A SIR Model with Time-Varying Parameters

Gustavo M. de Athayde[1] Ruben M. Damião[2]

Abstract: Making use of a State-Space framework, we present a generalization of the SIR-D models, where its parameters (mortality, contamination and recovery rates) evolve with time. The model has captured very well the effect of the lockdown on contamination, and the evolution of the mortality rate. It also provides us with forecast (with confidence intervals) the future values of these rates, as well as the total number of deaths, peaks of future waves, etc. Estimates of the percentage of the population divided into Susceptible, Infected and Recovered are also developed, as well as their predictions.[3]

Key words: SIR model, State-Space models, Kalman Filter, contamination, mortality, forecast.

1) Introduction

The SIR Model (Karnack and McKendrick 1927), and its extension that deals with deaths - called SIRD are models that divide the population (assumed to be constant) into compartments, in which people keep migrating through time. In essence, the population is divided in time into susceptible(S_t), infected(I_t), recovered (R_t) and dead (D_t). The dynamics occur in a deterministic way. Its discretized version is described as follows:

$$S_t = S_{t-1} - \beta S_{t-1} I_{t-1}$$

$$I_t = I_{t-1} + \beta S_{t-1} I_{t-1} - \gamma I_{t-1}$$

$$R_t = R_{t-1} + \gamma(1-\mu) I_{t-1}$$

$$D_t = D_{t-1} + \gamma \mu I_{t-1}$$

[1] Insper (Instituto de Ensino e pesquisa)
EESP-FGV (Escola de Economia de São Paulo)
athayde.gustavo@gmail.com

[2] rmdamiao@gmail.com

[3] we are indebted to many people for helpful comments and encouraging support, specially Airlane Alencar, Carolina Athayde Mendonça, Lie Pablo Grala Pinto, Berç Rustem, Renato G. Flôres, Izabel Athayde S.C. Salgado and Marco Aurelio Sáfadi.

In the original framework, there was no distinction between deaths and recovered, they were both merged into a larger group called "removed". So we had only the contamination rate β and the remove rate γ. By adding the mortality rate μ, we are able to divide the fraction that died and the other that recovered (and by assumption is now immune).

Our purpose is to allow these three parameters to evolve in time, attributing them stochastic processes. We expected that in the case of Covid-19 these parameters might have changed significantly through time. For instance, incorporating the habit of wearing masks and social restriction should affect β, advances in treatments should impact γ and μ, and a health system collapse should alter μ considerably.

The technique is called the State-Space models, and the tool to approach it is called the Kalman Filter - see Kalman 1960 and Kalman and Bucy (1961). The filter was developed for linear systems, and a few extensions have been made to deal with non-linear systems, as the one we will be facing here. We have chosen to use the extended Kalman Filter (EKF), which basically deals with a linearized approximation via Taylor series of the original non-linear system.

The use of State-Space models in the SIR models had been done before. However, the typical approach is quite different from the one we are proposing in this paper. As already mentioned, we allow the parameters $\langle \beta_t; \gamma_t; \mu_t \rangle$ to evolve with time. The former works did not allow that. They hey have chosen to model solely S, I and R as state variables (non-observed), so there would be a noise in measuring these variables. Unfortunately since the parameters are fixed, and the model is focused on simply capturing (and cleaning) the noises of S, I, R, this representation did not provide substantial gains in terms adding flexibility to the model. Examples of these approaches can be found in Dukic, Lopes and Polson (2012), Osthusk et all (2013) regarding Flus and more recently relating to Covid-19 we have Wang et all (2020) and Kobayashi et all (2020).

Section 2 describes the theoretical model, and some pragmatic suggestions to be applied when putting the model into practice. Section 3 shows the results for the City of São Paulo from March 16th 2020 to October 1st 2020, its predictions, and out of sample performance of the models. Section 4 concludes the work

2) A Theoretical Model

In this section we will present the theoretical framework of the model. All the practical results will be shown in Section 3.

2.1) The State-Space Representation

The SIR-D model with time-varying parameters will be described in the following manner:

$$S_t = S_{t-1} - \beta_t S_{t-1} I_{t-1} \quad (1)$$

$$I_t = I_{t-1} + \beta_t S_{t-1} I_{t-1} - \gamma_{t-1} I_{t-1} \quad (2)$$

$$R_t = R_{t-1} + \gamma_{t-1} I_{t-1} \quad (3)$$

$$D_t = D_{t-1} + \mu_{t-1} \gamma_{t-1} I_{t-1} \quad (4)$$

We want the parameters $\langle \beta_t ; \gamma_t ; \mu_t \rangle$ to evolve with time, and guarantee that they always be positive. In order to achieve that, we will assume that they are lognormally distributed, and follow a Geometric Brownion Motion. Thus we shall have:

$$\beta_t = e^{b_t}$$

$$\gamma_t = e^{g_t}$$

$$\mu_t = e^{gm_t}$$

$$b_t = b_{t-1} + v_{\beta,t} \quad (8)$$

$$g_t = g_{t-1} + v_{\gamma,t} \quad (9)$$

$$m_t = m_{t-1} + v_{\mu,t} \quad (10)$$

The terms $\langle v_{\beta,t} ; v_{\gamma,t} ; v_{\mu,t} \rangle$ are normally distributed white noises (and therefore independent of any past realizations), with zero mean and contemporaneous covariance matrix Q.

The observation equation will follow (4), because we have chosen to work solely on death data. Unfortunately, contamination data tend to be underestimated, causing errors and biases in the dynamic system. In case you are working with data from countries where you trust the contamination numbers are accurate, you can add (1) in the observation equations.

The other important aspect is that we will work with the log of (1)-(4). This not only facilitates the algebra, but also approaches the problem in a more proper way, since

diseases tend to spread exponentially, not linearly. Please remind that D_t is the fraction of the population that has died. Denoting the (log of the) number of deaths that happened at time t+1 asΔ_t, and the population as "pop", we will have

$$\Delta_t = ln[pop(D_{t+1} - D_t)] = ln(pop) + ln(D_{t+1} - D_t) + e_t$$

Making use of (4), our observation equation becomes:

$$ln(\Delta_t) = ln(pop) + m_t + g_t + ln(I_t) + e_t$$

Where e_t is a white noise, normally distributed with mean zero, variance R, and totally independent of $\langle v_\beta ; v_\gamma ; v_\mu \rangle$. Substituting (2) in the equation above, it becomes:

$$ln(\Delta_t) = ln(pop) + m_t + g_t + ln(1 + S_{t-1}e^{b_t} - e^{g_{t-1}}) + ln(I_{t-1}) + e_t \qquad (8)$$

The state equations are going to be separated into two groups. The first group is given by (1)-(3) and once the parameters are known, they will be defined deterministically. The second group is composed by (5)-(7), where all these parameters follow random-walks, we will group theses variables in a vector $\vec{z}_t = \langle b_t; g_t; m_t \rangle'$.

The reader may observe that the system is constructed in such a way that the number of deaths today will allow us to infer the parameters of yesterday. So we will always be one day behind.

2.2) The Recursion

Just to Remind the recursion equations in the Extended Kalman Filter, we must firstly define the (3x3) Mean Square error matrices for the state variables:

$$P_{t/t-1} = E\left[\left(\vec{z}_t - \vec{z}_{t/t-1}\right)\left(\vec{z}_t - \vec{z}_{t/t-1}\right)'\right] \qquad (9)$$

$$P_{t/t} = E\left[\left(\vec{z}_t - \vec{z}_{t/t}\right)\left(\vec{z}_t - \vec{z}_{t/t}\right)'\right] \qquad (10)$$

We start with the known initial known values $\langle S_0; I_0; R_0; \beta_{1/0}; \gamma_{1/0}; \mu_{1/0}; R; Q; P_{1/0}\rangle$. Following the Extended Kalman Filter recursions - and considering that $\langle S_{0/0}; I_{0/0}; R_{0/0}\rangle = \langle S_0; I_0; R_0\rangle$ - we will have the following recursive equations:

$$z_{t/t-1} = z_{t-1/t-1} \tag{11}$$

$$\begin{bmatrix} S_{t/t-1} \\ I_{t/t-1} \\ R_{t/t-1} \end{bmatrix} = \begin{bmatrix} S_{t-1/t-1}\left(1 - I_{t-1/t-1}e^{b_{t/t-1}}\right) \\ I_{t-1/t-1}\left(1 + S_{t-1/t-1}e^{b_{t/t-1}} - e^{g_{t-1/t-1}}\right) \\ R_{t-1/t-1} + I_{t-1/t-1}e^{g_{t-1t/t-1}} \end{bmatrix} \tag{12}$$

$$\Delta_{t/t-1} = ln(pop) + m_{t/t-1} + g_{t/t-1} + ln\left(1 + S_{t-1/t-1}e^{b_{t/t-1}} - e^{g_{t-1/t-1}}\right) + ln\left(I_{t-1/t-1}\right) \tag{13}$$

The forecast error for Δ_t made at t-1 is given by:

$$\Delta_t - \Delta_{t/t-1} = m_t - m_{t/t-1} + g_t - g_{t/t-1} + ln(1 + S_{t-1}e^{b_t} - e^{g_{t-1}}) - ln\left(1 + S_{t-1/t-1}e^{b_{t/t-1}} - e^{g_{t-1/t-1}}\right) + ln(I_{t-1}) - ln\left(I_{t-1/t-1}\right)$$

Since all the values for t-1/t-1 had already been updated, we will only update those with subscript t/t-1. In order to do that, we must have the derivative of $\Delta_{t/t-1}$ with respect to $\vec{z}_{t/t-1}$. We shall call this vector $H_t{}'$, and it is given by

$$H_t{}' = \left[S_{t-1/t-1}e^{b_{t/t-1}} / \left(1 + S_{t-1/t-1}e^{b_{t/t-1}} - e^{g_{t-1/t-1}}\right);\quad 1;\quad 1\right] \tag{14}$$

$$E\left(\Delta_t - \Delta_{t/t-1}\right)^2 \cong H_t{}' P_{t/t-1} H_t + R \tag{15}$$

$$\vec{z}_{t/t} = \vec{z}_{t/t-1} + P_{t/t-1}H_t\left(H_t{}' P_{t/t-1} H_t + R\right)^{-1}\left(y_t - y_{t/t-1}\right) \tag{16}$$

$$\begin{bmatrix} S_{t/t} \\ I_{t/t} \\ R_{t/t} \end{bmatrix} = \begin{bmatrix} S_{t-1/t-1}\left(1 - I_{t-1/t-1}e^{b_{t/t}}\right) \\ I_{t-1/t-1}\left(1 + S_{t-1/t-1}e^{b_{t/t}} - e^{g_{t-1/t-1}}\right) \\ R_{t-1/t-1} + I_{t-1/t-1}e^{g_{t-1t/t-1}} \end{bmatrix} \quad (17)$$

$$P_{t/t} = P_{t/t-1} - P_{t/t-1}H_t\left(H_t{'}P_{t/t-1}H_t + R\right)^{-1}H_t{'}P_{t/t-1} \quad (18)$$

$$P_{t+1/t} = P_{t/t} + Q \quad (19)$$

2.3) Maximum Likelihood Estimates and Some Practical Tips

Considering that the errors $\Delta_t - \Delta_{t/t-1}$ are normally distributed, with variance given by (15), and the last observation is Δ_T, the log-likelihood function will be given by :

$$LL = -\frac{1}{2}\sum_{i=1}^{T}\begin{Bmatrix}\left(H_t{'}P_{t/t-1}H_t + R\right)^{-1}(\Delta_i - \Delta_{i-1})^2 \\ +ln(2\pi) + ln\left(H_t{'}P_{t/t-1}H_t + R\right)\end{Bmatrix} \quad (20)$$

We should choose $\langle S_0; I_0; R_0; b_{1/0}; g_{1/0}; m_{1/0}; R; Q; P_{1/0}\rangle$ that maximizes (19). Let´s call LL_0 this maximum value of (19). In case we want to make some hypothesis tests about the parameters, we could perform for instance a Likelihood Ration Test. Let us say that you want to test r restrictions. Then maximize (19) with these r restrictions. This restricted maximum will be called LL_r. Using the property that $2(LL_0 - LL_r)$ will converge to a Chi-Square Distribution with r degrees of freedom, we can reject or not the restrictions.

Some suggestions, in a more practical level, might be useful. A good warm start values for the covariance matrix Q is to assume a diagonal matrix with variances around 0,01%. Also, we would suggest that $P_{1/0}$ should equal Q plus a positive definite matrix. In order to guarantee Q and $P_{1/0}$ to be positive definite, define them as the product of triangular matrices and their transposes and estimate the parameters of the triangular matrices.

Secondly, the first observations tend to be a little problematic as we are dealing with logs. For instance an increase from 1 death to 2 will be like an increase from 1.000 to 2.000. So, depending on the beginning of your sample, it might be worthy to exclude the first values. In our case, the first data has 5 deaths, followed bay 2 . That would be a huge decrease of 60% from day 1 to day 2, so we excluded these first 2 observations

Finally, due to the lockdown, we had a huge drop in β in the begging of our sample, but not observed in γ or μ, which showed that this dynamic approach was capturing the expected behaviour. The problem was that due to this huge fall, the variances of this parameter became extremely high. By adding a dummy variable during 10 days in the dynamics of β, we had a good improvement in our estimations. So the equation that described the dynamics of the coefficient b_t became:

$$b_t = b_{t-1} + dum_t + v_{\beta,t} \tag{23}$$

2.4) Forecasting out of Sample

Since we cannot update the variables for out of sample forecast, we will project the future values from our last observation T. Since z_t follows random-walks, we shall have our forecasts for the n days ahead given by:

$$z_{T+n/T} = z_{T/T} \tag{24}$$

$$\begin{bmatrix} S_{T+n/T} \\ I_{T+n/T} \\ R_{T+n/T} \end{bmatrix} = \begin{bmatrix} S_{T+n-1/T}\left(1 - I_{T+n-1/T}e^{b_{T+n-1/T}}\right) \\ I_{T+n-1/T}\left(1 + S_{T+n-1/T}e^{b_{T+n-1/T}} - e^{g_{T+n-1/T}}\right) \\ R_{T+n-1/T} + I_{T+n-1/T}e^{g_{T+n-1/T}} \end{bmatrix} \tag{25}$$

And their Variance of errors

$$P_{T+n/T} = P_{T/T} + nQ \tag{26}$$

As for the variance of $\langle S_{T+n/T}; I_{T+n/T}; R_{T+n/T} \rangle$ and their confidence intervals, it was necessary for us to make use of monte Carlo simulations due to the non-linear recursive dynamics of these variables.

2.5) Identification

Identification is a typical problem in State-Space models. In Essence, it happens when the system can be rewritten in various ways with the same variables, resulting in the same (maximum) likelihood values. For example, if there were no equations (1)-(3), we would have such a problem with variables g_t and m_t.

From the theoretical point of view, since all the three variables $\langle \beta_t; \gamma_t; \mu_t \rangle$ have different dynamics and effects in the whole system due to (1)-(3), we wouldn´t have at first sight such a problem. However, there are some solutions whose likelihood are so close to the maximum, that they are statistically the same. Although this is not an exact identification problem, we might be very close to one. This would lead to different solutions every time we would run with one more observation, or start the optimization from different neighbourhoods. Usually this means that we are having many degrees of freedom. From our experience, the difference these different solutions deliver is not significant in terms of forecasting the number of deaths. However they tend to differ considerably in the dynamics of $\langle \beta_t; \gamma_t; \mu_t \rangle$, and even more in $\langle S_t; I_t; R_t \rangle$, because the latter are accumulating the effect of former.

When we are faced with such a problem, it is recommended to choose towards the simplest model. Due to our recent and short experience with Covid-19, we would expect the contamination rate β_t to vary considerably according to lockdown, hygiene policies, etc. The mortality rate μ_t might be our natural candidate to not vary much with time. So we decided (in the case of the city of São Paulo, at least) to simplify the system by imposing$\gamma_t = -\ln(14)$ (the recovery rate is expected to be around the inverse of the number of days it takes on average for the patient to be fully recovered). Our system now has only two state variables $\langle b_t; m_t \rangle$ and 7 parameters less to estimate ($P_{1/0}$ and Q are now 2x2 matrices). The likelihood obtained is almost identical to the unrestricted model. Please keep in mind that although it worked for our sample this might not apply to other cases.

2.6) Smoothing

In order to obtain the smoothed estimations, we start with our last observation T and the recursion goes backwards:

$$P_{t/T} = E\left[\left(\vec{z}_t - \vec{z}_{t/T}\right)\left(\vec{z}_t - \vec{z}_{t/T}\right)'\right] \quad (27)$$

$$J_t = P_{t/t} P_{t+1/t}^{-1} \quad (28)$$

$$\vec{z}_{t/T} = \vec{z}_{t/t} + J_t\left(z_{t+1/T} - z_{t+1/t}\right) \quad (29)$$

$$P_{t/t} = P_{t/t} - J_t\left(P_{t+1/t} - P_{t+1/T}\right)J_t{}' \tag{30}$$

$$\begin{bmatrix} S_{t/T} \\ I_{t/T} \\ R_{t/T} \end{bmatrix} = \begin{bmatrix} S_{t-1/T}\left(1 - I_{t-1/T}e^{b_{t/T}}\right) \\ I_{t-1/T}\left(1 + S_{t-1/T}e^{b_{t/T}} - e^{g_{t-1/T}}\right) \\ R_{t-1/T} + I_{t-1/T}e^{g_{t-1/T}} \end{bmatrix} \tag{30}$$

2.7) Adding Bounds

Due to the fact that we are dealing with a non-linear system with local minima problems, the traditional algorithms may lead to spurious solutions. In order to bypass this problem, we may impose some upper and lower bounds to our variables. One suggestion is to replace the assumption that the rate(s) instead of following a geometric Brownion motion (and therefore are lognormally distributed, as we have done so far), we may model them as (The value L and U are for the respective lower and upper values):

$$\beta_t = (L_b + U_b e^{b_t})/(1 + e^{b_t}) \tag{31}$$

$$\gamma_t = \left(L_g + U_g e^{g_t}\right)/(1 + e^{g_t}) \tag{32}$$

$$\mu_t = (L_m + U_m e^{m_t})/(1 + e^{m_t}) \tag{33}$$

In the former lognormal case L=0, and U→ ∞. For instance, we may not desire any of these rates to be above 100% (it would be counterintuitive). In our case, we have assumed that$L_b = 0; U_b = 100\%; L_m = 0; U_m = 1\%.$

If this approach has been adopted, keep in mind that our vector H_t will now be given by:

$$H_t = \begin{bmatrix} S_{t-1/t-1} I_{t-1/t-1} (U_c - L_c) e^{b_{t/t-1}} / \left[\left(1 + e^{b_{t/t-1}}\right)^2 I_{t/t-1} \right] \\ U_r / (U_r + L_r e^{-g_{t/t-1}}) - 1/(1 + e^{-g_{t/t-1}}) \\ U_m / (U_m + L_m e^{-m_{t/t-1}}) - 1/(1 + e^{-m_{t/t-1}}) \end{bmatrix} \quad (34)$$

3) Results

We have been running this model for the city of São Paulo. Our sample starts on March 16th and ends on Oct 1st. Firstly we will present our current estimates and forecasts. Secondly we shall show how the model had predicted (we choose every first day of the month) through time vis-a-vis what really happened.

3.1) Current Estimations

In this section we will present all the results obtained at Oct 1st 2020, not only in terms of interpreting the recent in light of the SIR model, but also our projections for the deaths in São Paulo.

We start with the number of deaths. The thin red line describes the actual deaths, the thick blue line the fitted model (t/t-1), and the dotted lines represent the 95% confidence interval. All the data after Oct 1st are projections. The total deaths reported until Oct 1st are 14.514. The total expected to the end of the pandemic is 17,290 with 95% confidence interval between 16.201 and 20.084

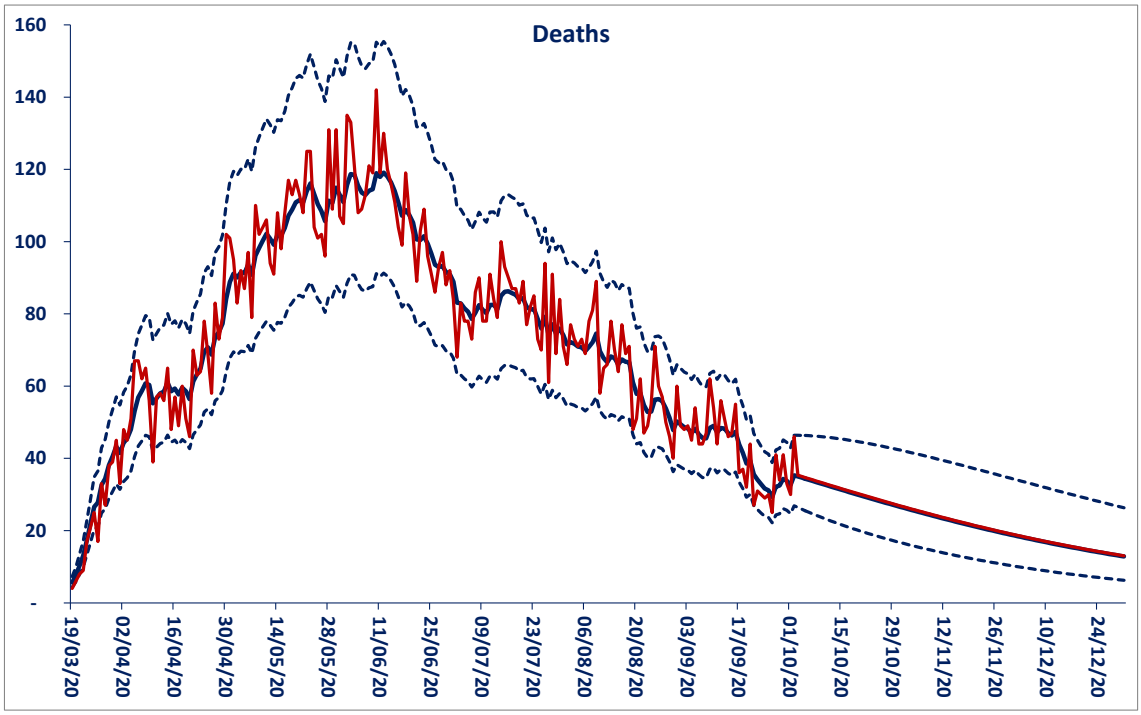

Figure 1 – Actual, Fitted and Predicted Deaths

For the evolution of the contamination rate, its smoothed estimation is shown below (the dotted lines represent the 95% confidence interval):

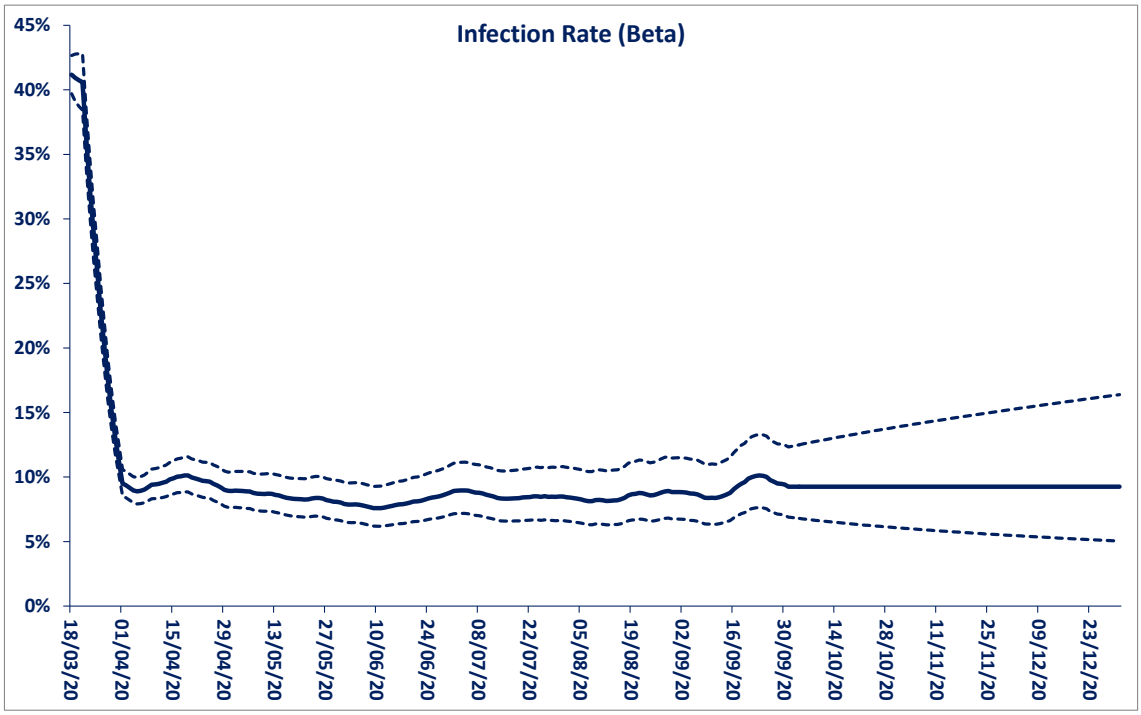

Figure 2 – Infection Rate

This huge drop in the last 10 days March coincides with the beginning of the lockdown in São Paulo. The model seems to have captured very well the effects of this policy. As already mentioned, in our first estimation, this huge drop was present, but due to the abruptness of the drop, the variances exploded. When we added a dummy variable in these days, the variance stabilized to reasonable values. In light of these estimations, the effectiveness of the lockdown is unquestionable.

In our case, in order to avoid (over)identification problems, we have assumed $\gamma_t = 1/14$ without any significant loss in our Likelihood Function. This means that the replication rate - which is given by β_t/γ_t - is simply β_t multiplied by 14:

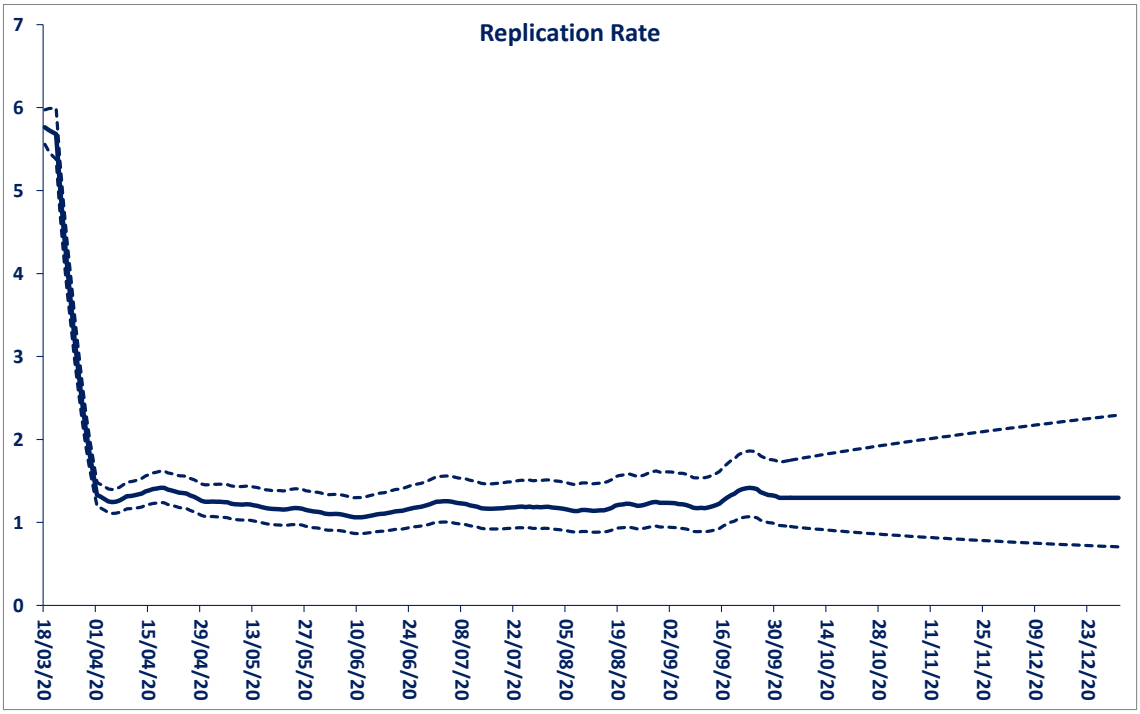

Figure 3 – The Replication Rate

One may be surprised by the high level in the beginning of the sample (the expected value without lockdowns would be around 3). However, the deaths around these first dates represent those who probably got infected in the end of February/beginning of March, which coincides with the period of Carnival in Brazil, in which no restrictions were enforced by the government. So the Carnival flowed just like any normal year. Therefore, we shouldn´t be surprised that this rate was around twice the expected.

Regarding the mortality rate, it kept oscillating between 0,3% and 0,4%. The model showed a peak in the beginning of June, which was when the ICUs in São Paulo hit their maximum values. Fortunately, it wasn´t enough to bring a collapse to the health system in São Paulo. From that period on, the mortality rate started to slowly decline:

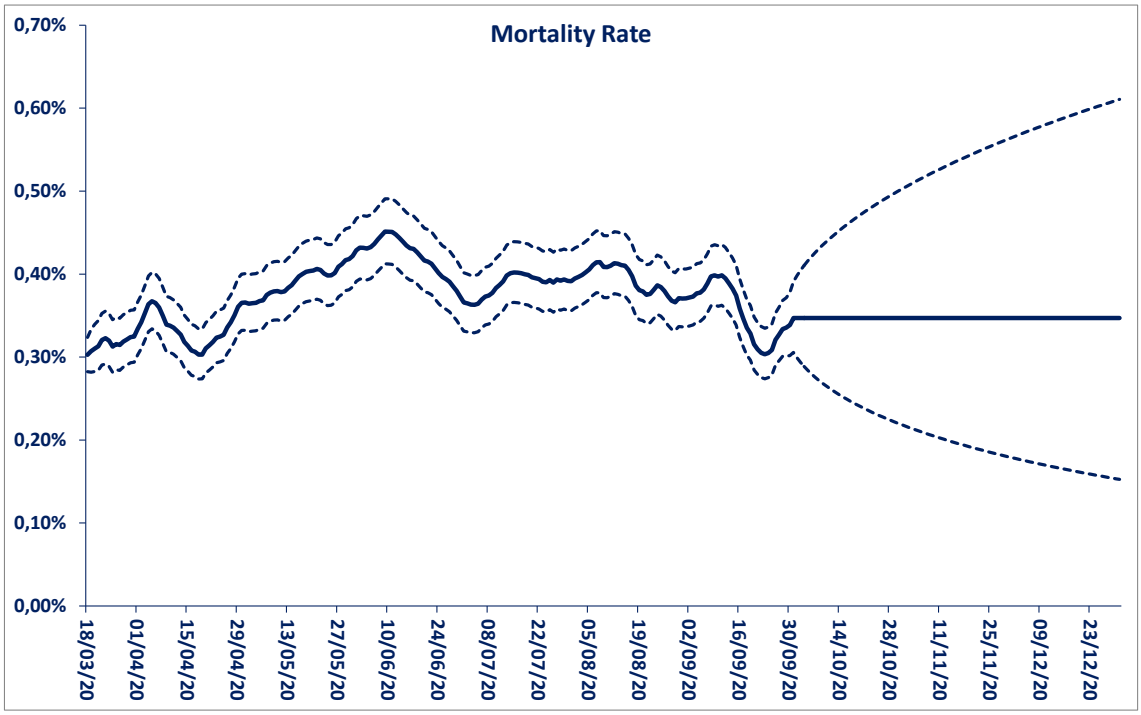

Figure 4 – Mortality Rate

Finally, there´s the evolution of the fractions of the population (smoothed estimates) divided in Susceptible (Blue line), Infected (Red line) and Recovered (Green line), as well as their 95% confidence intervals. After Oct 1st they are projections

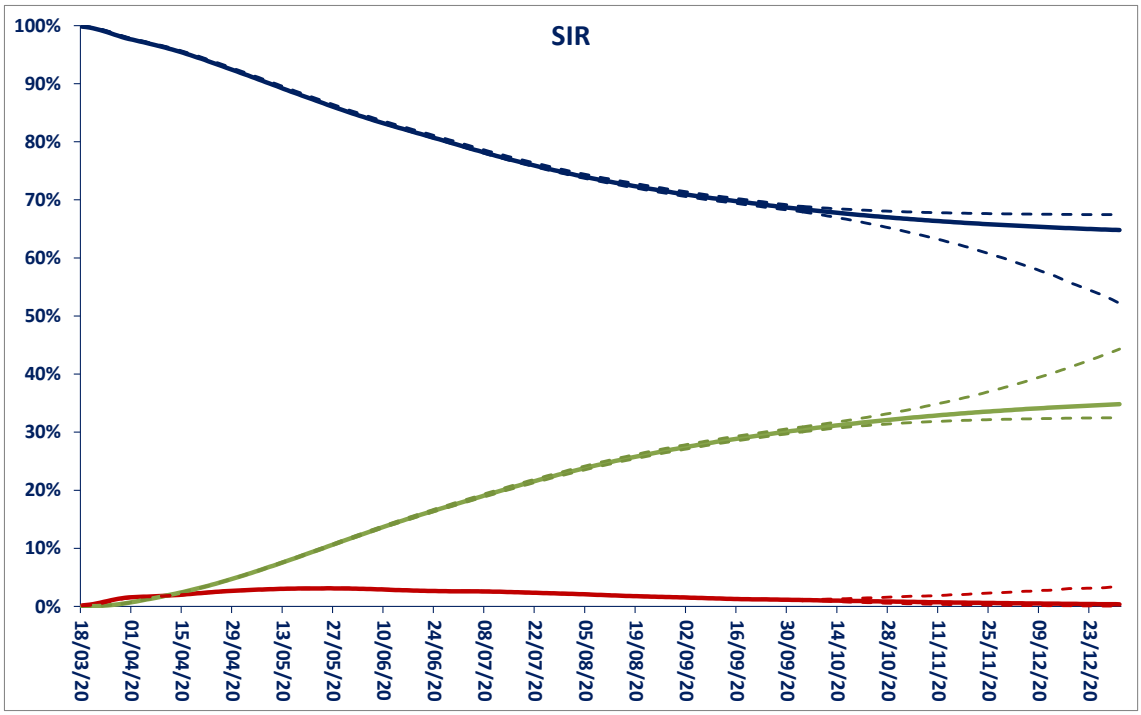

Figure 5 – Susceptible, Infected and Removed

According to our model, the susceptible fraction is now around 68,50%, the infected/infectious around 1,5% and the removed around 30%. The contamination estimations are higher than the average numbers we tend to see in the news.

One can see that the distribution of the predictions are much skewed, and far from being symmetric. This is quite intuitive actually. Let us consider the susceptible. Today they are around 68,5%, and expected to converge in the long run to 64%. The highest possible value it can achieve is 68,5% (this group cannot increase with time), while its lowest possible value is zero, the contamination effect of the (accumulated) β_t is quite asymmetric indeed. The same argument goes for the remaining part of the population, but the effect is on the opposite direction. Please observe that the border stressed trajectory describes a second wave forming in the infected group (red line at the bottom of the graph) that is not decreasing up to December.

3.2) Evolution of Predictions

We now want to show the adaptability of the model, and see how it learned and forecasted throughout the months. We will always take the estimations and projections of the model in the first day each month. The thin red line is the real deaths to Oct 1st, and the thick blue line is the fitted/projected curve and its 95% interval. A dividing line was put in the graph so the reader can see more clearly the date the estimations were made.

In April 1st there was only two weeks data, so we will start on May 1st. As can be seen, at this date we were facing a huge upward movement, and deaths were accelerating again, after it began to stabilize in the beginning of April. Naturally, the model forecasted a more stressed scenario than what really happened:

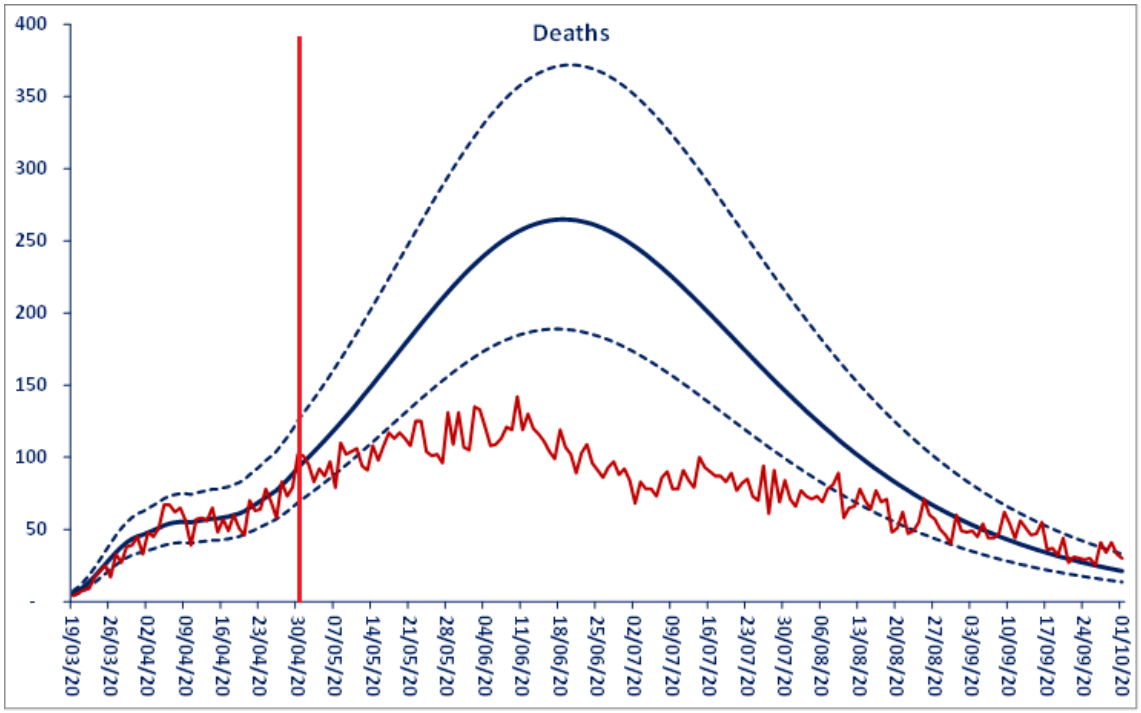

Figure 6 – Predictions made at May 1st vs actual

We now move to June 1st. We are close to the highest values, and the model foresaw the peak and the decay according to the last values. It is easy now to see that the downfall that occurred in June was more radical than the projection, but from July on was quite adherent to the projection interval.

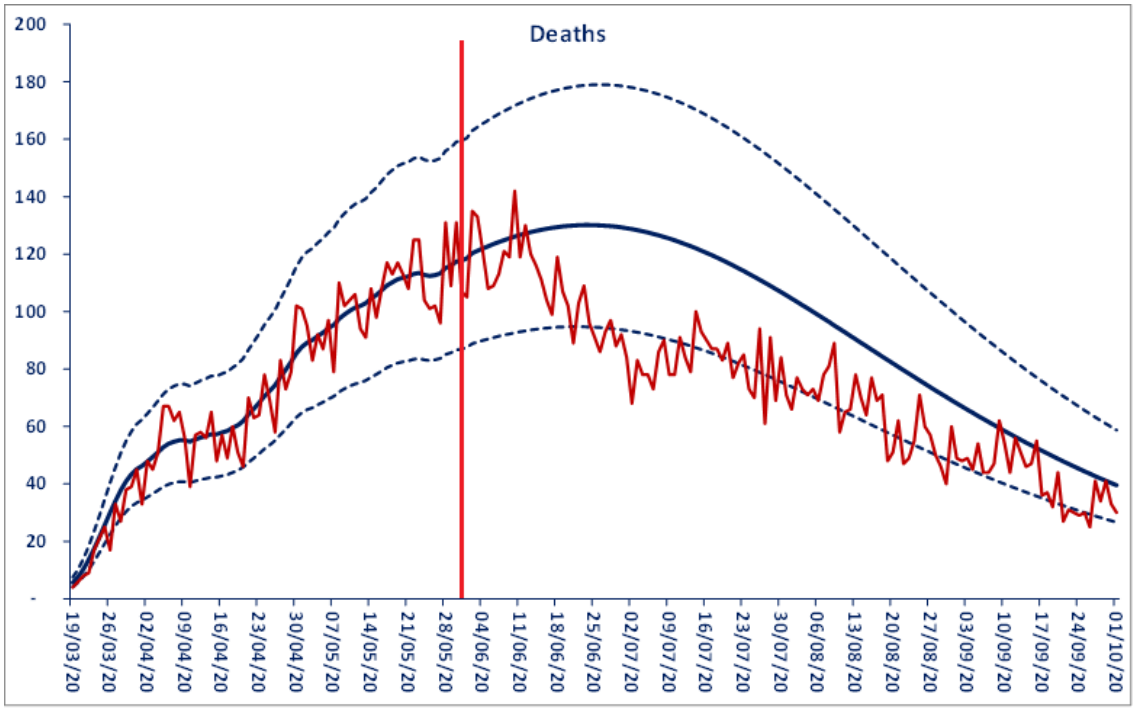

Figure 7 – Predictions made at June 1st vs actual

We are now at July 1st. The model understood that deaths were decreasing now quite fast during the month of June. Consequently, it projected a more optimistic downward movement that would lead to underestimations.

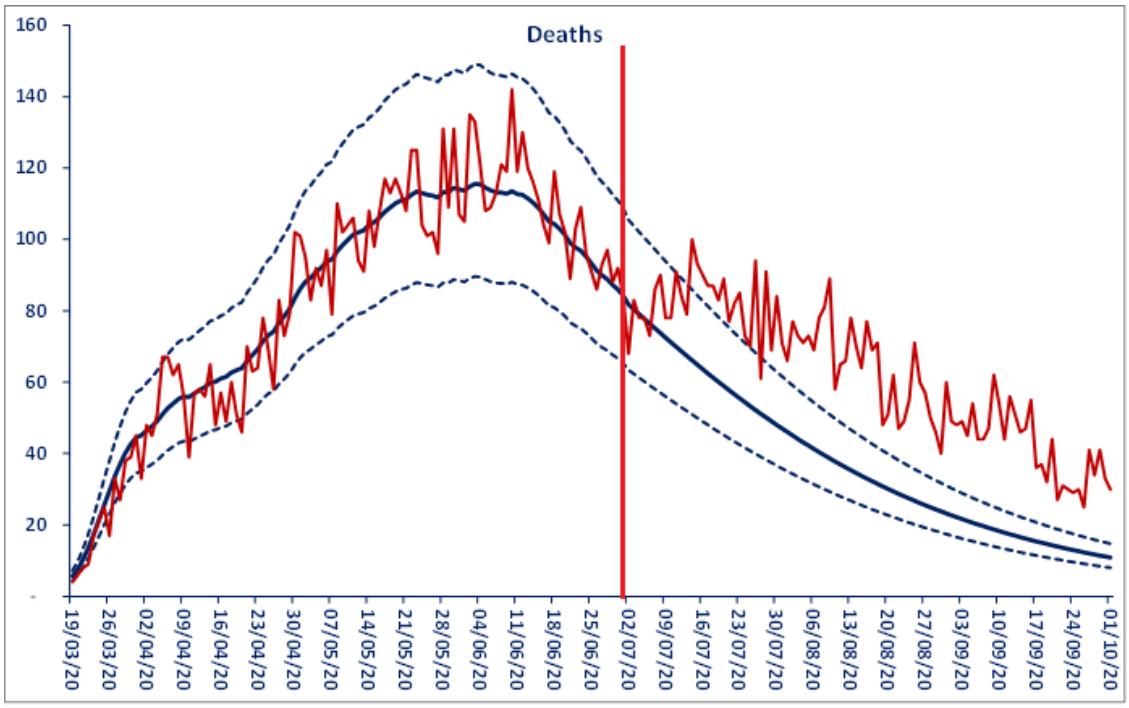

Figure 8 – Predictions made at July 1st vs actual

We now move to Aug 1st. The model learned that the speed deaths were decreasing was lower than before, and it made quite a good forecast for the next two months:

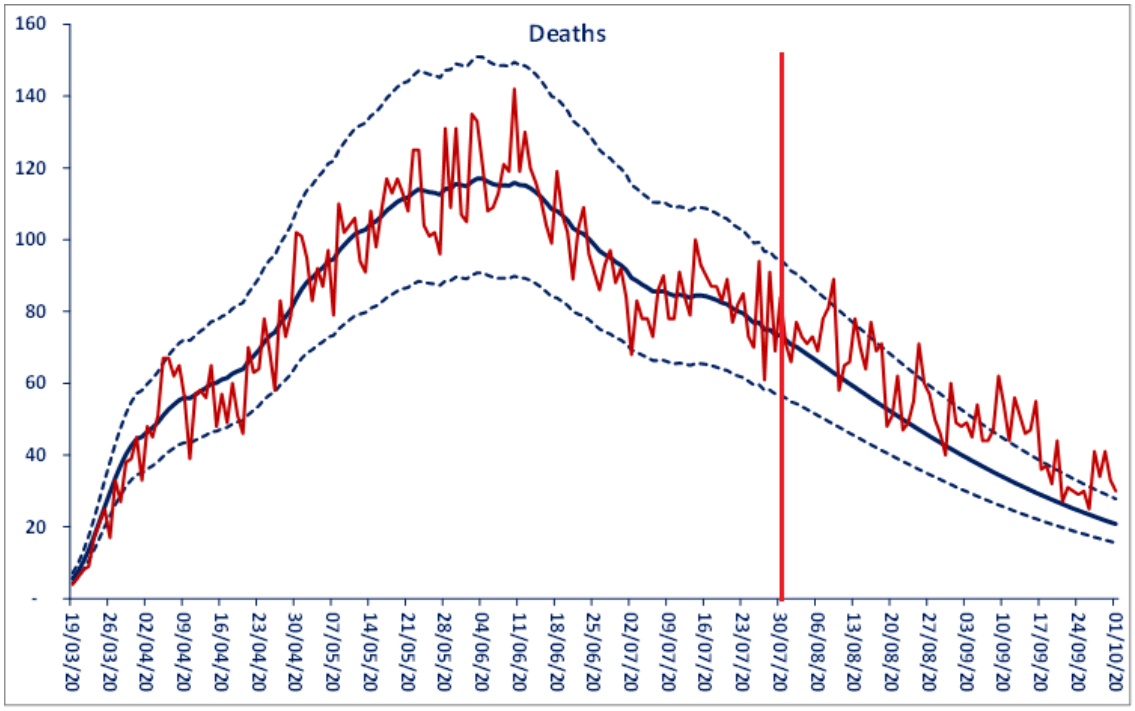

Figure 9 – Predictions made at Aug 1st vs actual

Finally, we move to Sep 1st. The behaviour of the month of August did not differ much from July. Therefore the model predicted basically the same pattern. The last month out the sample behaved quite in line with the projections.

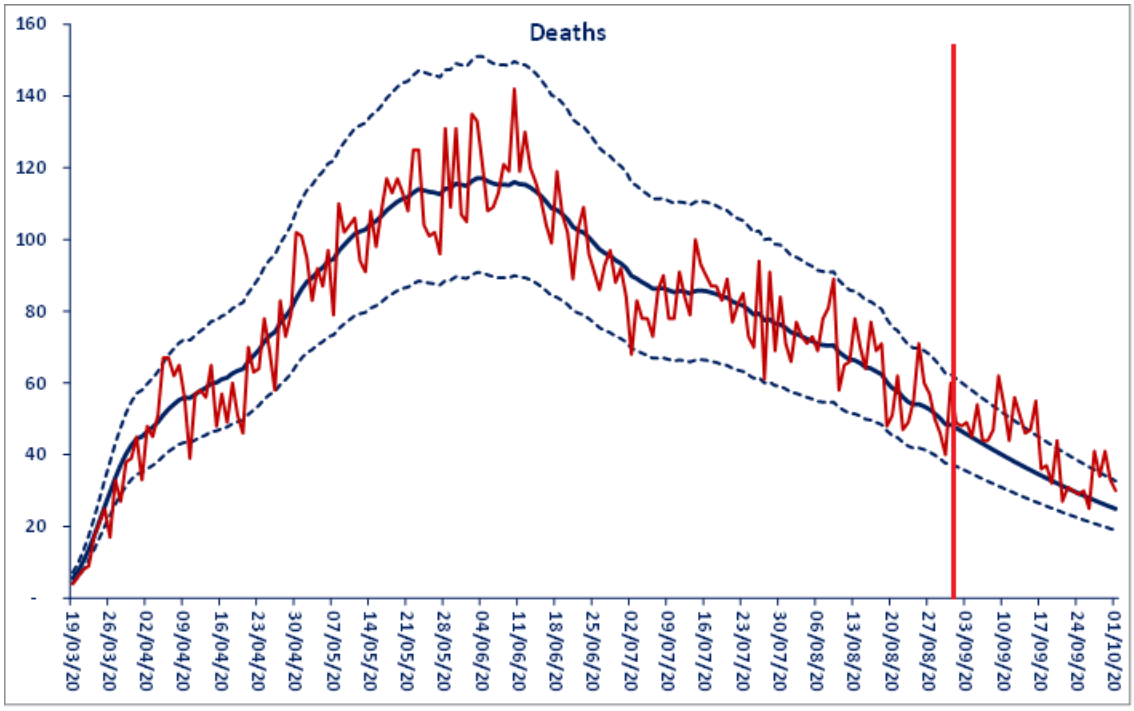

Figure 10 – Predictions made at Sep 1st vs actual

Observing the evolution of the forecasts, we can see that the model learns fast, adapts quickly to changes and seemed to get better predictions as the sample increased and thus more information.

4) Conclusions

We have proposed a model that allows the parameters of the SIR model to evolve with time. By only using deaths data due to Covid-19, we could infer the history of the mortality, contamination and recovery rates, and also been able to make projections.

The model seemed to capture very well and discriminate the effects of the lockdown on contamination, and described quite fairly the evolution of the mortality rate as we expected according to ICU occupations. This ability to discriminate different effects on each specific rate (lockdown on contamination rate and occupations of ICU on mortality rates) was a very nice surprise.

A whole analysis was made of how the model had been adapting, learning and forecasting, showing interesting out of sample results. We believe that in case of new waves, promising treatments, and maybe vaccines, the model have proven to be quite adaptable.

5) References

1) Dukic, V., H. F. Lopes and N.G. Polson (2012). Tracking Epidemics with Google Flu Trends Data and a State-Space SEIR Model. *Journal of the American Statistical Association*, 2012, 107, 1410-1426
2) Kalman R.E. (1960): A new Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering 82 (1):* 35-45
3) Kalman R.E. and R.S. Bucy (1961): New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering 83 (1):* 95-108
4) Kermack W. and A. McKendrick (1927): Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London , Series* A 115, 700-721
5) Kobayashi G., S. Sugasawa, H. Tamae and T. Ozu (2020): Predicting Infection of Covid-19 in Japan: State Space Modelling Approach.
6) Osthusk D., K.S. Hichmann, P.C. Caragea, D. Rigdon and S. Del Valle (2013): Forecasting Seasonal Influenza via a State-Space SIR Model. *The Annals of Applied Statistics 11(1)*, 202-224

7) Wang L., Y. Zhou, J. He, B. Zhu, F. Wang, L. Tang, M. Eisenberg and P. X. K. Song (2020): An Epidemiological Forecast Model and Software Accessing Interventions on Covid-19 Epidemic in China. sthusk D., K.S. Hichmann, P.C. Caragea, D. Rigdon and S. Del Valle (2013): Forecasting Seasonal Influenza via a State-Space SIR Model. *The Annals of Applied Statistics 11(1)*, 202-224