Exploring COVID-19

Nils Bertschinger

March 28, 2020

> The world stands still ... desperately observing the unfolding of the global COVID-19 pandemic.

Data exploration

The John Hopkins university and other institutes publish daily numbers of cases and death tolls. Here, we build on their data sets and provide some simple explorations and modeling.

Country comparison

Fig. 1 shows the raw data for several countries [1].

[1] Here, we only consider these countries in the following

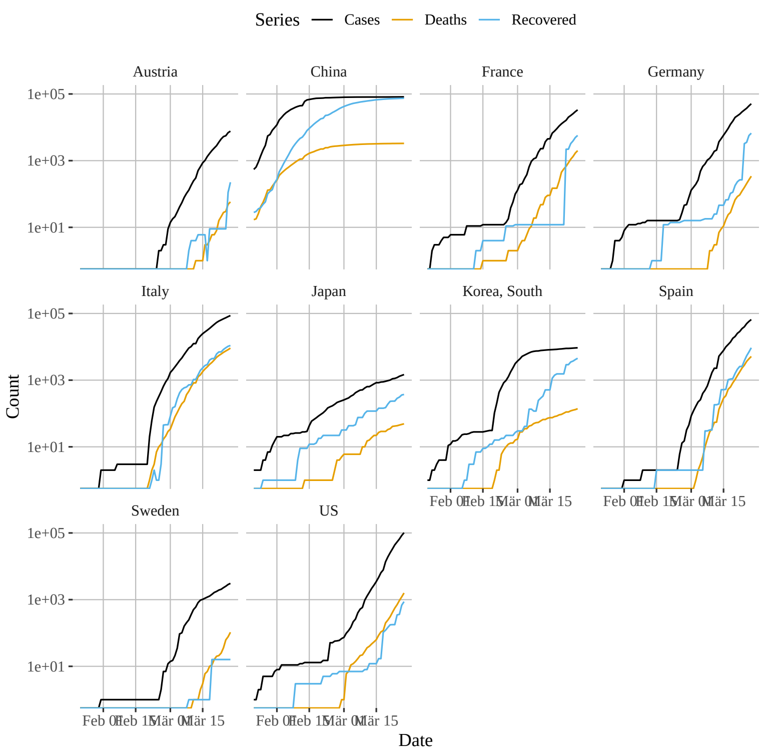

Figure 1: Data as provided by the John Hopkins university for some selected countries.

As the beginning of the epidemics is different in different countries a direct comparison is difficult. Furthermore, especially the count of cases is highly debated and plaqued with several uncertainties. Here, we assume that the *death counts are reliable* and essentially correct. Thus, in order to compare different countries we align all curves such that day 0 corresponds to the first day that the death count reaches a specified threshold (either absolute or relative per million inhabitants).

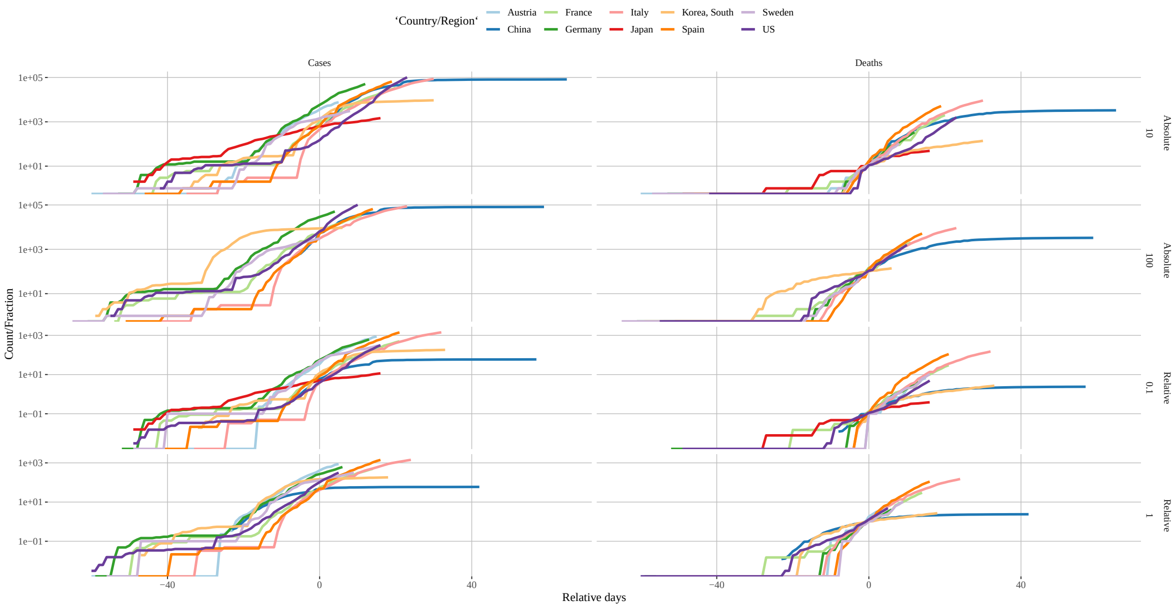

Figure 2: Case and death counts aligned to first day of more than a specified threshold (either absolute or relative per million inhabitants).

Fig. 2 shows a clear data collapse, especially at higher thresholds[II] – which are less noisy – of 100 (absolute) or 1 per million (relative). Furthermore, it is evident that the current growth rate of China, South Korea (which are very similar in relative terms) and Japan is markedly slower than of the other countries.

[II] Note that some countries might not have reached these higher thresholds and are therefore not included in every subplot.

Interestingly, also the case counts are somewhat aligned even though day zero has been defined purely based on the death counts. Furthermore, there appear to be two groups of countries with different systematic delays between case and death counts. In particular, Austria and Germany seem to report deaths consistently later than France, Italy and Spain. This suggests that the surprisingly low death rate reported for Austria and Germany could be an artefact as reported numbers are simply some days older compared to other countries! This *delay effect* makes comparing numbers from different countries difficult and also leads to unreliable estimates when naively comparing numbers from same days only. Similarly, judging the effectiveness of containment measures, e.g. social distances, requires time as well within the second week after the intervention has been established a majority of observed cases had probably been infected already before the intervention.

Statistical modeling

Phenomenological growth model Overall, I believe it unlikely that the death rate is very different across different countries[III]. Next, I build a phenomenological model on the assumption that the *death rate is constant across all countries* and differences purely arise from delays in reporting positively tested cases and deaths. The model assumes the following:

[III] There are certainly demographic, medical or other aspects though.

- The probability of death is the same for all countries whereas the testing prevalence is country specific.

- Observed counts are negative binomial distributed – as an over-dispersed Poisson – and delayed wrt the actual cases.
- Actual case and death counts in each country grow according to a sigmoid function[IV] with country specific parameters[V].

As shown in Fig. 3, this model[VI] indeed finds a consistent difference in the delay of death counts between Austria, Germany and France, Italy, Spain. Furthermore, the death rate – assumed constant across all countries – is estimated as $4.8 \pm {}^{1.2}_{1.2}\%$. Estimates in the range of 3 to 5% appear to be rather robust for the present model, yet seem to be somewhat too high[VII]. It remains to be seen if my model estimate holds up over time and wrt estimates derived from more realistic models ...

Detailed model predictions for all considered countries are shown in Fig. 4. The predictions are mostly reasonable, but the model has difficulty of matching the rapid leveling off observed in China and South Korea. Interestingly, the model predicts that the curve has already slowed markedly in Germany and Italy even though this is barely visible in the raw numbers by now – another example of why the delay effect is important in understanding the dynamics of the COVID-19 pandemic. Yet, this model prediction relies heavily on the assumption of sigmoidal growth and I would not be too optimistic about it.

[IV] A more realistic model should build on epidemic dynamics such as SEIR ... see below.

[V] The basic model assumes

$$c(t) = a(1 + e^{-\beta\nu(t-\tau)})^{-\frac{1}{\nu}} ,$$

i.e. a generalized logistic function.

[VI] Full code of this and other models can be found at my accompanying repository.

[VII] Note that a naive estimation of the death rate, i.e. dividing contemporal case by death counts is biased downwards by the delay effect as actual deaths only realize about a week later. Accordingly the death rate estimate should be based on the substantially lower case counts a week ago.

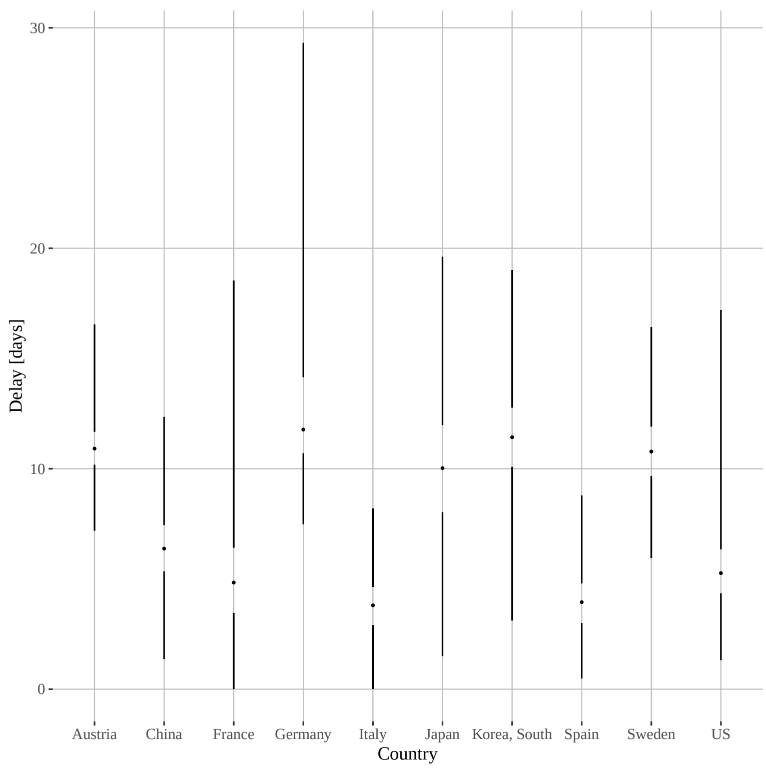

Figure 3: Model estimated delay between reported case and death counts.

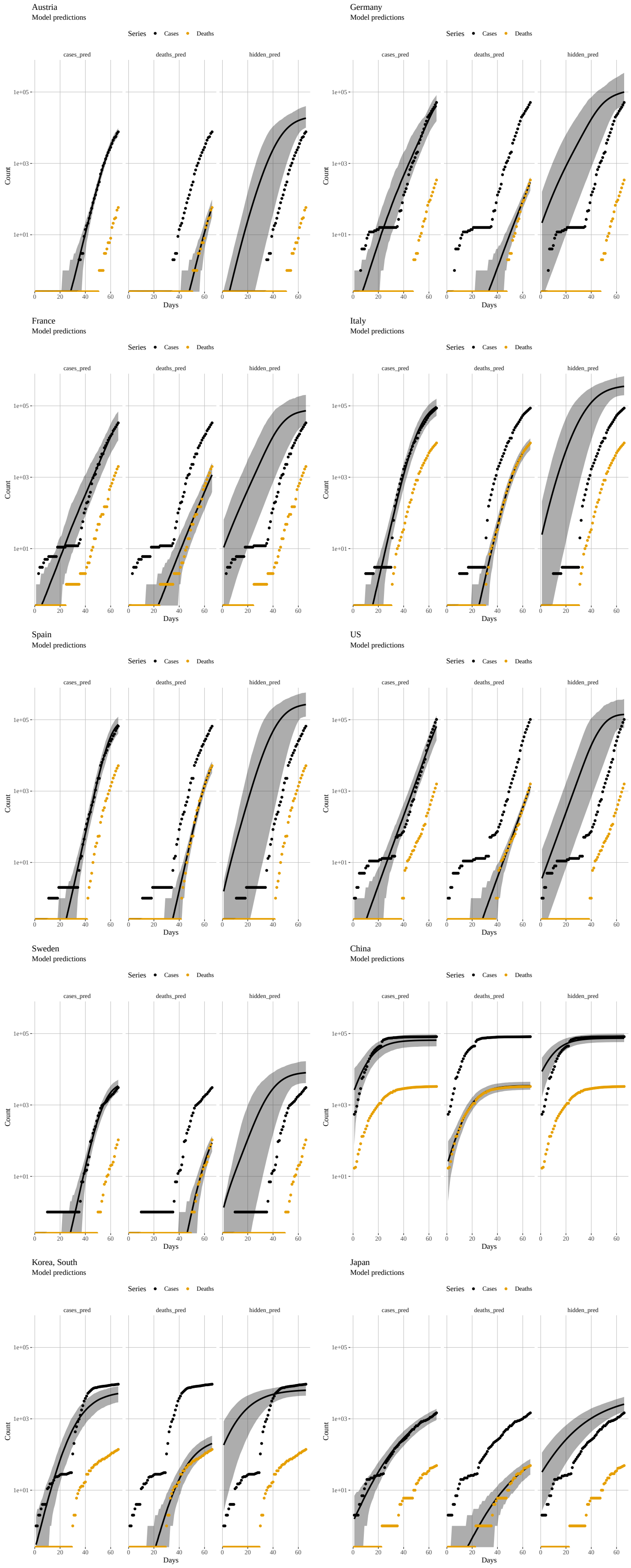

Figure 4: Model predictions (mean and 95% credible region) of actual (hidden) and observed case and death counts for different countries. Note the log scale on the vertical axis.

Epidemic model The basic SIR model, assumes that an infection unfolds when susceptible (S) individuals become infected (I) – which in turn infect further susceptible individuals. Finally, infected individuals recover (R) (or die) and are no longer susceptible. In continuous time, the dynamics can be described by the following system of ordinary differential equations (ODEs):

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{I_t}{N} S_t \\ \frac{dI}{dt} &= \beta \frac{I_t}{N} S_t - \gamma I_t \\ \frac{dR}{dt} &= \gamma I_t\end{aligned}$$

where $N \equiv S_t + I_t + R_t$ is constant over time. Model parameters are

- the infectivity β
- and the recovery rate γ.

In this model, the average time of infection is γ^{-1} giving rise to a *base reproduction number* of $R_0 = \beta\gamma^{-1}$.

Build and discuss model with delay effect …

Financial markets

Especially the containment measures implemented in many countries around the world, have major economic consequences. Non-essential productive activity has come to a hold and financial markets around the world tanked. Furthermore, uncertainty is high and the volatility index VIX has reached levels as during the financial crisis of 2007/8.

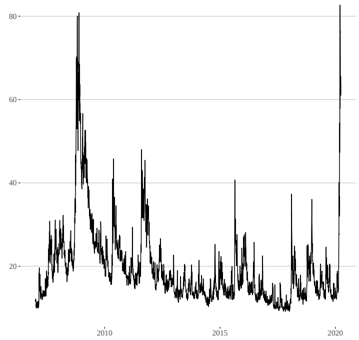

Figure 5: Closing values of VIX.

Implied risk-neutral densities

To get an impression of the forward-looking market outlook, I investigate call and put options on the S&P 500 index. In particular, out- or slightly in-the-money options are actively traded and also utilized in the computation of the VIX. Fig. 6 shows the corresponding prices of call and put options for different expiration dates.

By risk-neutral pricing, the current value of an option on the underlying S_t paying $v(S_T)$ at maturity T is given as

$$p_t = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)} v(S_T)] \, ,$$

where the expectation is taken over the risk-neutral measure $\mathbb{Q}$. In particular, assuming a log-normal risk-neutral distribution $q(s_T)$ the well known Black-Scholes-Merton (BSM) formulas provides the analytic price of European call and put options depending on the risk-neutral interest rate r, maturity T, strike price K, spot price S_t and volatility σ. Unfortunately, the BSM has several short-comings. Especially the assumption of constant volatility does not hold in actual option prices giving rise to the famous volatility-smile.

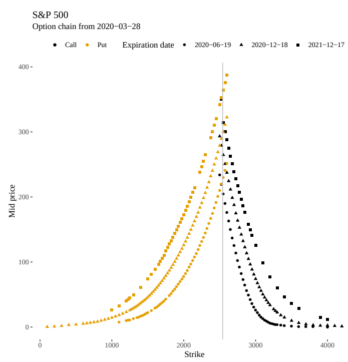

Figure 6: Mid prices of call and put options on the S&P 500. Only out- or slightly in-the-money options with positive bid price are shown. The grey line marks the current spot price.

Nevertheless, the model is easily extended towards mixtures of log-normal distributions. Then, by linearity of expectation values the theoretical price is simply given as a weighted sum of BSM prices for the different mixtures components. Here, we fit the above option prices with a mixture of two components[VIII].

[VIII] Preliminary results show that more components only marginally improve the fit.

Compared to a single component, the two component model is substantially better and also nicely interpretable. In particular, Fig. 7 shows the implied components of the risk neutral density. These can readily be interpreted as a good and bad market outlook.

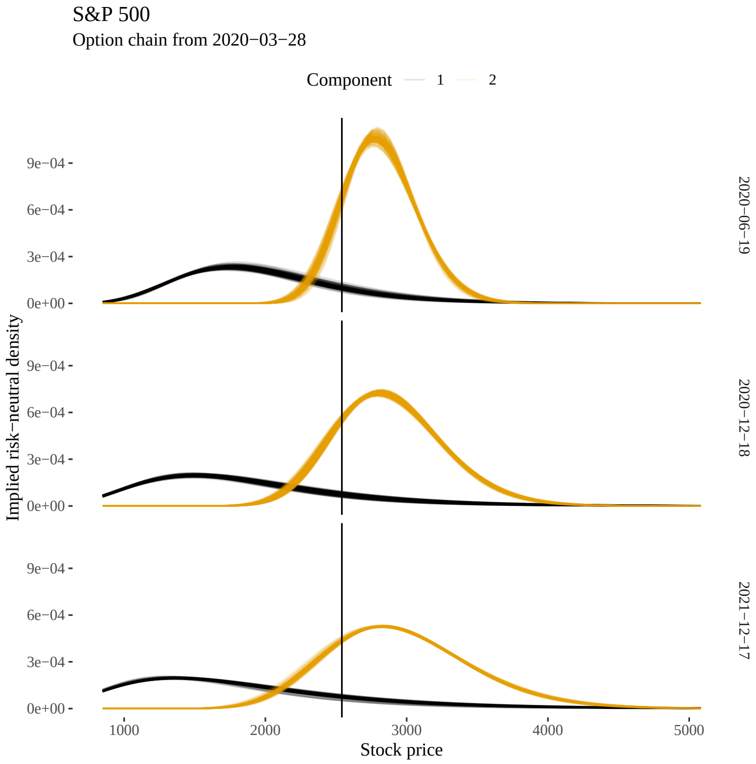

Figure 7: Components of fitted implied risk-neutral density. Shown are 100 draws from the posterior distribution.

Indeed, Fig. 8 shows the implied crash probabilities, i.e. weight assigned to the bad market component. These are surprisingly high and even rise over longer maturities. In this sense, option markets already price potentially long and large economic distortions. Timely updates of this analysis will be provided and hopefully the outlook will become more optimistic any time soon . . .

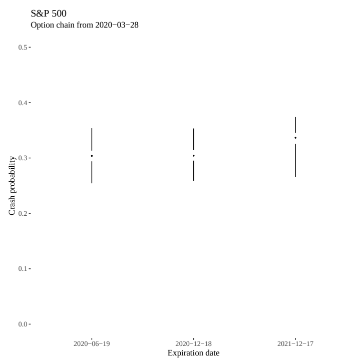

Figure 8: Market implied crash probability derived from two component mixture model.