GENERALI INVESTMENTS

Focal Point

Covid-19: proprietary models set to monitor pandemic evolution

19/06/2020

Authors: Giovanni Millo, Mattia Mammarella

- In this report we present stylized facts and regularities of the spread of the virus deriving from 4 proprietary statistical models which focus on: short-term forecast of contagions and deaths, longer term Gaussian model for deaths estimate and, finally, current estimate for total real cases.
- We advocate basing cases estimates on deaths as a stable measure, less affected by testing issues than cases.
- The evolution in the pattern of contagion rates in Western countries is typical and quite predictable: an initial outbreak with very high contagion rates followed by linear moderation as habits change and containment measures step in.
- Some Eastern countries which succeeded in containing the virus have patterns that either skip the initial acute phase (Japan) or contain it very rapidly (Korea), leading to dramatically lower total cases and lesser strain on the Health system.
- Gaussian models of the epidemic's evolution predict the peaks well (and hence the maximum strain on intensive care) but the dying out phase is slower than the bell-shape curves would predict.
- The actual number of infected people is a multiple of the official figure, and can be roughly estimated at around 100 times the official deaths.
- The typical trajectory of an eventual new outbreak (second wave) would probably skip the initial acute phase due to the improved degree of cultural and material preparation, looking more like the Japanese case than what was experienced by Western countries.

Introduction

At the end of 2020, a new coronavirus started spreading in China. A few months later, it was already all around the world, changing the life and habits of billions of people. Soon after the covid-19 started spreading, a large number of models and reports were published, fed by the need of understanding what was happening. Both from a sanitary and from an economic point of view, it is key to understand the spreading dynamics of the virus and to forecast the evolution of the contagion around the world.

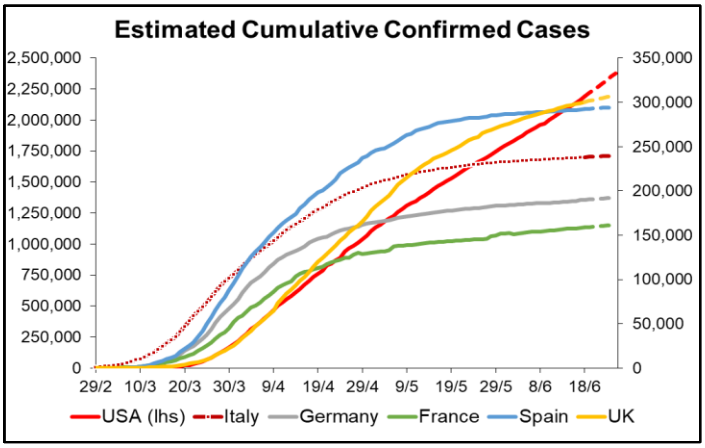
In this paper, we present **four of the coronavirus-related models** we developed: two short-term forecasting model of **infection rates**, one for cases and one for **deaths**, a forecasting model to estimate the **peak and the evolution of deaths**, and a model to estimate the **real number of cases**, which are generally underestimated.

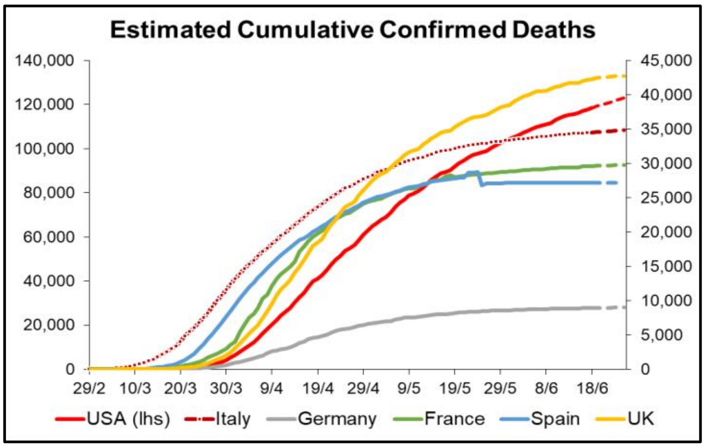
Generally, the relevant measure we use in the models is the death count. The number of deaths is in all likelihood the most precisely reported among the quantities of interest, the infected cases being so heavily dependent on the testing strategy and overall testing possibilities of a country or region. Despite widespread rumors of severe underreporting in some particular instances (Italian province of Bergamo, to say one), one can be confident that most of the Covid-19 deaths is reported as such in most countries and regions. That said, we must remember that not all countries report covid-19 deaths in the same way: in some cases only deaths registered in hospitals are considered and those resulting from the worsening of pre-existing conditions are excluded.

Short-term forecasting of infection and death rates

The basic infection rate of a population, in absence of previous infection and containment measures, is such as to lead to explosive growth. We can already see the effect of the policies in force even in some countries where the outbreak is still of exponential nature. We will endeavour estimating an **exponential model of contagion** (on total official cases and deaths) in order to measure the average infection rate in each major country; then we will estimate on a rolling window to assess the change, if any, in the infection rate brought up by different measures.

The ultimate goal of the exercise is to project the near-term evolution of the disease, therefore projections will have to be based on the most recent estimates of the infection ratio rather than on an average value over the whole story, which is not a valid representation of the true velocity of the contagion anymore. The transmission rate, and therefore the strength of exponential growth, is generally not constant even if the exponential model as a *functional form* still holds. Recursive estimation will provide a description of how the transmission rate changes along the history of the infection.





The typical evolution of a **successfully resolved epidemic** can be seen in the full story of **China** (and Wuhan in particular, which counts for the vast majority of cases): there are very high transmission rates at the peak, then a sharp reduction thereof as important measures are undertaken (possibly with a lag due to the time distance between actual infections and observation: estimated at 12 dd. in the case of China); then infection rates linearly decreasing and approaching zero as containment and social distancing measures work and the illness is won.

Korea, another success story, has a peculiar pattern: building on China's experience, they have contained contagion almost immediately (witness infection rates below 0.1 from the very beginning). Then the action of a single super spreader has provoked thousands of cases in the cities of Daegu and Seul, with a soaring to almost 0.5, the highest value estimated in this exercise, only to be reduced again to practically nil in the next two weeks.

The timeline of infection rates in most **Western countries** shows a **different, less favourable common evolution** pattern. After an initial explosion in contagion rates, which can reach near or over 0.4, a linear descent becomes, which takes the rate towards zero (in the best case scenarios, like **Switzerland**) or anyway towards very modest values (**Italy, Germany, France**); or in the worst cases towards moderate but still important values (**USA, UK**).

Countries where the application of containment measures has been slow, partial or erratic, like the **USA** and the **UK**, in fact show **slower convergences towards zero of the contagion rate**, and often a flattening on a nonzero asymptote. Nevertheless, the main tendency is the same as for countries where action was much firmer. We attribute this to the effect of information and awareness on the public's part, and the consequent modification in habits, even in the absence of a firm government stance.

Estimating the real total infected

In principle, a function describing the evolution of the spread should be based on true, not on official cases. This for two reasons: 1) official cases come with a lag (every case has to be detected, tested and reported) 2) not all infectious cases are actually reported.

The spread of the virus obviously depends on true, not on official cases. **The lag between true new cases and official new cases is estimated in (about) 12 days** based on a) medical evidence and b) statistical data on the evolution in China/Wuhan.

Still, it can make sense to project detected (official) cases in the end. One just has to remember that **exogenous interventions**, like the containment measures, will have effect with a given lag. Current cases can be guessed, as a first approximation, by applying the lag between true and official cases as estimated on Chinese data (i.e. pulling the new cases data back 12 days) but this method cannot provide an estimate of actual, not reported cases.

Nowcasting actual cases

Reported cases are probably a fraction of the actual total. What is worse, the share is believed to vary wildly across countries, from those who tested extensively and even randomly (Korea) to those where only the seriously ill get tested (Italy, up to a point).

Random testing is the only way to gauge the actual share of the population that has been infected. Unfortunately, the tested population is almost invariably heavily selected.

The most dependable measure are in all likelihood the fatality counts, although these are known to have been underestimated as the situation of the health system worsened.

In order **to estimate the actual number of infected**, it is therefore natural to **take the dead as the tip of the iceberg** and from this manifestation work out the **latent variable**: the number of cases at the time of contagion (which can be taken to be about 20 dd. back).

As in **Pueyo** (2020), observed deaths - perhaps the most precise of available measures - can be used for assessing the number of current true cases, reported and unreported, given a number of parameters. Some of these depend on the **nature of the disease** (time to death, incubation etc.) others depend on the **country**, i.e. on the **density** of social groups, the **health system**, **fatality rate**, doubling time, etc.

To move from deaths to (current) true cases in one day t take the current mortality rate μ, suppose 1%. This means 1 death in t indicates 100 infected *at the time of infection*. Now we need to calculate the evolution from 100 infected back then when the dead person got infected until today. We need the time to death τ (estimated at 17.3) and then the doubling time (the time it takes for cases to double) δ estimated at 6.2 (5.1, 8.2). The number of true cases today for each death today would then be

$$\frac{y_t}{d_t} = \mu 2^{\frac{\tau}{\delta}}$$

hence $y_t = d_t \mu 2^{\frac{\tau}{\delta}}$. This procedure has the benefit of estimating the current true cases, including asymptomatic spreaders.

This reasoning depends crucially on the type of spread. If k dead are from a cluster, they can count as one dead so that the above count will be underestimated (see the example of Washington State in Pueyo, 2000); if on the converse there are super-spreaders, the count will be underestimated. This puts large uncertainty on the calculation. Nevertheless, these peculiarities are probably going to cancel out in the general population.

Estimating the true fatality rates

The fatality rates of Covid-19 reported across the world have been the most diverse, peaking at over 13% in Italy, while the general consensus between epidemiologists is that it be one order of magnitude lower (nearly one tenth of the observed rates), even adjusting for the age structure of the population and for other risk factors like pollution and concurrence of other lung diseases.

Even over homogeneous populations, **obtaining a consistent estimate of mortality** (deaths over the entire population) **and fatality** (deaths over the infected population) rates obviously **depends on the sample**: notoriously, during the Italian outbreak the vast majority of the people who got tested were seriously ill, to the point of showing strong symptoms or even of needing intensive care.

Ideally, one would want to do a random trial on, respectively, the entire population or a randomly chosen population of infected, without the milder cases going unnoticed. What is perhaps the closest to an experiment of this kind happened when the Covid-19 spread on the cruise ship Diamond Princess (see here). The ship contained 3,711 passengers and crew, of which 700 got infected. While the infection rate got magnified by the close confinement conditions, the sample of infected people was reasonably close to randomness (despite some prevalence of older people) and, especially, everybody on board was tested.

Is the dead count truly a dependable measure?

There were 7 dead in the cruise ship (1% of cases), a number too low for dependable estimation. That said, when researchers used the methodology, i.e. the share of people with mild symptoms, or even asymptomatic (18%), to estimate the true number of infected in the Wuhan outbreak, they come up with a similar **"true" fatality rate** of 1.1%, against the official estimate of 3.8%.

While in "normal" conditions, for the reasons given above, one can expect "most" of Covid-related deaths to be reported, experience taught us that **there will always be some degree of underestimation**. This effect will typically be **non-linear**, as unreported Covid-19 deaths will begin to occur massively during the peak of the emergency, as (if) the health system is overwhelmed and resources are not enough to test all the dead for Covid antibodies: especially those who do not even get to the hospital and die at home.

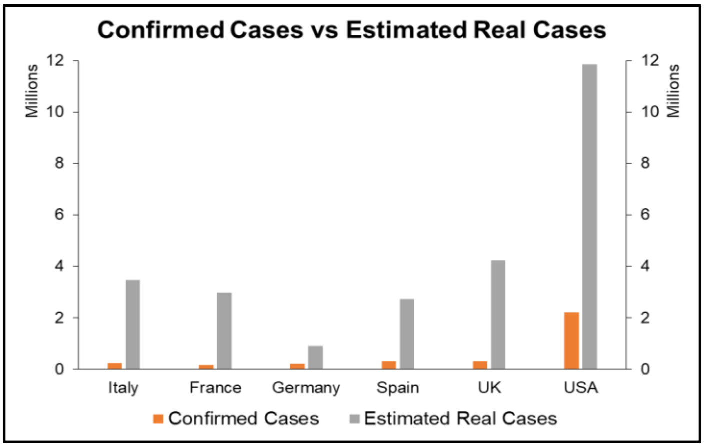
The most popular way to assess the actual Covid-related mortality has been to compare reported Covid-19 deaths to the increase observed over the "normal" (average) mortality of the relevant region. According to this method, for example, in the Italian province of Bergamo the statistical **"extra mortality"** has been more than double the reported number of official Covid-19 deaths.

According to recent coverage by FT, the size of the underreporting might well be as high as 60%.

Based on this line of reasoning, one could **estimate the order of magnitude of the total infected** in different countries taking the total number of dead, perhaps allowing for an underreporting rate of between 10 and 60%; **assuming a case fatality rate of about 1.5% in the Western world** -- where the share of the elderly and of the people affected by other risk-enhancing pathologies (obesity, diabetes, asthma) is higher – **and 0.5 to 1% in emerging countries and developed South-East Asia**.

From this gauge the true number of infected, symptomatic and asymptomatic, gross of the recovered, as being around **100 times the dead in the West, and between 150 and 200 times in Africa and SEE**.

In any case, we can estimate the actual infected, current and past, to be two orders of magnitude higher than the dead: **which would take the count, for Italy, to about 3.5 million; or, for the USA only, to the impressive figure of around 12 million** people who are, or have been, carrying the virus. Too many to be safe, too few to grant any kind of herd immunity.



Estimating the peak

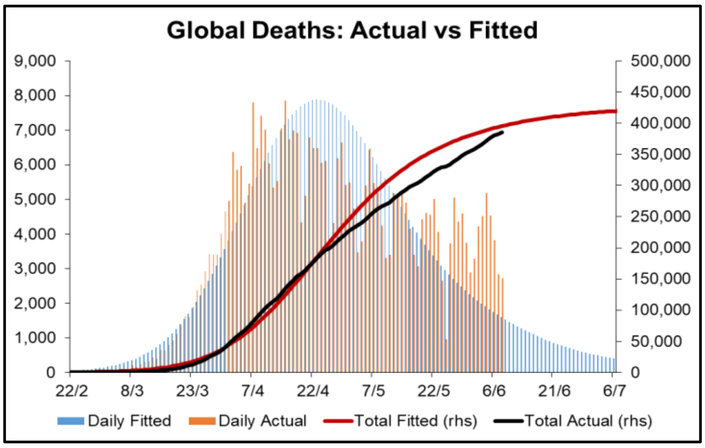
Gaussian curves have often been estimated as a theory-free, statistical approximation to the full development history of an epidemic, if the explained variable is the daily number of cases. The Gaussian curve relates to the usual, sigmoidal-shaped evolution of total recorded cases, as the latter is the cumulant of the former, so that one maps to the other.

The recent Covid epidemic has seen many attempts at fitting the sigmoid-shaped curves describing -- and predicting -- total cases. If the main interest lies in estimating the peak time of the infection, though, fitting the top of a Gaussian curve is more reliable than trying to identify the flex in a sigmoid; moreover, it directly returns many quantities of interest. According to Schüttler et al. (2020), **the Gaussian is described by three parameters**, two of which have an interesting direct interpretation: **the maximum daily cases at**

peak, the peak time, and a width parameter. From these both the maximum number of cases at the end of the epidemic and, by symmetry, the time of the end (first day with no cases) can be obtained.

Again, following Schüttler et al. (2020), we take the number of deaths as the target variable. The number of deaths is in all likelihood the most precisely reported among the quantities of interest, the infected cases being so heavily dependent on the testing strategy and overall testing possibilities of a country or region. Despite widespread rumours of severe underreporting in some particular instances (Italian province of Bergamo, to say one), one can be confident that most of the Covid-19 deaths be reported as such in most countries and regions. In turn, the number of deaths has been taken as a robust (backwards) predictor of the total number of infected, so that, besides being important in its own right, projecting the evolution of deaths in time can give a good estimate of the evolution in total cases, and hence of the number of active cases needing intensive care and such variables of interest to policymakers.

We have applied the indirect Gaussian fit described above to daily Covid-19 deaths across all countries having at least 500 fatalities to date. In-sample it performed well in most cases. Of course, its predictive performance as regards the future evolution depends heavily on the symmetry assumption, by which the epidemic, country by country, is supposed to decrease with the same speed with which it spread at the beginning.



From the initial off-peak trajectories of the many countries which already passed the maximum of Covid-19 daily deaths, this assumption seems to be questioned by the relatively slow speed of descent: in fact, and unfortunately, **the rightmost part of the bell curve for many countries is proving less steep than the model would predict**. As a preliminary conclusion, therefore, the Gaussian model looks better suited to predicting the peak time (and number of infected!) than the time the epidemic will effectively end.

Conclusions

Despite the novelty of the disease, next to the medical experience, there are **lessons to be learned from statistical regularities in the recent Covid-19 outbreak**. The **evolution in the pattern of contagion rates in Western countries has been typical**, the very high contagion rates of the first weeks moderating linearly under the influence of containment measures by Governments on one hand, and habit change in the population on the other as prevention culture spreads, spontaneous social distancing takes place and public hygiene improves.

Combining statistical evidence from different geographies and settings, it is also possible to **estimate the actual number of infected people**. This is very likely to be a multiple of the official figure; a well-founded rule of thumb for estimating it **is 100 times the official deaths**, which takes the figure to many millions in several of the major countries.

What does this teach us for the future, given that the "permanent" solution under form of a vaccine, although clearly in the offing, is unlikely to be available, tested and widespread before the Fall; and that containment measures are working well in most countries, but are not going to eradicate the disease in the presence of such a huge number of active cases worldwide?

There are **substantial chances that another outbreak happens in the next Fall**, after the mitigating effects of the warm season have died out, and people start spending less time in the open. **The typical trajectory of the initial outbreak is not very likely to repeat itself**, anyway. The initial phase of an eventual second wave would probably be milder due to the improved degree of cultural and material preparation by the population, the daily centralized control state-regions and the much improved management of the infected by hospitals.

References

Schüttler J, Schlickeiser R, Schlickeiser F, Kröger M (2020). Covid-19 predictions using a Gauss model, based on data from April 2. MedRxiv. doi:10.1101/2020.04.06.20055830. MedRxiv/2020/055830.

Pueyo T, (2020). Coronavirus: Why you must act now. URL: https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca

Imprint

Head of Research Vincent Chaigneau (vincent.chaigneau@generali-invest.com)

Head of Macro & Market Research: Dr. Thomas Hempell, CFA (thomas.hempell@generali-invest.com)

Team:
Elisabeth Assmuth(elisabeth.assmuth@generali-invest.com)
Elisa Belgacem (elisa.belgacem@generali-invest.com)
Radomír Jáč (radomir.jac@generali.com)
Jakub Krátký (jakub.kratky@generali.com)
Michele Morganti (michele.morganti@generali-invest.com)
Vladimir Oleinikov, CFA (vladimir.oleinikov@generali-invest.com)
Dr. Martin Pohl (martin.pohl@generali.com)
Dr. Thorsten Runde (thorsten.runde@generali-invest.com)
Dr. Christoph Siepmann (christoph.siepmann@generali-invest.com)
Dr. Florian Späte, CIIA (florian.spaete@generali-invest.com)
Dr. Martin Wolburg, CIIA (martin.wolburg@generali-invest.com)
Paolo Zanghieri, PhD (paolo.zanghieri@generali.com)

Head of Insurance and AM Research: Michele Morganti (michele.morganti@generali-invest.com)

Team:
Raffaella Bagata (raffaella.bagata@generali.com)
Alberto Cybo-Ottone, PhD (alberto.cybo@generali.com)
Mattia Mammarella (mattia.mammarella@generali-invest.com)
Roberto Menegato (roberto.menegato@generali.com)
Giovanni Millo, PhD (giovanni.millo@generali.com)
Antonio Salera, PhD (antonio.salera@generali.com)
Cristiana Settimo (cristiana.settimo@generali.com)
Federica Tartara, CFA (federica.tartara@generali.com)

Issued by: Generali Insurance Asset Management S.p.A., Research Department

In Italy:
Generali Insurance Asset Management
S.p.A Società di gestione del risparmio

Piazza Tre Torri
20145 Milano MI, Italy

Via Niccolò Machiavelli, 4
34132 Trieste TS, Italy

In France:
Generali Insurance Asset Management
S.p.A Società di gestione del risparmio

2, Rue Pillet-Will
75009 Paris Cedex 09, France

In Germany:
Generali Insurance Asset Management
S.p.A. Società di gestione del risparmio

Tunisstraße 19-23
50667 Cologne, Germany

www.generali-investments.com