Information on the Efficacy of Intervention Strategies

KAI LIN[+], CHRISTOPHER JOYE[+], NATHAN GIANG[+] AND DR ADAM RICHARDSON[+]

[+]Coolabah Capital Investments

First Draft: 23 March 2020

This Draft: 2 April 2020

Abstract

As the novel coronavirus and its associated disease COVID-19 started to rapidly transmit around the world in early 2020, the financial, social and health impacts represented a 1-in-100 year shock, the likes of which had not been observed since the last global pandemic in 1918 and the Great Depression in 1929. A key question for policymakers, medical researchers, and financial market participants was how the disease would propagate in an environment in which it was left unconstrained as compared with preferable alternatives where nation states implemented assertive efforts to mitigate the disease's adverse effects. Medical researchers seeking to advise governments produced theoretical forecasting models, drawing on the epidemiological literature, which have often been too inflexible and abstract for use by financial markets. For this niche user group, empirical, agile, and intervention-aware forecasting methods are paramount, especially those that can accommodate the subjective judgements of different users. This paper outlines two such empirical forecasting frameworks for the daily confirmed case counts, eventual case counts, and time to peak daily new case counts for major countries. The first framework uses a linear mixed effect model for the case growth rate, accounting for the presence of intervention measures and idiosyncrasies of individual countries. The second framework allows users to forecast the case trends of a target country by substituting in the observed effects of interventions from qualitatively similar countries with customisable calibrations to reflect lower efficacies. Combined, these two frameworks are especially useful in the early days of the outbreak, when the effects of different countries' imminent interventions have not yet shown up in observed data, but which can be inferred from similar countries further along their intervention path. When first applied and published on March 23, these models projected the peak in daily new COVID-19 case counts for the US and Australia would arrive in early-to-mid April 2020. To the best of our knowledge, this was one of the first early-to-mid April peak projections published globally. Whilst not theoretically founded in the mechanisms of infectious disease, such empirical forecast frameworks offer versatile and parsimonious projections for financial market participants seeking to make decisions under conditions of uncertainty apropos the efficacies of different intervention measures around the world.

Keywords: Coronavirus, COVID-19, Forecasts, Financial Markets

Global Empirical Forecasts of COVID-19 Trajectories Under Limited Information on the Efficacy of Intervention Strategies

KAI LIN+, CHRISTOPHER JOYE+, NATHAN GIANG+ AND DR ADAM RICHARDSON+

+Coolabah Capital Investments

As the novel coronavirus and its associated disease COVID-19 started to rapidly transmit around the world in early 2020, the financial, social and health consequences represented a 1-in-100 year shock, the likes of which had not been observed since the last global pandemic in 1918 and the Great Depression in 1929.

A key question for policymakers, medical researchers, and financial market participants was how the disease would propagate in an environment in which it was left unconstrained as compared with preferable alternatives where nation states undertook assertive efforts (ie, containment policies) to mitigate the disease's adverse effects.

Medical researchers seeking to advise governments have produced theoretical forecasting models, drawing on the epidemiological literature[1], which have often been too inflexible and abstract for use by financial market participants. For this niche user group, empirical, agile and intervention-aware forecasting methods are paramount, especially those that can accommodate calibrations based on users' judgements regarding the relative intensity of different nation states' containment policies.

This paper outlines two such empirical forecasting frameworks for the daily confirmed case counts, eventual case counts, and time to peak daily new case counts for major countries. The first framework uses a linear mixed effect model for the case growth rate, accounting for the presence of intervention measures and the idiosyncrasies of individual countries.

The second framework allows users to forecast the case trends of a target country by substituting in the observed effects of interventions from qualitatively similar countries. Combined, these two frameworks are especially useful in the early days of the outbreak, when the effects of different countries' imminent interventions have not yet shown up in observed data, but which can be estimated from similar countries further along their intervention path.

This paper's methodology and associated forecasts were first published on March 23 [6], with the application specific goal of forecasting trends in key developed countries, including the US, the UK, Italy, France, Spain and Australia. Our methods were specifically designed to address the limited information available at the time regarding the efficacy of different governments' containment strategies.

All figures presented in this paper were produced with the information available up to and including March 23 to reflect the state under which the methodology decisions were made. Note that these methodologies are also applicable to many other countries.

These methods were coded in Coolabah Capital Investments' data science systems to function in real-time with automated live updates that presented the projections to portfolio managers via graphical user interfaces.

The paper is organised as follows. Section 1 outlines the data visualisation and intuitions that culminated in the statistical model for a country-level case growth rate. Section 2 outlines the application of this model for case forecasting, the results of which are discussed in Section 3. All data on daily counts of confirmed cases, deaths and recoveries for COVID-19 are sourced from the John Hopkins Coronavirus Repository [2].

1. Data Visualisation and Model

In this section, we outline the visualisation of trends in Covid-19 confirmed cases, and intuitions that culminated in the statistical model for a country-level case growth rate

1.1 Case Growth Trajectory

The first widely known observation is that cumulative total confirmed case counts for COVID-19 in countries around the world mostly follow an exponential growth curve, once the infection has taken hold within the country (arbitrarily defined as having at least 100 confirmed cases). Refer to Figure 1. A straight line on this plot, where the y axis is on a log scale, means that infections are experiencing exponential growth. Of course, active government policy decisions to contain the virus can exert substantial influence on future infection trajectories. In this plot, the vertical dotted lines indicate the advent of major public intervention policies, defined here as mass closures of services, limitations on human mobility, contact and proximity, and/or the introduction of an extremely comprehensive testing regime.

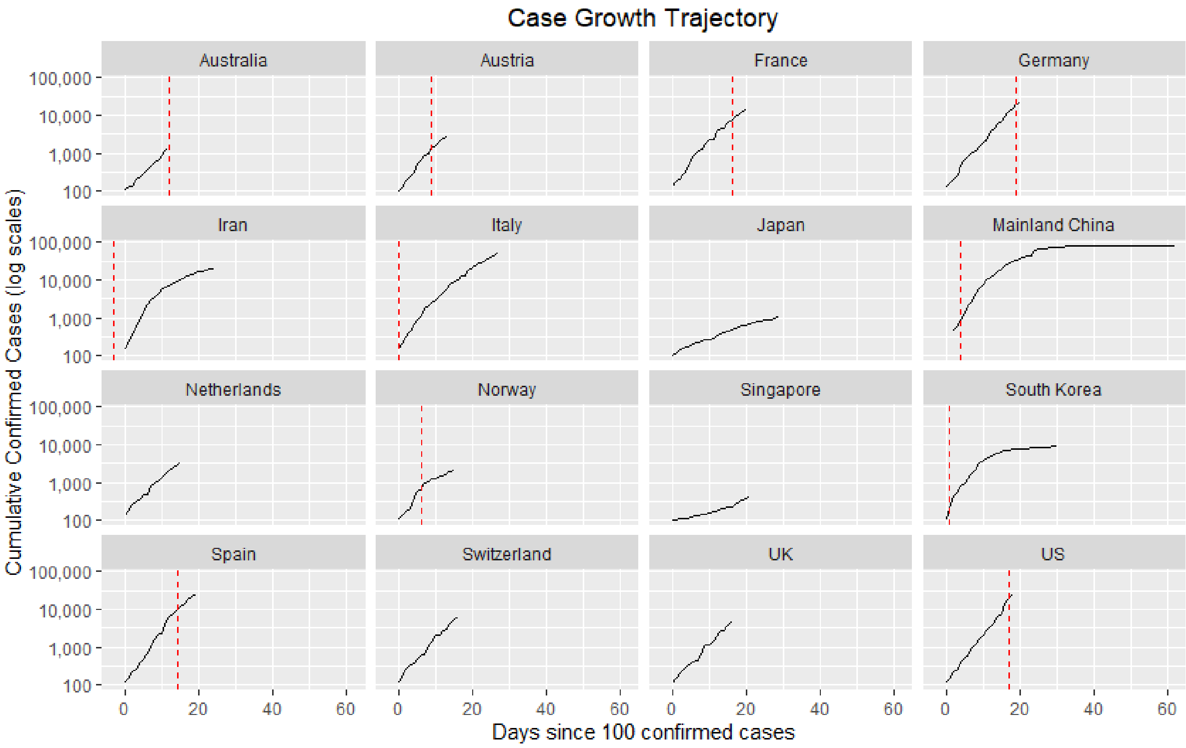

Figure 1

1.2 Evolution of the Case Growth Trajectory

Next we investigate the evolution of the above case growth trajectory. It is also a widely known phenomenon that case counts of an infectious disease cannot grow exponentially forever. Many empirical researchers have applied 'S shaped' functions to infectious disease case trajectories, including for COVID-19 [3]. This is also exhibited in our data visualisation. As noted above in Figure 1, the trajectories are rarely a single straight line in the log domain. Most have a slight trend towards flattening. Some appear to have bigger reductions in steepness in response to public policy intervention measures. This indicates that the infection growth rate is decreasing over time.

One measure of infection growth rate is the number of daily new cases as a percentage of the number of outstanding active (ie, infectious) cases. In Figure 2, we plot this against the number of days since 100 confirmed cases. Note that the number of outstanding cases is defined as the number of cumulative confirmed cases, less the cumulative deaths, and less the cumulative recovered cases.

We see that this percentage mostly holds constant in countries without active intervention measures at the time (e.g., Australia and US), indicating unchecked exponential infection growth. Note that for this metric, a geometric smoothing over 2 days is used as case data reporting can often be delayed by a day in many countries.

Countries with extreme intervention measures, such as China and South Korea, have a curve with steep negative slope, indicating the process of bringing the epidemic under control. There is visual suggestion that intervention measures reduce the slope of this line, to be more negatively sloped. Many European countries also exhibit a slightly negatively sloped line even before major interventions as at the date these data were published (i.e., March 23).

What is useful about this curve is that once plotted in this manner with the y axis in log domain, it appears piecewise linear reflecting active public policy interventions to reduce the virus's transmission rate. This allows the characterisation of countries' trajectories using up to two coefficients, namely the slope of the line in this plot before and after government intervention.

These slopes are fitted via a linear mixed effect model described in the next section. This further permits calibrated "what-if" analysis, by applying the coefficients of other countries (eg, countries with successful containment strategies such as South Korea) to countries in the earlier stages of applying intervention (e.g., the US) with adjustments to reflect the latter's expected containment intensity and efficacy.

In Figure 3, projected blue lines are extrapolated from the country's model fitted slope, which reflect the global average of observed policy interventions. The grey lines are projections applying the coefficient of specific countries as at the date of the forecast (e.g., applying coefficients from South Korea, China, Italy and so on to, say, the US from March 23 onwards).

Note that the idea that the growth rate being log linear is also described by the Gompertz Curve was subsequently brought to our attention by [5], and has been applied in empirical research on COVID-19 [3]. While previous applications of the Gompertz curve focused on the rate of change in confirmed cases, this paper's method defines the rate of change differently as the number of new cases as a percentage of outstanding cases because it is the outstanding cases that are still infectious and hence exponentially driving new cases. We believe that this is a critical insight for forecasting purposes.

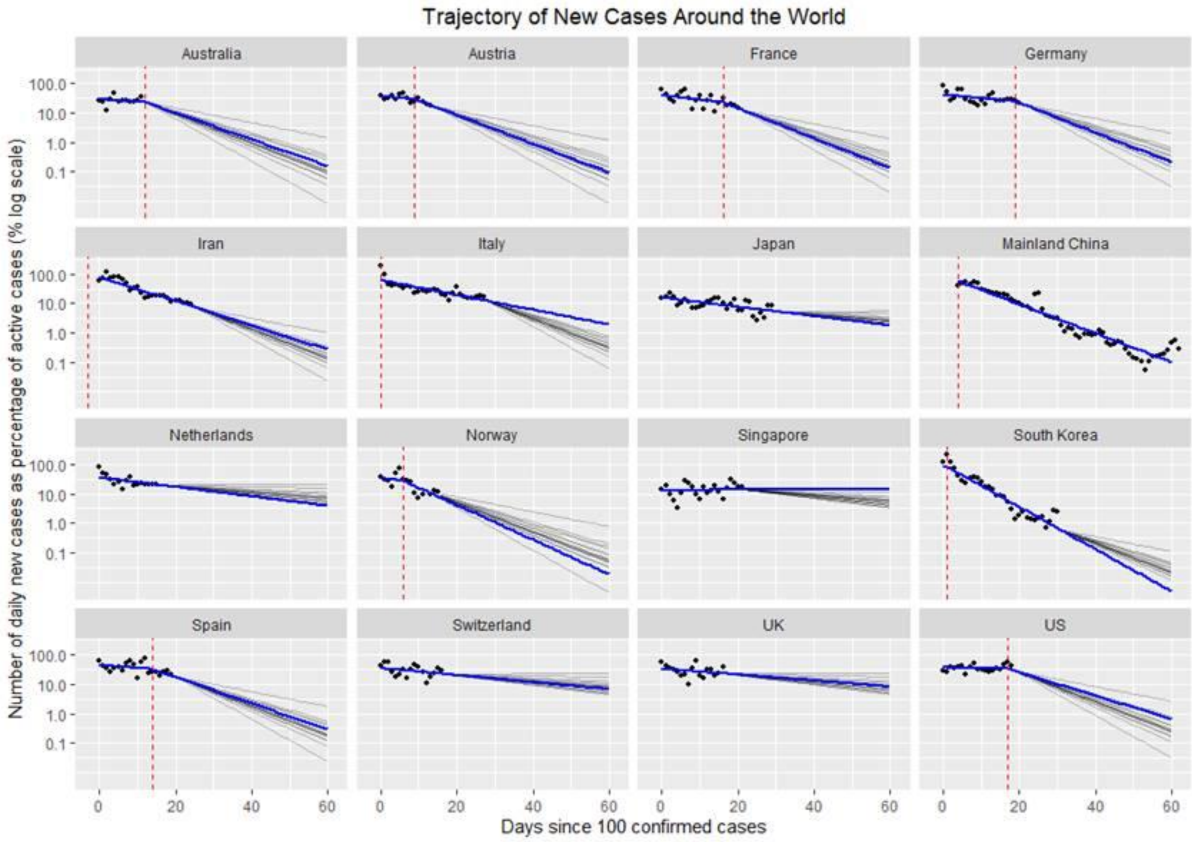

Figure 2

1.3 Model

As noted earlier, a linear mixed effect model was built for the logarithm of the growth rates of COVID-19 confirmed cases. It is defined as follows:

$$\log(r_m[t]) = \alpha + \beta t + \gamma i + a_m + b_m t + c_m i + \epsilon$$

Where

- $r_m[t]$: number of daily new cases as a percentage of outstanding cases, for country m, t days since the date of surpassing 100 confirmed cases
- t: days since 100 cases
- i: days since introduction of intervention, with negative values capped at 0
- α, β, γ: fixed effects
- a_m, b_m, c_m: random effects for country m, assumed to be distributed as a zero mean multivariate normal
- ϵ: residual error, assumed to be independent and identically distributed zero mean normal

An alternative but equivalent model definition in the programming language R is also given:

$$\log(r) \sim 1 + t + i + (1|m) + (t - 1|m) + (i - 1|m)$$

Where:

- r, t, i, m: as defined above

- $(x - 1|m)$: denotes a random effect for x, grouping by country
- 1: intercept
- -1: exclude intercept

2. Forecast of Cases

By utilising the trajectory evolution predictions from Section 1, we can forecast COVID-19 case numbers for different countries.

The growth rate(number of daily new cases as a percentage of outstanding cases) for each country is forecast by extending the fitted straight lines from the linear mixed effect models (including the fitted random effects).

Iteratively for each new forecast day, the number of new cases is calculated using this rate and the previous number of outstanding cases, arriving at a forecast for cumulative number of cases.

The number of forecast recoveries and deaths are estimated as the number of cases 14 days ago, as the typical number of days from symptom onset to deaths or recovery is around 18.5 to 22 days [4]. Our method parameterises 14 days to account for some time lag between onset and case confirmation.

Note that these imperfect estimates of deaths and recoveries are sufficient in this method and do not materially impact the ensuing projections. The current number of forecast outstanding cases is calculated by subtracting the forecast cumulative deaths and recoveries from the forecast cumulative case count. The iteration then continues for each new day of the forecast into the future.

Below we plot the forecast cumulative case counts for the US and Australia. For the US in Figure 3, one method (light red line) extrapolates off the mixed effect model fitted US trajectory, which uses the global average effect of interventions due to the lack of post-intervention observed data for the US at the time of the forecast (March 23). The second method (light blue lines) uses the trajectory evolution curves from substitute countries, such as Korea, China and Italy, at 25%, 50%, 75% and 100% of the substitute country's efficacy.

As mentioned earlier, this substitution effect is achieved by replacing the slope coefficients of a target country m (b_m, c_m), with the slope coefficients of a substitute country s (b_s, c_s). Any desired discount to the substitute country's efficacy is achieved by multiplying these slope coefficients with the discount factor.

For example, the light blue line from the bottom right corner shows that if the US has 100% of South Korea's efficiency in driving down the case trajectory from the date of forecast (March 23), there will be 139.6k cases eventually, assuming the interventions are not lifted. Similarly, in the top right corner plot, the light blue line shows that if the US has 100% of Italy's intervention response efficacy, the eventual case count would be around 1.2 million. Otherwise, if the US only has a globally average intervention efficacy, then the eventual case count will be around 383k. Note that dotted lines in the plot denotes projections.

We hypothesise that financial markets are likely to focus on the peak in daily new cases, viewing it as the inflection point in the case trajectory and hence the course of the pandemic in the world's largest economy. We have therefore also plotted the forecast daily new cases for the US in Figure 4.

If the US has, for example, 50% of South Korea's efficacy in driving down new cases, then the peak will be around April 3 . Conversely, if the US has, say, 75% to 100% of Italy's efficacy, then the peak will in mid-April. We have also extended this analysis to Australia (Figures 5 and 6).

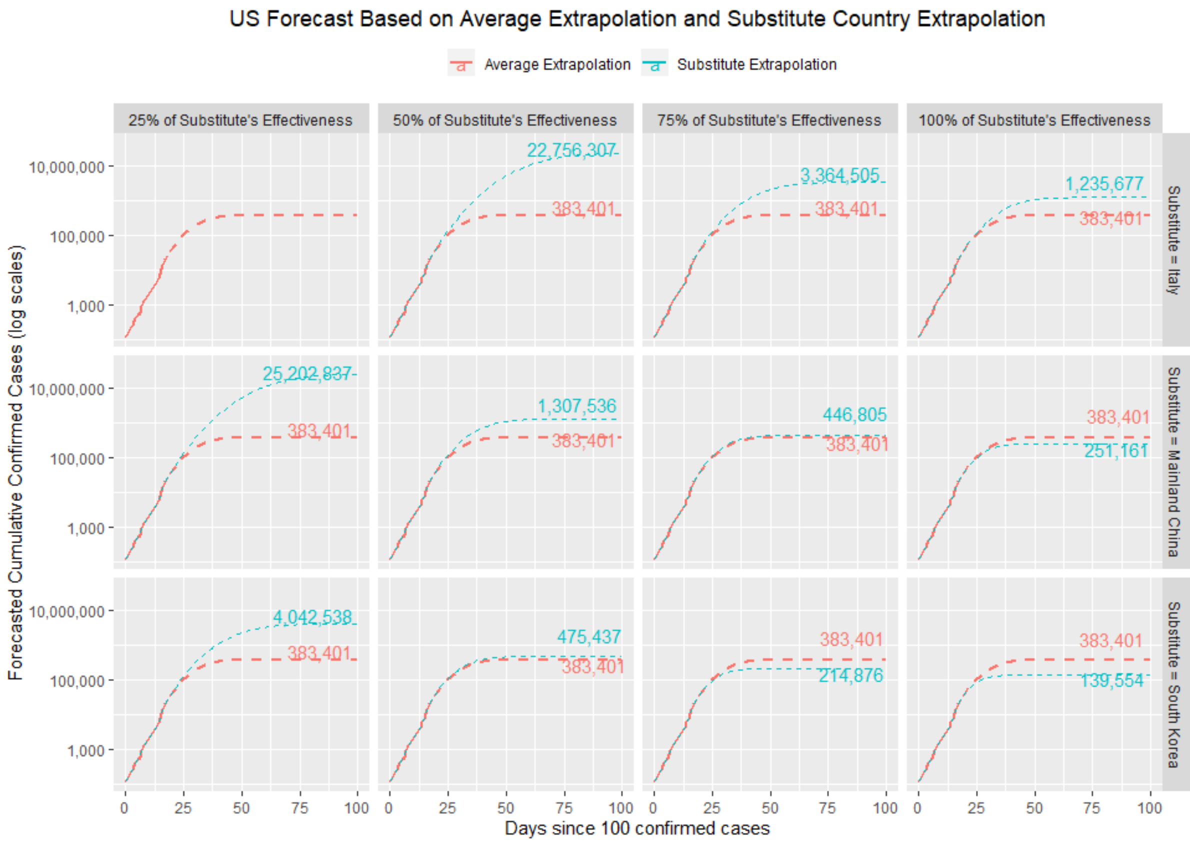

Figure 3

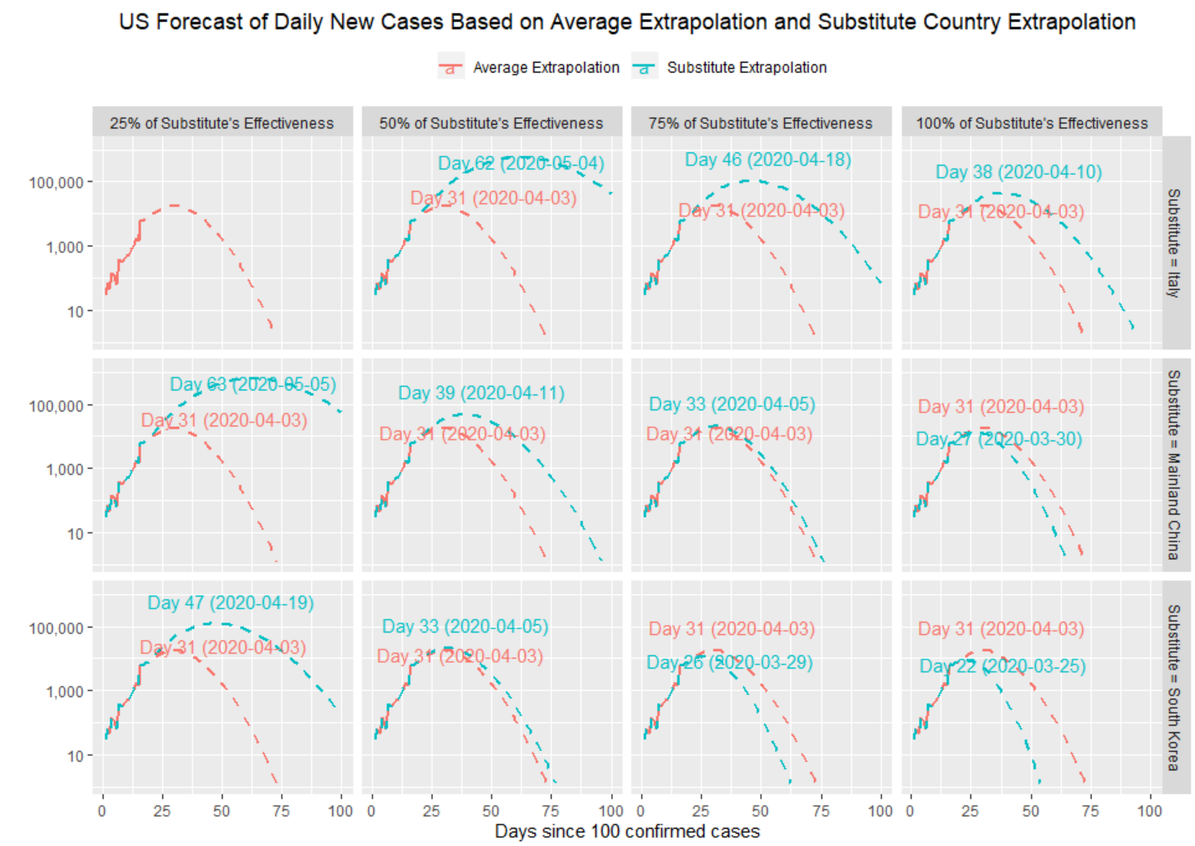

Figure 4

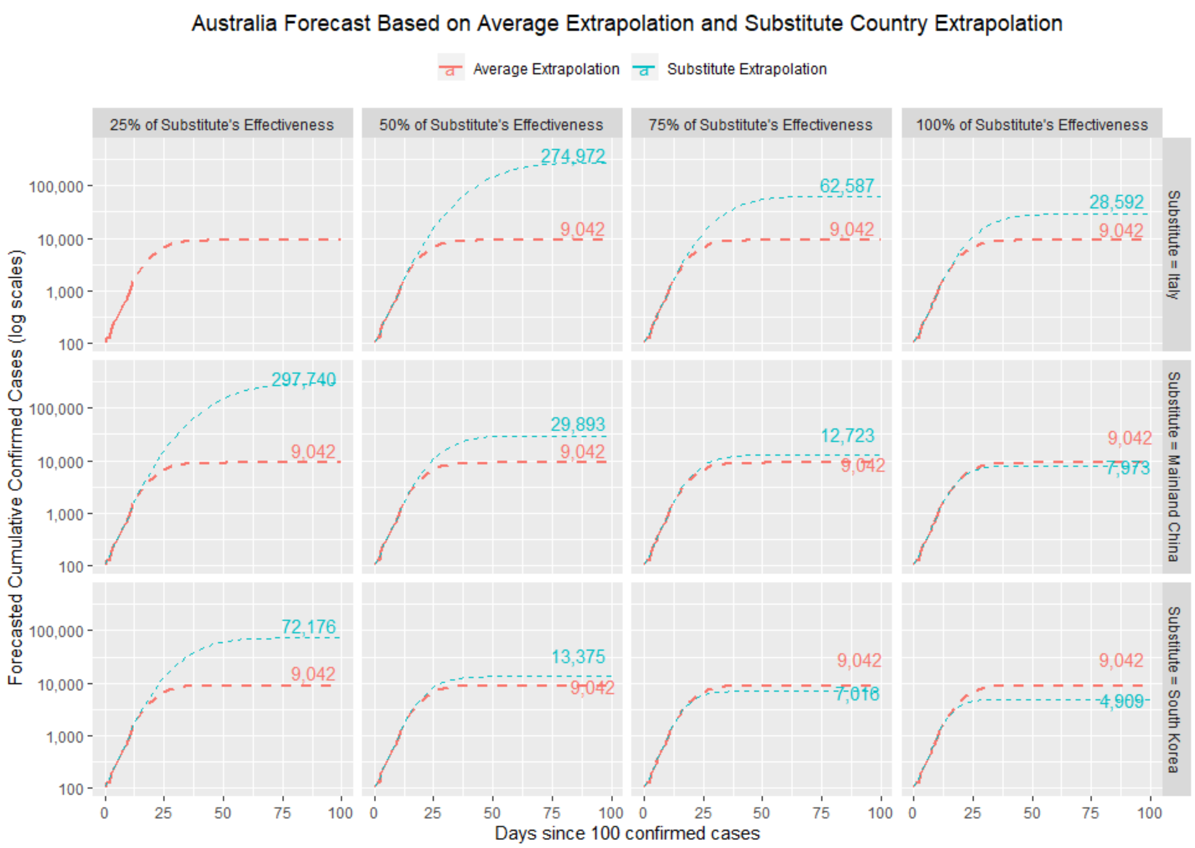

Figure 5

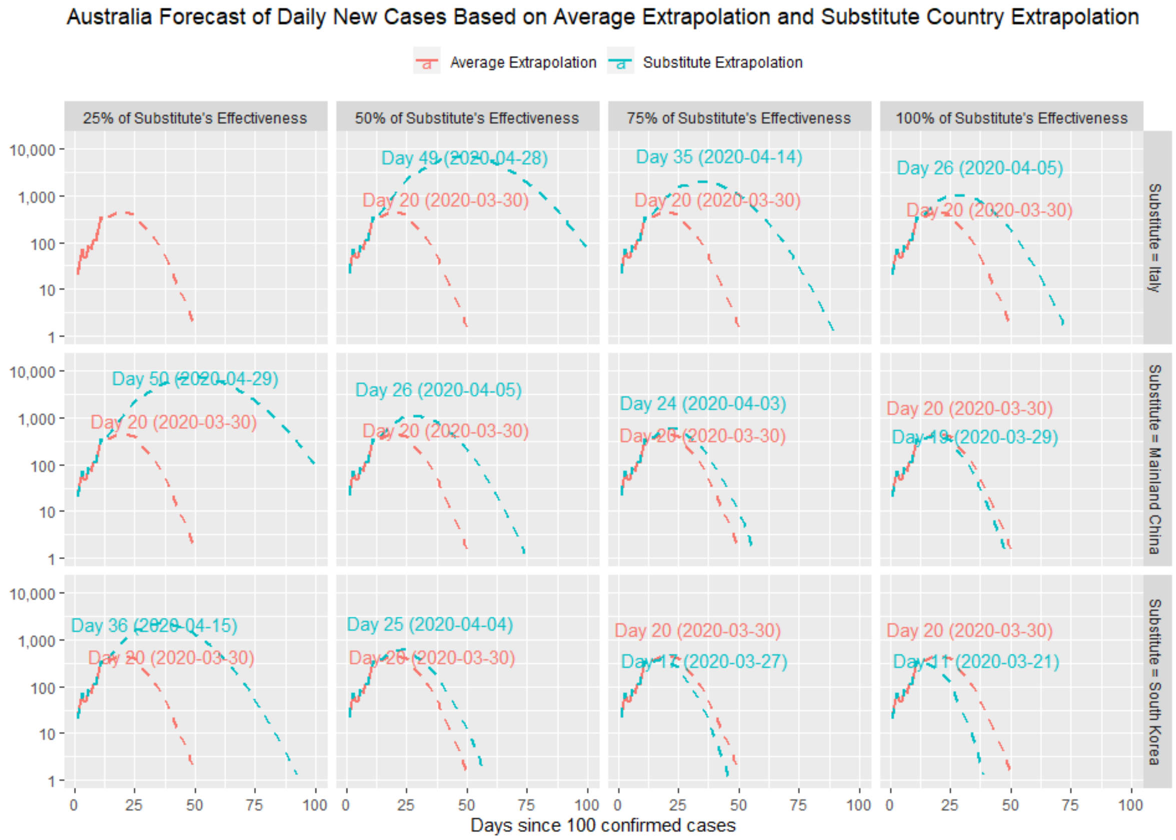

Figure 6

3. Discussion of Key Forecast Findings and Conclusion

With the virus apparently contained in China, financial market participants are likely to be especially sensitive to innovations in the infection trajectory in the world's largest and most important economy, the US.

We calibrate our forecasting framework to condition on a range of potential containment paths spanning, in a best case scenario, the US government displaying containment intensity that is parameterised at between 25% and 50% of South Korea's efficacy to, in a more gloomy case, a scenario where the US is only 50% to 75% as effective as the Italian reaction function. This implies a reasonable range for the peak in the observed number of new cases in early to mid April with a demonstrable decline in new infections evidenced thereafter.

When we apply the same methods and calibrations to other target countries, like Australia, we project that infection rates are similarly likely to peak in early to mid April with a reduction in new cases observed in the second half of the month.

This forecasting framework assumes that each country is a homogeneous entity, and that a single coefficient can be employed to model the trajectory of each nation's case growth. A potential source of inaccuracy stemming from this is that if a country, such as the US, adopts interventions a few jurisdictions at a time, as opposed to a coordinated nationwide effort, then the time to peak daily new case count could be substantially elongated.

In this scenario, the US government may be making intervention decisions based on the goal of keeping the national daily case count manageable but constant until a vaccine is ready, in which case the time to peak daily case count may be, by construction, the time to a vaccine (eg, 12 months). This would obviously be sub-optimal.

Whilst our methods are not founded on epidemiological theory, and are predicated on an empirical approach based on intuitions obtained through expert data visualisation, we believe this approach offers a versatile yet parsimonious model for predicting infectious disease trajectories.

This is especially useful when the effects of imminent non-pharmaceutical public policy intervention measures on the case trajectory of a target country of particular interest are not yet known.

In the absence of strong qualitative insights from the user, our approach also enables forecasts using an averaged global effect of policy interventions fitted via a linear mixed effect model.

In the presence of user calibrations, this framework facilities ‘what-if’ analysis, producing forecasts for target countries conditional on the intervention efficacies of other nations that may be further along their own policy response path. These calibrations can be further adjusted to reflect imperfections in the target country’s application on its own responses (eg, with inferior intensity).

Combined, these two approaches offer sophisticated financial market participants versatile and parsimonious projections in the absence of observed data on imminent intervention efficacies in specific countries.

When first applied and published on March 23, these models projected the peak in daily new COVID-19 case counts for the US and Australia would arrive in early-to-mid April 2020. To the best of our knowledge, this was one of the first early-to-mid April peak projections published globally.

These methods were coded in Coolabah Capital Investments’ data science systems to function in real-time with automated live updates that presented the projections to portfolio managers via graphical user interfaces.

References

1) Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z.U.L.M.A., Cuomo-Dannenburg, G. and Dighe, A., 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand.
2) JHU CSSE. 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. GitHub. https://github.com/CSSEGISandData/COVID-19. Continuously updated. Accessed March 23, 2020.
3) Jia, L., Li, K., Jiang, Y. and Guo, X., 2020. Prediction and analysis of Coronavirus Disease 2019. arXiv preprint arXiv:2003.05447.
4) Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X. and Guan, L., 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet.
5) Jones, K., 2020. Gompertz Growth and COVID-19. LinkedIn. https://www.linkedin.com/pulse/gompertz-growth-covid-19-kingsley-jones/. Published March 4, 2020. Accessed March 31, 2020.
6) Lin, K., Joye, C., 2020. Peak Virus: Forecasting the Peak in COVID-19 Infections in the US and Australia. https://www.livewiremarkets.com/wires/peak-virus-forecasting-the-peak-in-covid-19-infections-in-the-us-and-australia. Published March 23, 2020. Accessed March 23, 2020.