



Munich Personal RePEc Archive

COVID-19 with uncertain phases: estimation issues with an illustration for Argentina

Ahumada, Hildegart and Espina, Santos and Navajas,
Fernando H.

Universidad T. Di Tella, Argentina, Fundacion de Investigaciones
Economicas Lationamericanas (FIEL), FIEL, Universidad de Buenos
Aires Univ. Nac. de La Plata

20 June 2020

Online at <https://mpra.ub.uni-muenchen.de/101466/>
MPRA Paper No. 101466, posted 05 Jul 2020 18:58 UTC

Ahumada, Hildegart and Espina, Santos and Navajas, Fernando, "COVID-19 with Uncertain Phases: Estimation Issues with An Illustration for Argentina "(June 20, 2020). Available at SSRN: <https://ssrn.com/abstract=3633500> or <http://dx.doi.org/10.2139/ssrn.3633500>

COVID-19 with uncertain phases: estimation issues with an illustration for Argentina

Hildegart Ahumada* Santos Espina Mairal^{#a} Fernando Navajas^{#bc}

UTDT

FIEL and UCEMA

FIEL, UBA and UNLP

June 20, 2020

Abstract

We use an approach to assess COVID-19 performance that starts from what we consider is the most likely set of hypotheses about the uncertain evolution of the pandemic, that envisage a sequence of different cycles with unknown duration and magnitude over 18-24 months. This pattern implies a research strategy where short-term time series forecasting of the evolution of observed cases and deaths play a central role in both detecting transitions from phase to phase of infections and the estimation of necessarily changing structural parameters and indicators of a SIRD model. We illustrate our approach with Buenos Aires City performance, which represents a significant share of the Argentine case with an early introduction of a lockdown followed by a second wave latter on. This approach can be extended to include measures of the intensity and compliance of lockdowns, as well as the heterogeneity across areas. We find that mobility (as a proxy for the effectiveness of the lockdown) has an impact on observed cases in Buenos Aires City with a lag of 8 days and deaths relate with new cases registered 16 to 19 days before. Mobility has a clear impact on the growth rate of cases and by extension deaths. Our approach and results have implications for policy dialogue issues.

*Universidad Torcuato Di Tella, # Fundación de Investigaciones Económicas Latinoamericanas, a Universidad del CEMA, b Universidad de Buenos Aires, c Universidad Nacional de La Plata
Corresponding authors: hahumada@utdt.edu ; navajas@fiel.org.ar

1. Introduction

Far from being a well understood phenomenon, a lot of uncertainty prevails in the performance and observed outcomes of the COVID-19 pandemic because it is surrounded by many unknowns, as one should expect from this newly born global pandemic. This makes the epidemiological process shaping COVID-19 dynamics uncertain in terms of both time and rate of change or evolution, also because the development of a vaccine fundamentally alters the scenario. This scientific uncertainty is faced from all angles of study including economic-epidemiological models on the optimal duration and intensity of lockdowns which use for their simulation purposes parameter values or interval values taken from epidemiological studies (Acemoglu *et al*, 2020; Alvarez *et al*, 2020; Garriga *et al* 2020; Gonzalez Eiras and Niepelt, 2020 to mention a short but representative list). Some scientific synthesis based on epidemiological models and empirical and historical evidence expects a pattern with cycles ahead in a span period of 18-24 months (Moore *et al*, 2020) given or conditional to the absence of a vaccine. From our perspective, this view should have important methodological implications for measurement and estimation efforts to follow and forecast COVID19 dynamics, which are useful for adjusting parameters used in simulations and modelling or evaluations. It clearly implies that we are not witnessing, except perhaps for the takeoff, an unmitigated trajectory for most country-cases, due to the interplay with interventions or lockdowns and the, averting or not, behavior of society. Take offs and initial mitigation effects to flatten the curve have been estimated in a way closely attached to the structural parameters of the SIR model (*eg* Harris, 2020), with forecasting techniques (*eg* Castle, et al, 2020), with sophisticated epidemiological frameworks (Imperial College, 2020) or through mixed reduced form econometric models (Hsiang et al, 2020). However, the recognition that we may be witnessing only an initial phase of a multi-cycle trajectory implies a completely different strategy to avoid a singled-peaked SIR model that could lead to forecasting and, possibly, policy mistakes. The pattern in many countries is that there may be no single peaked process or cycle but rather phases of a temporarily repressed dynamic that will move in time, but this is also an uncertain hypothesis.

Mitigation policies, human choices and shocks, make forecasting of the cases and deaths be subjected to distribution shifts, leading to parameter instability and serious forecast failure. This is a fact that has been discovered some time ago in the field of macro econometric times series modelling (see Hendry, 2000; 2020) and has been recently applied to COVID-19 modelling start-ups (Castle *et al*, 2020). They use forecasting modelling as a convenient strategy to predict the path of COVID-19 cases in the short term while comparing their forecasts with those of a conventional epidemic modelling like the SIR model, with parameters obtained from nonlinear

estimation. They suggest a complementarity between forecasting algorithms for the short term and epidemiology models for the analysis of the expected trajectory. However, for reasons reflected in the epidemiology literature (Moore *et al*, 2020) the overall process of the pandemic is uncertain in dimension, time span, speed and recurrence cycles. This late pattern may affect the prediction of long term forecast with the SIR model or with empirical logistic (Gompertz) curve models that assume a single peaked trajectory (see for example Batista, 2020; Sanchez-Villegas, 2020, Lee et al, 2020). Other more complex models based on detailed information and using either variants of nonlinear growth models (IHME, 2020) have been exposed by the specialized press to criticisms for forecast failure (underestimation) in the US (Wallace-Wells, 2020) and neural network forecast models applied to Brazil have also led to significant underestimation (Pereira *et al*, 2020) or have been defeated by exponential growth time series models (Martinez *et al*, 2020) . Finally, returning to a short term forecasting strategy one of the most complete and recognized models for forecasting COVID-19 with results for many countries (Imperial College, 2020) has also a short term horizon, centering its working on the evolution of observed deaths rather than observed cases, due to the measurement problems with the later, and using a SEIR framework.

In this paper we submit that there is a connection to exploit between short-term time series forecasts on the one hand and parameters of a SIRD model on the other, insofar as the detection of the presumable growth regimes that the epidemic may exhibit in an uncertain multiple waves scenario. This is motivated by the observation that a dynamic regression model of the (log approximated) rate of growth of COVID-19 observed cases relates well with the infectious rate parameter of the SIRD model estimated in a linearized (instead of nonlinear) fashion as provided for instance in Harris (2020). Thus, forecasts of a time series model that takes advantage of impulse and step saturation techniques (Hendry and Doornik, 2014) to detect breaks allows helping in the process of detecting changes in the evolving stop and go nature of the COVID-19 process and directly translate into SIRD parameters, providing a convenient dialogue from a methodological point of view. A complementary forecasting approach can be performed in the case of observed deaths, which are an necessary element for policy design and debate.

We illustrate our modelling strategy for the case of Buenos Aires City which serves our purpose and also represents so far the central and richer driver of the observed dynamics. This can be described, up to the moment of writing this version, in broadly three stages with a four likely to come soon, each corresponding to a different month from March to June. In mid-march 2020, as some European countries were showing the worst face of COVID-19 and given the undeniably links with Italy and Spain in genes, culture and, in particular, travel ties, the government in Argentina felt the abyss ahead, as imported cases and related contact cases started to generate a pandemic dynamics. Supported by a solid group of virologists and health care experts, a quarantine started officially on March 20 and has been extended since

then with relaxation rules since late May. The disease tended to concentrate more and more in Buenos Aires City and the Great Buenos Aires Area, explaining more than 80% of cases and deaths as data showed that the interior of the country was in a different situation, except in a couple of provinces. The country numbers reacted very fast to the early lockdown as this was able to quickly flatten the curve of reported cases. Days after Argentina was ranked on April 7 by Imperial College London (2020a) survey of 42 cases as a stabilizing ($R_t < 1$) and relatively small (deaths < 100) case. From the beginning the country displayed a relatively low number of test per million of inhabitants, a much noticed difference with neighbor Chile, which initially showed a very similar dynamic in terms of new additional cases and of deaths to move later on to a far worse performance. Subsequently, tests numbers increased and methods improved but tended towards a priority-testing or focalized strategy, which can be rationalized given limited testing capabilities (Kasy and Teytelboym, 2020). The country stayed very far from organizing an aggressive policy of testing and tracing even though some suggested isolationist strategies using hotels (eg Galiani and Stambulski, 2020) were somehow implemented. During April a clear plateau was achieved, bringing some comfort to authorities but with bigger concerns in the business sector for the observed knock-on effects on economic activity and the perceived macro-economic and sectoral consequences of what seemed to become a prospective very long lockdown, given that it started at the very beginning of autumn. Observed cases in the April plateau fluctuated around 120 daily observed cases, while daily deaths kept running on average at a single digit, with critical hospitalizations at very low levels compared to available (and building up) hospital capacity. This April plateau was shaken in May when on top of an initial outburst in care homes in Buenos Aires City that was controlled hidden cases erupted in one of the largest poor-neighborhoods, following a problem with water supplies. Priority testing towards this area resulted in an astonishing increase in the percentage of positive cases but with relatively low hospitalizations, critical cases and deaths, which can be explained by the demographics of poor neighborhoods as compared to the rest of the City (Panadeiros, 2020). Thus, Buenos Aires (but also Argentina to a lesser extent) was all the way reducing the observed case fatality rate towards values below 3% (compared to 5.5% for the world average, 5.1% for Brazil and 1.6% for Chile, which shows that a very low observed case fatality rate may be compatible with a seriously stressed health system if the absolute number of critical hospitalizations reaches a limit). From this point, a visible second wave started to build up progressively, with some rebalancing between poor and non-poor neighbors suggesting also perhaps a transmission pattern, with observed cases multiplying by a factor of 5 and deaths by a factor of 3 between the end point of the plateau for despite the trend towards lockdown relaxation shown in mobility indicators all over the country. The hazard for a next stage of the pandemic seems likely to be located in the Great Buenos Aires Area, as our data -up to June 10- indicates. Thus, Buenos Aires City and the Great Buenos Aires area came in June to a uncomfortable point where, after a ten-week lockdown, authorities were forced to

accommodate to the need and demand for lockdown relaxation¹ while facing an acceleration of total observed cases and an effective reproduction number above, but with a low average and marginal observed fatality rate. The current setting may represent a precarious equilibrium if the social and economic behavioral response remains close to the pre-lockdown pattern.²

The paper is organized as follows. Section 2 develops our analytical framework connecting the parameters of a SIRD model with a forecasting approach. Section 3 provides a simple data representation of COVID-19 performance in Argentina as one with a comparatively successful initial performance but with a second wave of infections as the effectiveness of the lockdown weakened. Section 4 provides our econometric forecasting analysis of observed cases and deaths in Buenos Aires City. Section 5 concludes with the relevance of our approach and results for policy dialogue and checking, as well as for enriching model and simulations with parameters consistent with observed data.

2. COVID-19 estimation and forecast issues and the SIR model

The reference model for the COVID-19 modelling is the SIR model³, that we extend to a SIRD version to include the dynamics of deaths. Expressions (1) to (4) represent the model. This comprises a set of four differential equations where a susceptible group S_t within a population of size N_t is being affected by a contagious disease giving rise to a group I_t of infected individuals that as the disease progresses lead to R_t recovered and to D_t deaths. By definition $N_t = S_t + I_t + R_t + D_t$ while $C = I + R + D$ says that observed cases C is a variable that add up Infected, Recovered and Death persons. The equations illustrate the transition rates from S_t to I_t to R_t and D_t which are governed by an infection rate α , a recovery rate β and death rate γ and their

¹ Expected losses in terms of output have been updated sharply in Argentina, besides fiscal and monetary spillovers that threaten to weaken macroeconomic stability given the inherited economic outlook and thin room of maneuver. Thus, both macroeconomic constraints and informal sectors put a strong pressure for a comprehensive policy response to avoid a disorganized scenario. The so far contained hospitalizations, critical cases and deaths evolution indicators as compared to existing capacity has been used to argue in favor of an strategic reduction of the quarantine (Urbiztondo, 2020) with a flexible servo-mechanism to react to surprises. An opposite view is expressed by simulations that rely on dynamic effects and on high uncertainty about worse-case scenarios and the capability to react in due time (Castro, 2020). From a theoretical perspective Barnett et al (2020) show that uncertainty about the case fatality rate bias the optimal choice of a policy maker in an asymmetric way. He tends towards more stringent actions if his prior underestimated the observed risks while it he tends to a no change reaction if his prior was overestimating risks.

² Behavioral science studies are behind the changing pattern of social behavior in face of COVID-19 (Bavel, et al, 2020 ; West et al, 2020). Economic behavior towards market participation when agents do not internalize health effects on others is very relevant to outcomes or optimality (Eichelbaum, et al, 2020; Chang and Velasco, 2020). Simple simulations that assume that people learn and correct behavior show significant differences in outcomes for Argentina (Sturzenegger, 2020). Social attitudes surveys for Argentina (Bozzoli et al, 2020) are preliminary but suggest that economic concerns may affect the way people behave.

³ Due to Kermack and McKendrick (1927); Heathcote (2000) is a rigorous overview. See also Atkeson (2020) and Stock (2020) which were followed by many economics papers; Harris (2020, Appendix) and Fernandez-Villaverde and Jones (2020) for an estimation-oriented representation and Yates (2020) for open divulgation.

interaction with the Susceptible and Infectious groups. Thus, $\alpha I/N$ is the average number of contacts of a susceptible person with the infectious each time period (eg days), and α is the average number of potentially transmissible contacts of one person with another person, while β is the daily share of the total number of days that takes a person to remain infected (for contagious purposes, i.e. is the inverse of the infection period) until it overpasses the contagious window of the disease and γ is the rate at which infected people at time t dies, which is not to be mistaken by the case fatality rate which is the (average or incremental) ratio of deaths per observed cases.⁴

$$\dot{S}_t = -\alpha I_t S_t / N \quad (1)$$

$$\dot{I}_t = \alpha I_t S_t / N - \beta I_t - \gamma I_t \quad (2)$$

$$\dot{R}_t = \beta I_t \quad (3)$$

$$\dot{D}_t = \gamma I_t \quad (4)$$

Econometric estimation

Most of the papers in the recent literature take parameters (α , β , γ) from (deterministic or stochastic) values in order to compute or simulate the evolution of the variables from given initial conditions. In correspondence with these parameters there are associated values of the initial ($R_0 = \alpha/\beta$) and effective (R_t) reproduction numbers, which give a much referred and useful parameter to assess the evolution of the disease, beyond the caveats on the use and abuse of this indicator (Aronson *et al*, 2020; Biggerstaff *et al*, 2014; Delamater *et al*, 2020). Alternatively, parameters of (1) to (4) may be estimated from observed data, given the observable nature of $C_t (=I_t+R_t)$ and D_t . From an econometric perspective there are two ways to proceed with this estimation. The first one is to use non-linear square methods, as done in Batista (2020) and Castle *et al* (2020) which will lead to a forecasted evolution of the SIR model given these parameters or to use those estimates to adjust parameters in simulation exercises. A second alternative, as shown in Harris (2020), is to derive a linearized form of the log of daily cases ΔC_t in order to estimate (by OLS or Poisson regression) the rate of infection α and R_0 (for assumed values of β). This method is quite useful to measure the start-up of the disease transmission and test for the flattening of the curve (as represented by the break in the $\log \Delta C_t$ linear trend), which will normally occur as a result of strict Non Pharmaceutical Interventions (NPI) such as lockdowns. This requires a sufficient number of observations. Beyond that point, given the structural break produced by the NPI, the estimated α or R_t is adjusted to the data but in a different stage to be

⁴ This is a SIRD version made as simple as possible. See Fernandez-Villaverde and Jones (2020) for a richer general SIRD model where they distinguish a “resolved” stage between infections and deaths, thus deaths do not depend on I_t but on a fraction of the resolved group. In the simplest SIR model (eg Harris, 2020) D is not even explicitly modelled although is subsumed in the definition of R_t as Removed from the disease.

defined. If a new stage or wave of the contagious process were to occur, a new testing will have to be performed once enough data is available to adjust the parameters.

Short term forecasting of reported cases and deaths

Forecasting the evolution of observed cases C_t is one of the tasks that has attracted many efforts particularly at the start of the pandemic to predict the intensity of COVID-19 dynamics. Once again, this can be done in several different ways, by simulations with assumed parameters, with a SEIR model based on observed deaths (as in the report by Imperial College, 2020) or by forecasting techniques (Castle et al, 2020). The model used by Imperial College considers that reported deaths are the only accurate observable variable and assumes a 1% stable relationship with true (unobserved) cases and a 21 day period from infection to death with a window of 6.5 days of infections. A death at time t is related to an infection assumed 21 days before and an accumulation of cases (given a consistent R_t and a doubling time period) that start 15 days ($t-15, \dots, t$) before and which will lead to additional deaths ahead ($t+6, t+7, \dots$). Thus, the time series behavior of deaths is the main driver of the model and is embedded in a SEIR model, where E_t stands exposed as a stage between Susceptible and Infected and is an addition needed so as to model the time process of the disease and to allow for some NPI to affect the infectious process (see Castro, 2020 for models for Argentina, and Werning *et al* 2020 in the context of economic optimal lockdown models). In the Imperial College model, forecasts of deaths (and of cases) come from the interaction between the underlying R_t and dynamic of previous cases rather than as a time series forecasting model.

To our knowledge the relative forecasting performance of different methods has not been thoroughly examined; one exception is Castle *et al*, 2020) who have made some comparison of performances. The forecast horizon is usually short for econometric models (week or fortnight) as the data process is subject to shifts which may require re-estimation or updating, particularly if the underlying econometric model has prior specifying restrictions that do not conform to the Data Generation Process (DGP) in some country cases, such as assuming a declining trend for countries that may be well before a given (global or local) disease peak (Liu *et al*, 2020). Short term forecasts are usually needed for updating policy-response decisions concerning the degree of NPI. Imperial College forecast for selected countries has a 28-day window. However, short term forecast do not inform on the overall underlying process (duration, peak, convergence to self-control, effects of an expected vaccine success) of the disease, in spite of being useful for updating such exercises. Longer period forecasts are mostly associated with simulations that describe a convergence process in terms of cases and deaths as it should be expected from a SIR model that has an associated almost natural cycle without intervention. A problem with these forecasts is the existence of cycles in the underlying disease process, a fact that has been well documented in previous pandemics and may well still be a likely

possibility for COVID-19 (Moore et al, 2020). The performance of some forecast models has been openly criticized in the US (Wallace-Wells, 2020).

In this paper we adopt an approach that seek to perform an econometric estimation and short-term forecasting of observed cases and deaths in a way that can be related to structural parameters to the previous SIRD model. Short term forecasts are flexible to accommodate shocks effects along with the variable degree and effectiveness of NPI including the response of agents. This may lead to uncertainty concerning the occurrence time and intensity of a new outbreak of the disease.

Starting from equation (2) and using definitions and equations (3) and (4) above we can write $\dot{I}_t + (\beta + \gamma)I_t = \dot{I}_t + \dot{R}_t + \dot{D}_t = \dot{C}_t = \alpha I_t S_t/N$. Thus, the growth rate of observed cases relates to the infectious rate parameter α as,

$$\Delta \log C_t \cong \frac{\dot{C}_t}{C_t} = \alpha \frac{I_t}{C_t} S_t/N \quad (5)$$

A short term (eg weekly) forecast of $\Delta \log C_t$ is consistent with a forecasted value of α given the relative stability of the computed values of the ratios I/C and S/N over the period. It also relates to a forward looking (rather than backward looking or past observed) doubling time value of cases, given by approximation by $\log(2) / \log(1 + \Delta \log C_t)$, also computed as $\log(2) / \alpha$ at the start up of the epidemic (eg Harris, 2020), given that $C_0 \cong I_0$ and assuming $S_0 \cong N$. Also, by extension, a value of the forecasted effective reproduction rate $R_t = (\alpha/\beta) (S_t/N)$ can be obtained from (5) as in expression (6), using the forecast of $\Delta \log C_t$ as an input and using values for (β) taken from epidemiological studies, which have a wide range of values and is a source of uncertainty (Moore et al, 2020). We have seen papers or presentations where the inverse of β ranges from 3 days (Castro, 2020, for simulations in Argentina); 6.5 days (Harris, 2020, based on Ferguson et al, 2020); 10 or 11 days (Wolfel *et al*, 2020; NCID, 2020) and to 18 (or more) days (most of the economics papers quoted before, based on Atkeson, 2020 which estimate is based on Wang et al, 2020). Thus, R_t estimation is subject to parameter uncertainty due to α and β estimates⁵; use of $\Delta \log C_t$ may be better than assuming α .

$$R_t \cong \frac{(\Delta \log C_t) C_t / I_t}{\beta} \quad (6)$$

⁵ In an exchange about the reliance on estimated parameters (that correspond to econometric or epidemiological estimates) to estimate R_0 , Rodolfo Manuelli suggested to us that there might be in principle, for simulation purposes, a (non-traditional) way to estimate R_0 that avoids using those parameters and is based on observation of infected (I_t) and Susceptible (S_t) people from serological information, which we found could also be applied to the estimation of R_t . This is derived from equation (1) and (2), forming a ratio and using the definition of R_0 , normalizing (or not) $N=1$, as this may be more tractable.

Short term forecasts of reported deaths can also be performed from the evolution of reported cases. From equation (4) and using (5) we can derive the following relation between the rate of growth of observed deaths and observed cases,

$$\Delta \log D_t \cong \frac{\dot{D}_t}{D_t} = \frac{\gamma}{\alpha} \frac{C_t}{D_t} \frac{N}{S_t} \Delta \log C_t \quad (7)$$

A relationship like (7) between $\Delta \log D_t$ and $\Delta \log C_t$ does not have lags as one should expect from the evolution between cases and deaths simply because this is assumed away in writing equation (4) above which does not have lagged values of I_t , neither have non-linear effects of I_t on D_t to capture likely congestion problems in the health system (eg Alvarez *et al*, 2020) which have been well documented at the dramatic startups of Italy and Spain. This makes our estimation different from Imperial College (2020a,b) approach as they (do not postulate a SIRD model like here and) start by modelling deaths and then work from that to a complete SEIR model by taking a relation between deaths and real cases. Instead, a forecast of $\Delta \log D_t$ may start from an estimated dynamic equation that allows for lagged and possibly non-linear effects of observed cases on deaths. Regardless this estimated equation ends up being represented by a growth rate model or by a long-run relationship (which may allow to infer the case fatality rate D/C implicit in the model) an estimated value of the γ can be obtained.⁶

NPI intensity and heterogeneity

Several papers that have modelled optimal lockdown policies (Alvarez *et al*, 2020; Acemoglu *et al*, 2020; Garriga *et al*, 2020) have introduced some important extensions to enrich the epidemiological model for simulation purposes. Apart from the non-linear effect of infections on the death rate already mentioned, these include the effectiveness of NPI such as lockdowns captured by a quadratic expression $(1 - \theta L)^2$ which corresponds to a quadratic matching model specification (Alvarez *et al*, 2020; see also Atkeson, 2020 and Stock, 2020), where L is the degree of the lockdown and θ an unknown parameter capturing effectiveness, with $L \leq \bar{L}$ denoting an upper bound to the lockdown. This term enters into equation (2) to affect the value of the infection rate α and in our formulation is easily introduced in the RHS of equation (5) or in the denominator of equation (6). For empirical purposes $(1 - \theta L)^2$ can be approximated by a mobility indicator M^7 (eg <https://www.google.com/covid19/mobility/>). The use of these type of indicators have been included in the toolkit of some estimation frameworks reviewed before (eg Imperial College, 2020) and are also a way to test the effects of partial or

⁶ The size of γ in relation to α in the initial stages of the pandemic ($N \approx S$) can be inferred with an illustration using (9). Assume that econometric evidence supports that the ratio of $\Delta \log D / \Delta \log C$ is close to one and that the observed D/C ratio is 0.05. Then from equation (9) $\gamma = 0.05\alpha$, ie is 20 times smaller than α . With a $R_0 = 2.5$ and an assumed value for β of $1/6.5$, then $\alpha = 0.38$ and $\gamma = 0.019$.

⁷ Mobility M may be bounded by $\bar{M} \leq M \leq 100$. This indicator dropped across countries from reference (100) values of "normality" to different levels depending on the intensity of the lockdowns and is a natural measure to follow changes in time.

complete withdrawal of NPI restrictions. Mobility indicators are important candidates to model an observed case equation for forecasting purposes, as the effect of changes in M will have a lagged impact on cases.

The issue of mobility is richer as larger datasets and interactions are included in modelling strategies (Monte, 2020; Nguyen *et al*, 2020). Heterogeneity issues come from the fact that a contagion process such as COVID-19 has several dimensions, involving areas and groups among them. Acemoglu et al (2020) develop a model to study, among other things, the consequences on optimal policies of allowing a separation of NPI by age groups, which they show to improve on uniform policies; other papers (Alon *et al*, 2020) have shown that as lockdowns are less effective in developing economies due to structural issues such as informality, these kind of segmentation policies may prove to be fruitful particularly in those countries. Others observers have raised doubts on the ability to segment by age in poor countries (Sturzenegger, 2020) and in Argentina the idea of implementing stringent lockdowns for people above 70 have received strong opposition. We nevertheless consider a useful heterogeneity specification coming from Acemoglu et al (2020) but applied to regions or areas instead of group of persons, which is particularly useful for the case of Buenos Aires City given the uneven performance between poor neighbors and the rest of the City. Simulations with richer data sets may allow broader geographically interacting areas (eg Castro *et al*, 2020). For simplicity and illustration let subindex 1 (2) be a poor (non-poor) neighbor and let ρ_{ij} , $i, j = 1, 2$ be an interaction coefficient within ($i=j$) and between ($i \neq j$) areas, reflecting different contact rates (with possibly $\rho_{ii} > \rho_{ij}$ for $i \neq j$). Then equation (2) is formed by two equations, one for each area,

$$\dot{I}_{1,t} = \alpha(\rho_{11}I_{1,t} + \rho_{12}I_{2,t})(1 - \theta_1 L_{1,t}) S_{1,t}/N_1 - \beta I_{1,t} - \gamma I_{1,t} \quad (8)$$

$$\dot{I}_{2,t} = \alpha(\rho_{21}I_{1,t} + \rho_{22}I_{2,t})(1 - \theta_2 L_{2,t}) S_{2,t}/N_2 - \beta I_{2,t} - \gamma I_{2,t} \quad (9)$$

This leads to equations (5) to (7) one for each area, provided that data is available and parameters ρ_{ij} could be estimated. Artificial neural networks (Castro *et al*, 2020) are being used to approximate these parameters in Argentina, while spatial econometric techniques (eg Baum and Henry, 2020) can be also be useful.

Critical hospitalizations and ITU supply-demand

While equations (1) to (4) do not model explicitly the hospitalization process leading from infections to deaths it is possible to include such a block and in fact there are many models, for analytical and simulation purposes, that do so. Among the ones quoted so far Imperial College (2020b) do model hospitalizations and ICU demand while some optimal lockdown models also do so (Acemoglu et al, 2020 ; Garriga *et al*, 2020) make an explicit distinction between infected people that do or do not go into an ICU. A simple extension in the framework above is to relate the change in deaths not to the level of those infected (as in equation (4)) but to ICU

patients with COVID-19, say $\dot{D}_t = \delta ICU_t$ which is completed adding a relation between the change in ICU patients and the level of infected, say $\dot{ICU}_t = \theta I_t$, although this might be an incomplete way of modelling deaths as they may not be 100 percent ICU mediated (Garriga et al, 2020). As these relationships can be estimated with observed data, the problem we face in our representation below is the poor quality of time series data on hospitalizations and ICU occupancy. Thus, general indicators rather than time series estimates may be used for simulations in those cases.

Dealing with health system capacity limits has also been approached by operative frameworks for assessment and management (Christen *et al*, 2020) or by models within a SIR framework (Charpentier *et al*, 2020), which have separated stages in a (one-wave) COVID-19 process (a second wave does not arise due to conservative lockdown strategy) where ICU capacity saturation along with a stable disease prevalence is part of an optimal path. More simple extrapolation exercises of excess ICU supply given the observed evolution of cases in a lockdown run the risk of overlooking lagged effects and non-linearities, particular in a context of uncertain stages. Lockdowns and the subsequent demand for easing may create a risk of signaling an initial excess of ICU capacity that may be spurious in face of uncertain waves ahead. This is so because ICU excess supply depends not only on the observed growth of ICU demand from COVID (as it is usually assumed) but also on the residual supply defined by the difference between ICU capacity and ICU demand for general (non-COVID-19) patients. The lockdown generates a controlled demand for ICU beds for COVID/19 patients. But lockdowns may also generate a repressed demand for surgical interventions which in turn demand (stochastically) ICU support. Thus a stock-effect may show up later on (under lockdown relaxation or not) which reduces the available capacity to COVID-19 cases, depending of course on the efforts for ICU capacity additions. Another problem is the existence of a seasonal (winter) peak in ICU beds in general (Gartner et al, 2001).

3. COVID-19 evolution representation for Argentina

In the Appendix we describe the data set used in this paper which comes from available official daily reports of observed cases and deaths from the National Ministry of Health and from the Government of Buenos Aires City. Official Ministry of Health data on COVID distinguish among provinces, thus The whole Metropolitan Area of Buenos Aires which includes Buenos Aires City and the Great Buenos Aires (which is aggregated in the data of the province of Buenos Aires) is not available from public sources. Thus a model of Buenos Aires City (CABA) and The Great Buenos Aires (GBA), along the lines of expressions (8) and (9) above cannot be implemented. Still, interactions do exist in the available partitioned data since a (rather stable) percentage of cases reported by Buenos Aires City statistics on COVID are classified as “non-residents” which we assume is mostly explained by people from GBA that come (to be diagnosed and presumably treated) to public hospitals of

CABA. For this reason we use residents in our data for CABA, this is compatible with national Ministry of Health statistics. This is also compatible with internationally assembled data for Argentina as for example in the John Hopkins University, EUCDC, OMS datasets, as well as other global looking sites such as Oxford Our World in Data or Worldometers. There exist however some issues with country-wide data reporting in Argentina in terms of the gaps between daily reports and consolidated data which refers to cases and deaths classified by “date of reporting” creating some daily (not cumulative) discrepancies and dynamics which are small for cases but significant in the case of deaths. We explain these issues in the Appendix. These differences are less important in the case of Buenos Aires City with a better reporting framework, which has now evolved towards detailed data covering neighbors and including testing efforts are positive results as they are located in certain neighbors following a target or priority strategy⁸. Daily recovered cases are reported daily but they do not mean hospital releases, as they may include people in hotels with light or no symptoms (about 50% of total cases). Daily hospitalizations is not explicitly informed but the (aggregate) level of critical cases in hospital is. However, ICU occupancy reporting in for example the province of Buenos Aires has a big (more than 100% in June 18) difference depending of a classification according to confirmed or unconfirmed. Thus, “clean” time series on ICU occupancy (in both public and private hospitals) in CABA and the Province of Buenos Aires are not available.

Figure 1 shows one way to represent the case of Argentina as one of different phases in comparison with the case of Spain which has seen a fast, single-peaked one phase process. Spain has about the same population as Argentina (so absolute values extend to per million values) and strong genetic, cultural and travel ties with Argentina, with Madrid-Buenos Aires being historical a central air corridor. Spain has accumulated almost 10 times the reported cases of Argentina, with 20 times more tests and with 32 times more deaths. Thus, the average case fatality rate in Spain more than 3 times that of Argentina and twice the registered world average. Figure 1 plots the logarithm of daily cases ($\log\Delta C$) in both counties since day 1 of the epidemic, which do not correspond to the same calendar date. This simple representation is, as reviewed above, one way to measure the progression of the contagion, as the trend represent the contagion rate α of a SIR model, estimated in a

⁸ Apart from, or rather than, increasing the effort on testing, Argentina and in particular Buenos Aires as a leading-case in the country, change the “testing policy”, with a priority strategy under the name of “Plan DetectAR” which directed tests towards places where there was evidence of an outbreak. This policy function change produced a sharp increase in positive cases since May, particularly in Buenos Aires City. The impact of this policy function change on the dynamics of testing cannot be evaluated as time series of testing and testing results are not long enough. Nevertheless, even it that were the case it is not straightforward to use test as an explanatory variable for forecasting cases, as tests have risen in reaction to a policy function that reacts to cases. Thus, there is a causality issue that needs to be tested. Doing that in a country like Chile that has not changed much its testing policy, but maintained a much higher effort than Argentina by far, we have found that tests Granger cause Cases, confirming the previous insight.

simple linearized version (eg Harris, 2020). The Figure also includes the dots corresponding to Buenos Aires City. Figure 1 shows two different performances separated by different initial strategies to introduce a lockdown. The delay represented in the case of Spain has been recognized as a key explaining factor that led to a very fast contagion process with a saturation of the health system, explaining Spain position (with Belgium and the UK) in the death per million world ranking. While NPI interventions tend to favor better outcomes across countries (Hsiang et al, 2020) their absence at the start-up of the epidemic do seem to explain very high mortality rates as in the case of Spain much better than arguments based on the unavoidable consequences of herd immunity (Okell et al, 2020). Argentina, on the other hand, triggered a quick and coordinated action and it was due able to flatten the curve rapidly, which to be fair was favored by being in the southern hemisphere and without compromising neighbors (at that time). Estimates of SIR model parameters of the start-up and until the lockdown (Ahumada *et al*, 2020, as in Harris, 2020) showed parameters α of 0.23 for Argentina (which implied a R_0 of 1.3 and a doubling time of 3.5 days) as compared to a value of 0.35 for Spain (with an associated R_0 of 2.3 and a doubling time of 2 days). These values flattened after the lockdown imposed on March 20 in Buenos Aires, while the first intervention in Spain (initiating the “estado de alarma”) reduced the speed of contagion process and was later starting to be controlled. Of course, estimates of R_0 for Argentina differ across studies depending on the model or the assumed (prior) parameters. Castro (2020) and Imperial College (2020b) work for values close to 3 for R_0 while the government has stated values close to 1.3 at about the time of the March 20 lockdown, as shown in some official presentations.

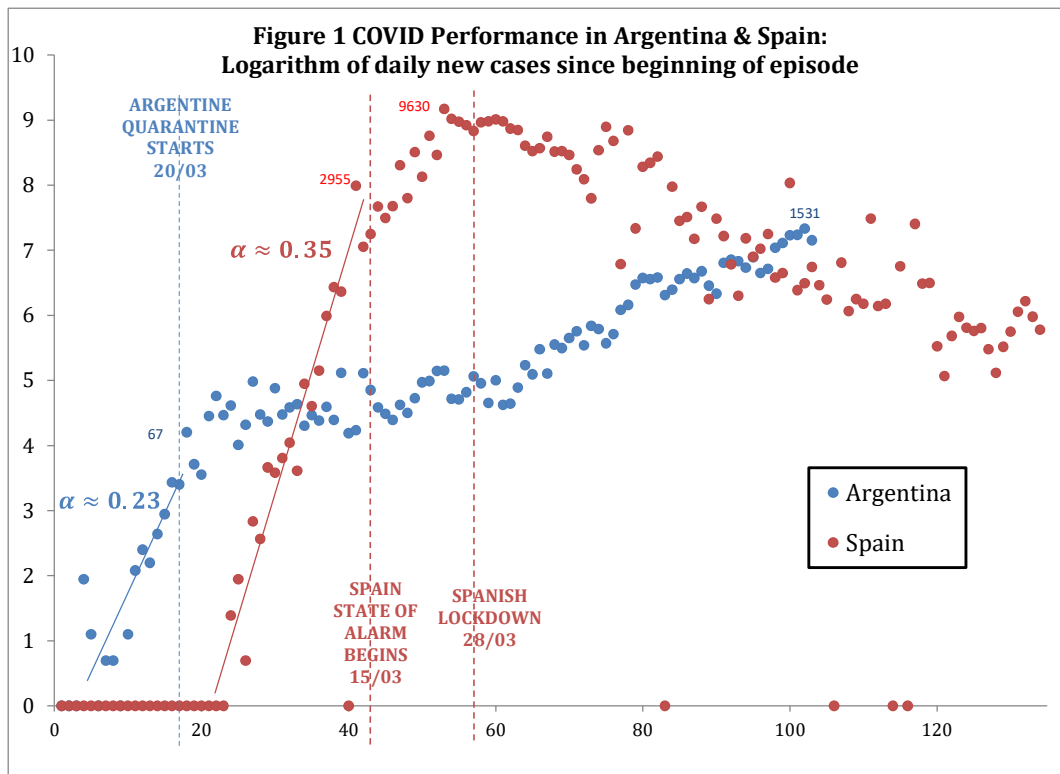
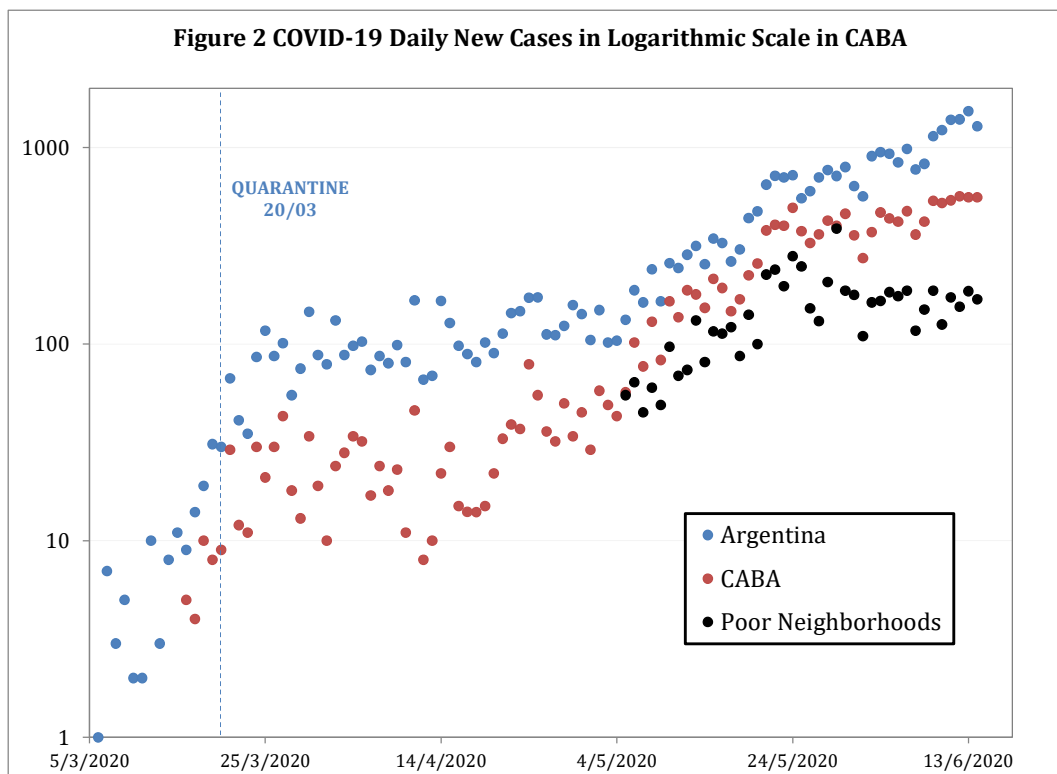
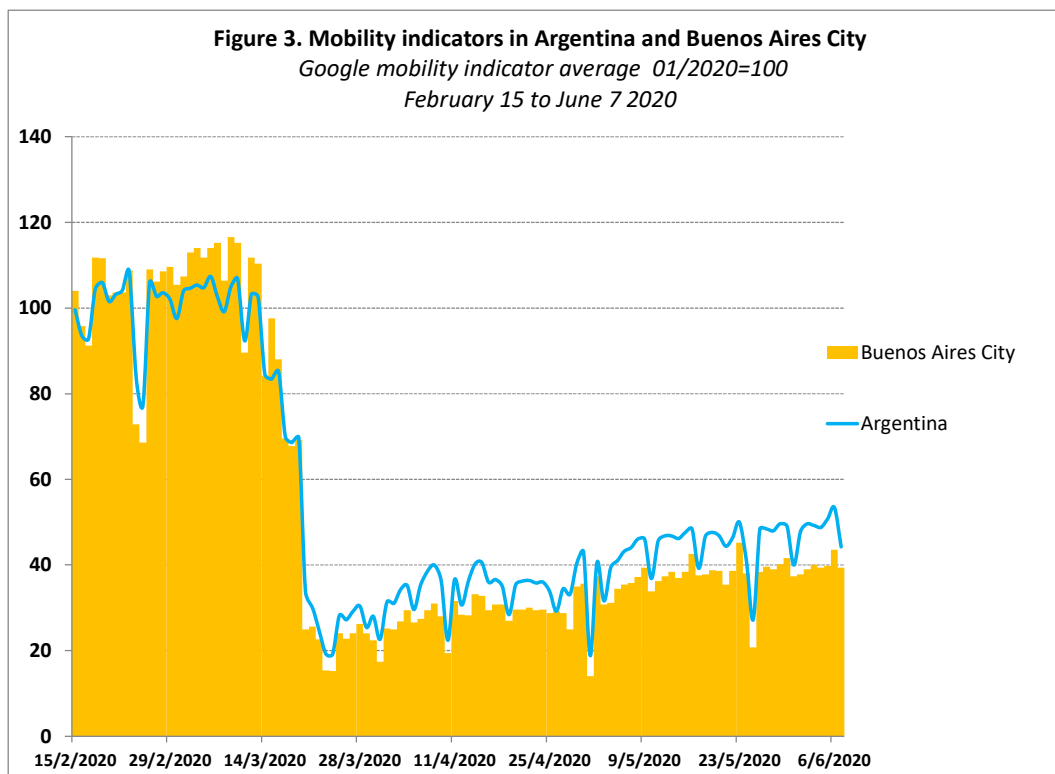


Figure 1 shows that, unlike Spain, Argentina shows clearly a two cycles process separated by a plateau created by the initial effect of the March 20 lockdown. Thus Argentina flattened the curve very early and with less than 100 cases per day but after a period the underlying dynamics of the disease, given the degree of compliance or effectiveness, showed a second wave. This was mostly albeit not exclusively located in Buenos Aires City due to a combination of new case in care homes but particularly in poor neighbors. Figure 2 replicates the plot of daily cases for Argentina along with those of CABA and of poor neighbors in CABA. While parameters estimates of the start-up of COVID-19 in Argentina are best captured using national data (because data has better time coverage and initial cases were less regionally concentrated than today), the plateau and second wave phenomena are well depicted by CABA data. Thus, we use CABA data to represent our forecasting approach of daily cases and deaths. The same can be done without difficulty using national data, Buenos Aires Province data (up to some problems in the quality of daily reporting) and the aggregation of CABA and The Province of Buenos Aires (which fairly approximates, albeit imperfectly, the dynamics in the Metropolitan Area of Buenos Aires). The advantage of the CABA data from daily reports is that it is the less imperfect, can be complemented by weekly reports with auxiliary indicators and finally that, as from the beginning of May, distinguishes poor neighborhoods from the rest given their initial contribution to the second wave of new cases in CABA. The last part of the data shows that the rest of CABA and GBA are providing most of the new cases.



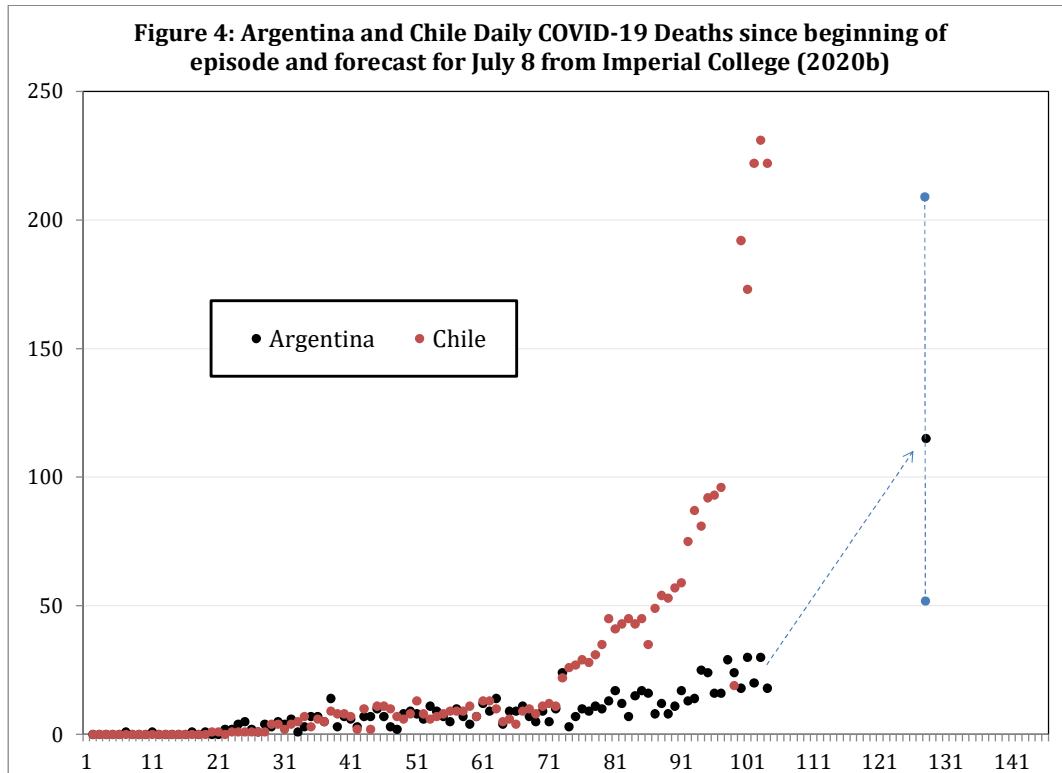
The existence of a sharp second cycle of observed cases within a lockdown call our attention on the problem of the dilution of strict NPI, as it happens with generalized controls, beyond a given time horizon. As publicly available mobility indicators for Argentina are at most disaggregated by provinces have an indicator for CABA on a daily basis. Figure 3 shows the evolution of google mobility aggregate indicator for CABA and Argentina up to the latest available figure (June 7, reported in June 11). It clearly shows a reduction in mobility starting a week before the lockdown of May 20 with a sharp reduction afterwards a plateau in April and a steady recovery since the beginning of May. The Figure is suggestive on the possible role of the relaxation of the lockdown on the move towards the second wave, as shown graphically by President Alberto Fernandez on an interview on June 17⁹, reflecting probably the view of his steering committee. This effect is of course contemplated in the models commented before although its existence and magnitude need to be assessed by the econometric evidence.

⁹<https://www.telam.com.ar/notas/202006/477987-alberto-fernandez-gobierno-nacional-coronavirus-argentina.html>



Turning into COVID-19 death performance Argentina has been showing a very low average case fatality rate (less than 3%, with 21 deaths per million inhabitants, 81th world position), which together with a relatively low occupancy (relative to existing nominal capacity) of ICU made room for optimist assessments of lockdown relaxation even under increasing daily reported cases and deaths. This picture or mood started to change at the time of writing this version, as those indicators, along with observed cases, showed a clear upward movement which led to some debate between CABA and Buenos Aires Province (PBA) Government on going backward on the observed relaxation of the lockdown, with the view of the Federal Government tilted towards PBA. Death forecasts for Argentina from Imperial College (2020) June 9 report tend to validate a scenario where deaths are on the rising. Assuming the same intervention policy, they forecast a more than four-fold increase in deaths from June 10 to July 8, reaching about 115 deaths per day, with a forecast interval very wide (from 50 to 200 cases). This is associated with a similar expected increase in ICU occupancy (which predates deaths) towards 1800 beds. This in turn will stress hospital capacity, depending on the demand of ICU for other uses. The average daily rate of deaths associate with this forecast, at about 5%, is much higher than the observed in recent past data, which is an issue that deserves some scrutiny. An acceleration of ICU hospitalizations and the death rate is not an unlikely event to happen in a relatively controlled situation with a very low case fatality rate (less than 2%) as the recent Chilean experience shows. In fact Argentina and Chile (where Santiago metropolitan area is about half that of Buenos Aires) had about the same

number of observed deaths for a month long period until deaths soared in Chile, as shown in Figure 4.



4- Short-run forecasts of reported cases and deaths in Buenos Aires City

Equations (4) to (6) establish a connection between short-term forecasts of reported cases and deaths with parameters and indicators of a SIRD model. In this section we estimate models, and evaluate their forecasting performance, for reported cases and deaths in CABA. We focus on the model ability to forecast these two main series to follow the disease evolution. Comparing ex-post forecasts and actual data (using pseudo out-of samples) can help to improve models, and thus ex-ante forecasts of these key series, in addition to quantify statistical uncertainty about the evolution of the disease.

4.1 Forecasting reported cases of COVID-19 in CABA

Forecasting changes in daily reported cases of COVID-19 in CABA in the short run (a week-ahead) can be done by estimating simple statistical dynamic models that allow updating and/or rapid break detection. This capability of the econometric modeling is necessary to follow the disease evolution due to different policy interventions, such as the degree of lockdowns and their effectiveness, as well as sudden shocks which could derive in a contagion process acceleration like the observed in the poor-neighborhoods of the city, where conditions for social distancing and infrastructure services such as water provision are deficient. Both kind of model's breaks are

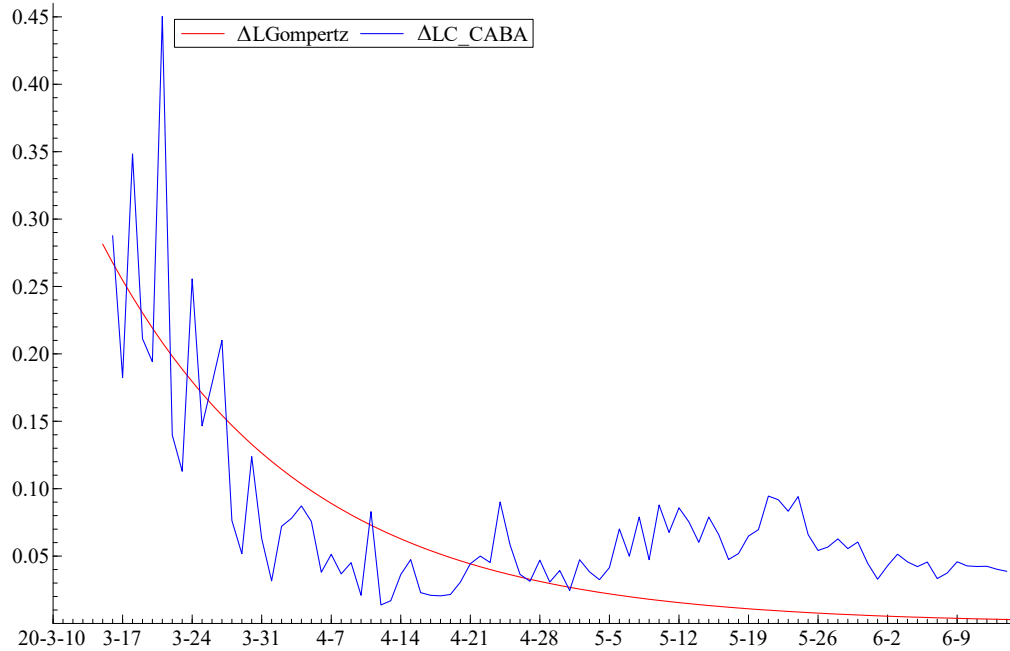
interacting in the case of CABA. To deal with those events and their consequences on the dynamics of reported cases, our approach includes *step saturation* and *impulse saturation* to deal with shifts and outliers and robust forecasts to rapidly adjust our ex-ante forecasts (see Castle et. al, 2015). Impulse saturation (of the form $0,0,0, \dots, 1, \dots 0$ and denoted by “I” below) and step saturation ($1,1,1, \dots, 0, 0, 0$ and denoted by “S1”) are part of an econometric approach that searches for the presence of these dummies for every observation of a given sample. In this way, data themselves are informative about the dummy type and location. This is essential for our purpose to detect changes in the contagion dynamics that can be informative about transitions between stages of the COVID-19 pandemic. Initially developed by Hendry (1999) through sample partitions, this dummy selection approach is part of the *Autometrics* algorithm (Doornik, 2009) that allows to estimate models with more variable than observations. Regarding COVID-19 forecasts, a saturation approach but mainly focus on trend saturation are developed by Castle, et. al. (2020).

We model $\Delta L(C_t)$, a log difference-approximation of the growth rate, given the persistent but shifting behavior (non- stationary) of the C_t series. To compare with the evolution of an epidemiological curve, we simulate a logistic (Gompertz) curve models that assume a single-peak trajectory (see for example Batista, 2020; Sanchez-Villegas, 2020, Lee et al, 2020) starting at the beginning of the contagion process, from which $\Delta L(C_t)^{10}$ was calculated as shown in Figure 5 along with the actual data. We can observe how the lockdown effect in March 20 and the growth jump in May deviate the actual data from the Gompertz-curve growth rate trajectory.

Figure 5 shows what we believe is a fundamental issue in a COVID-19 case such as Argentina. A single peaked process with one significant wave (as in the case of Spain shown in Figure 1) will normally imply a monotonically decreasing growth rate, which will eventually led to some disease control reflected in a reduction of the contagion rate α (see expression (4) above) and a corresponding reduction in the effective reproduction rate R_t (see expression (5)), both at non-pandemic or manageable levels (say $R_t < 1$). Instead, a two waves process may not necessarily have this feature and instead show relatively stable (non-decreasing) values of the growth rate of observed cases at levels which seem to be compatible with non-exponential growth in cases. However, as the process moves on, a given steady rate applied to an increasing stock of cases means a significant increase in new cases which may stress the health system and produce an increasing number of deaths too. This is, we believe, a central ingredient of the argentine case.

¹⁰ Note that ours is a (percentage) growth rate and not the absolute change, which is often defined as “growth rate” in the Gompertz curve literature on COVID-19 (see Utsunomiya *et al*, 2020). While the absolute change is increasing and then decreasing, the percentage rate is monotonically decreasing from its initial peak. See Saez *et al* (2020) for a percentage time series representation applied to testing NPI in Spain.

Figure 5
**Gompertz curve ($\Delta LGompertz$) and actual daily growth rate of observed cases (ΔLC)
in CABA**



Since mean location shifts are the most detrimental break for forecasting (Hendry, 2000) we initially focused on when, and to what extent $\Delta LOG(C_t)$ shifted by step saturation. We use the estimated model to forecast both the daily change (in %) and the daily cumulative (the median level of) cases for the following 7 days. For both we compute confidence intervals (95% for levels and 90% for the rate of change¹¹) Figure 6 plots estimated and actual series, with data used for the initial estimation, after eliminating initial outliers, from March 19, that is 16 days after “patient 0” (March 3) and a day before the lockdown was imposed. Nevertheless, it captures well the high initial daily rates around 20% that contain the inertial elements of the pandemic start-up. Equation 10 shows the forecasting model estimated with data up to April 30, which is a useful starting point to illustrate the approach we are following, before proceeding to test the second upcycle of the infection.

¹¹ Smaller forecast intervals will allow us to better detect forecast biases in the case of growth rates.

$$\begin{aligned} \Delta LC_CABA = & -0.173*\Delta LC_CABA_2 + 0.0466 + 0.13*S:03-27 + 0.0438*S:04-05 \\ (SE) & (0.061) (0.0052) (0.013) (0.0093) \\ & + 0.266*I:03-21 + 0.0568*DI:03-24 \end{aligned} \quad (10)$$

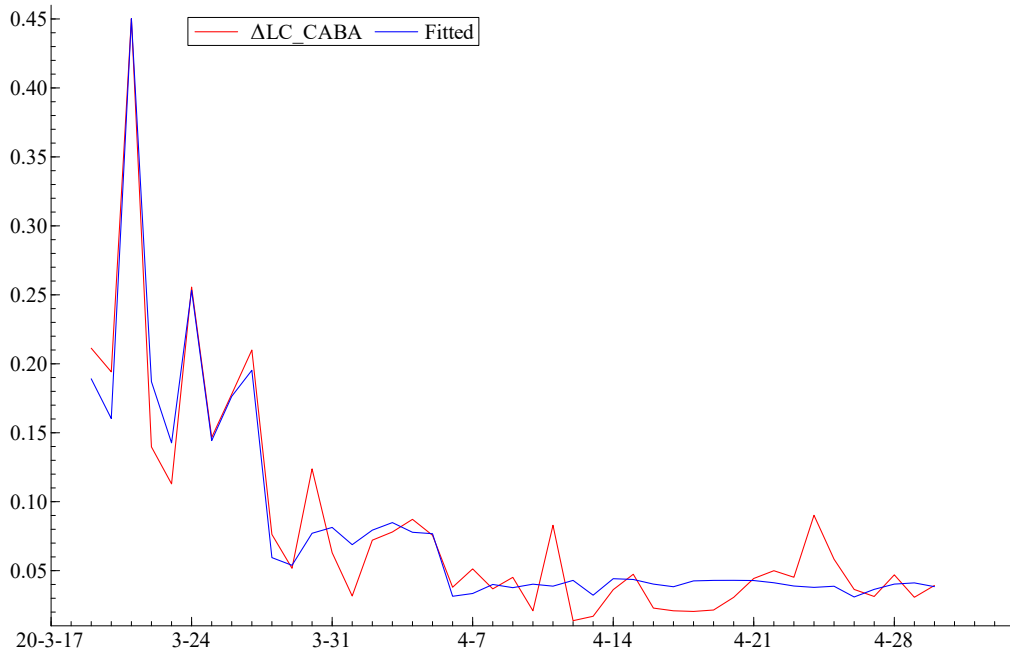
sigma = 0.0223 Adj.R^2 = 0.9276
no. of observations = 43 (from March 19 to April 30)

AR 1-2 test: F(2,35) = 0.66748 [0.5194]
ARCH 1-1 test: F(1,41) = 0.01596 [0.9001]
Normality test: Chi^2(2)= 2.4514 [0.2936]
Hetero test: F(6,35) = 0.48015 [0.8185]
RESET23 test: F(2,35) = 0.66785 [0.5192]

This equation is basically a model of C_t growth rate with an AR(2) term (ΔLC_CABA_2) and impulse dummies (I) for March 21, 23 and 24 (restricted as differences due to similar magnitudes and opposite sign of their coefficient estimates) for large outliers detected by saturation.¹² However, the main feature of the estimated model is the downward shift of the mean growth rate of reported cases, detected by step saturation, 8 days and 14 days after the March 20 lockdown was implemented (i.e. by S1:03-27 and S1:04-05, steps dated March 28 and April 6, as steps are defined as (1,1,1,...,0,0,0)). Although the date of the lockdown is known, the date when the series show the effect is an empirical issue given the epidemic dynamics/inertia, reporting lags and heterogeneous response. We can note also since the steps are defined starting with ones (1,1,1,...,0,0,0) the estimated downward effects are 13% and 4% daily, respectively. That is, the average initial rate, before lockdown, is estimated at 22% approximately (i.e. $0.04657+0.1304+0.04375$) This model has been updated every 7 days performing ex-post evaluation of the forecast performance and generating true out-of sample 7-day ahead forecasts.

¹² We use a 1% target size. Although impulse dummies for 30-03, 11-04 and 24-04 were also selected by the algorithm, no outlier correction was made since they deteriorate the forecasting performance of the model.

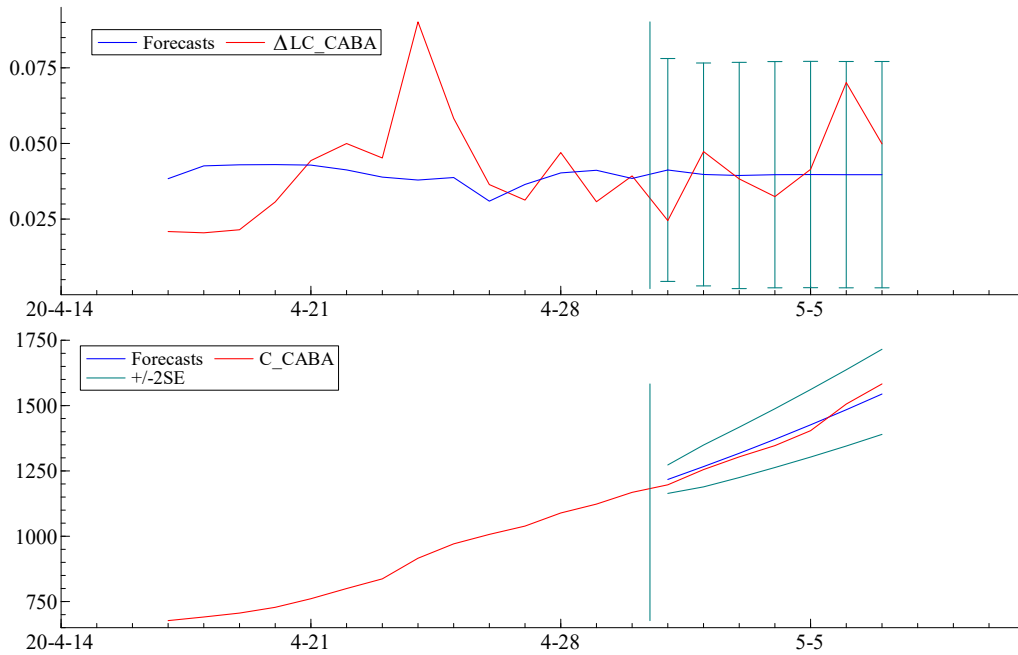
Figure 6
Actual and fitted daily growth rates of reported COVID-19 cases in CABA



The 7-day ahead forecasts from April 30 are shown in Figure 7, where we can observe that forecast and actual values are quite similar, well inside the confidence intervals. However, the two last observations called our attention. In this week new developments happened after reports of water supply shortages in one of the largest poor neighborhoods of CABA (Villa 31). Because of that, we were daily re-estimating the model and follow the daily evolution of forecast errors.

Figure 7

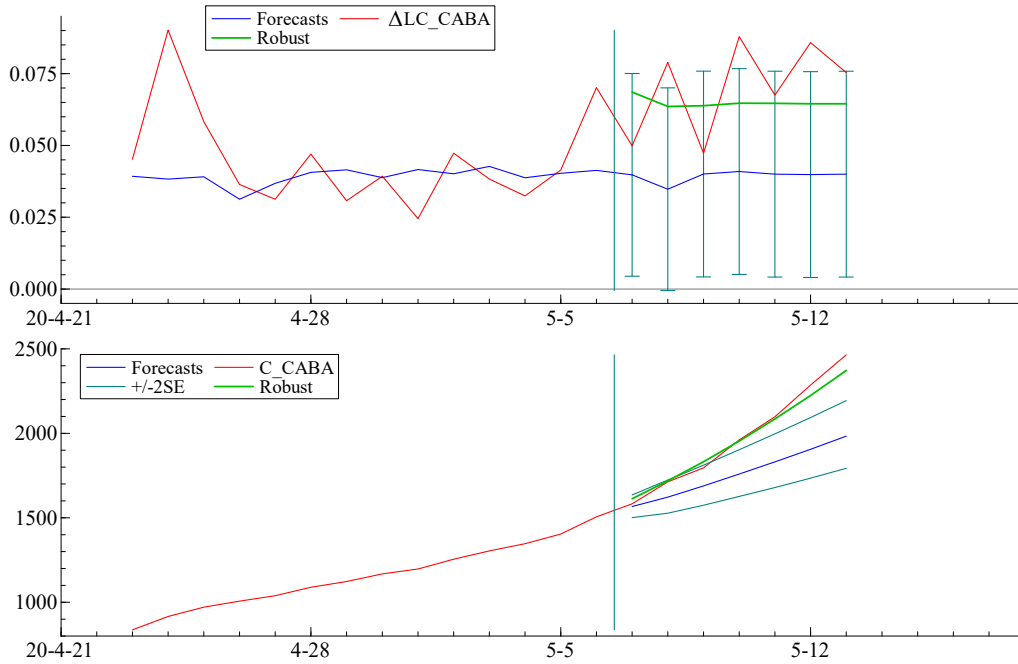
7 day-ahead forecasts of daily growth rates (upper panel) and levels (lower level) of reported COVID-19 cases in CABA- May 1 to May 7, 2020



Until May 6 model estimates were similar to that of equation (10) but systematic biases of forecasts were observed as shown by the blue line in the Figure 8. Beyond the daily arriving information, we were able to learn from the biases quantitatively and put forecasts back-on track, adjusting the forecasts for the initial biases, which greatly improved performance, as the green line in Figure 8 suggested. Robust forecasts could be similarly calculated for the “true” out of sample.

Figure 8

Robust forecasts of daily growth rates (upper panel) and levels (lower level) of reported COVID-19 cases in CABA- May 7 to May 13, 2020



By May 10, evidence of an outbreak in a poor neighborhood (Villa 31) was beginning to be accepted. This coincides with our estimation detecting a step in the growth rate of cases and by extension to the parameters of the SIR model. This estimation is shown in Equation 11 and 7-day ahead forecasts are presented in Figure 9.

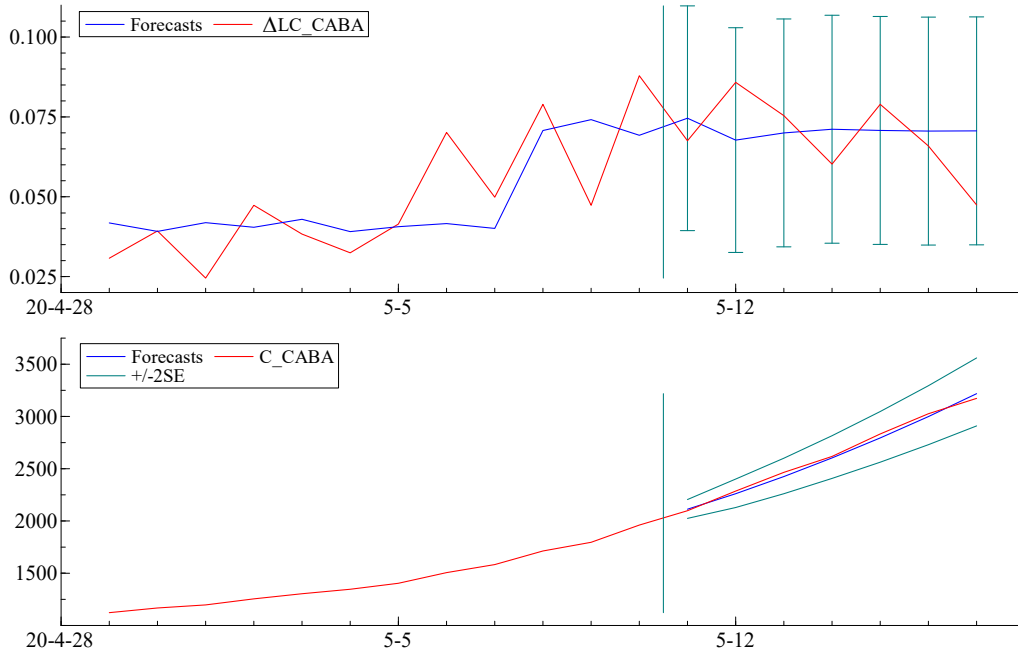
$$\begin{aligned} \Delta LC_CABA = & -0.169 \Delta LC_CABA_2 + 0.0826 + 0.13 * S:03-27 + 0.0428 * S:04-05 \\ (SE) & \quad (0.058) \quad (0.013) \quad (0.013) \quad (0.0087) \\ & + 0.266 * I:03-21 + 0.0568 * DI:03-24 - 0.0355 * S:05-07 \quad (11) \\ & \quad (0.023) \quad (0.015) \quad (0.013) \end{aligned}$$

sigma = 0.0213 Adj.R^2= 0.9213
no. of observations= 53 (from March 19 to May 10)

AR 1-2 test: F(2,44) = 0.47247 [0.6266]
ARCH 1-1 test: F(1,51) = 0.062186 [0.8041]
Normality test: Chi^2(2) = 2.2985 [0.3169]
Hetero test: F(7,44) = 0.58595 [0.7636]
RESET23 test: F(2,44) = 0.78545 [0.4622]

Figure 9

7 day-ahead forecasts of daily growth rates (upper panel) and levels (lower level) of reported COVID-19 cases in CABA- May 11 to May 17, 2020



The model estimates reported in equation 11 show not only a mean rate adjustment (the new estimated constant jumped to 8%) but also a new step. This estimation indicates an upward shift in May 08 (of about 4%), ie a duplication of the daily growth rate. This estimation was significant at 1%, improving the forecasting performance as observed in Figure 9 where outcomes and forecast were very closed again. Values of this forecast translate immediately to policy relevant parameters. Using expression (4) above this allows us (approximating I_t/C_t by its observed value of 0.7) to estimate a duplication of the post lockdown “observed” infection coefficient rate, $(1-\theta L)^2\alpha$, from 0.06 to 0.12 which can be associated (depending of the assumed values of $(\beta)^{-1} \in (6.5, 10)$) with a duplication of R_t from about (0.4, 0.6) to (0.8, 1.2); thus May 8 is a moment where the contagion process, even under a lockdown strategy may be considered as moving above, towards $R_t=1$. The doubling of the growth rate of C_t can also be related to a halving of the doubling time period for observed cases ($\log(2)/\log(1+\Delta LC_t)$) from 18 to 9 days. Finally, given that the mobility indicator shown in Figure 4 jumps from 0.2 to 0.4.

This motivates the inclusion of the mobility indicator in the estimated model of the growth rate of observed cases. This is done in equation (12) with estimates until June 5 and forecasts for an out-of-sample until June 12. For this extended sample, we evaluate how mobility M , affects the growth rate of observed cases along with

the effect of the COVID outbreak in the poorer neighborhoods. The log of the CABA Mobility index 8-day lagged (*LMobilCABA_8*) was significant at 1%. This implies that the 15 point increase in the Mobility Index between March 20 and the end of May (ie from 25 to about 40) added a 3.4% increase in the daily growth rate of cases, that is it explains about 75% of the rate of growth (4.5%) observed at the end of the sample. This is an important effect working against a strategy of slowing down the spread of COVID-19 and at the same time relaxing the lockdown in the current context, unless the policy strategy changes towards testing and isolation more aggressively.

We can note that the step detected for March 28 in the shorter sample (equations (10) and (11) is now incorporated in the mobility effect, which showed a huge downward shift soon after the lockdown started. However, apart from the mobility effect an additional fall in the rate was still observed in early April (by the step *S1:04-05*). Also the upward shift of the growth rate of the confirmed cases in the city after a testing program began in the first poor neighborhood was located in May 6 with a lower coefficient after controlling by mobility. A shift in the opposite direction was found in May 26.¹³

As an additional feature of the model estimated in equation (12), an intraweek effect was also found significant for the more recent sample, due perhaps to the new programs for focused testing, indicating a fall about 1.4 % in reported cases in Sundays and Mondays.

$$\begin{aligned} \Delta LC_CABA = & 0.152*\Delta LC_CABA_3 + 0.0703*LMobilCABAadj_8 - 0.014*dweekday1+7 \\ (SE) & (0.07) & (0.012) & (0.0042) \\ & - 0.214 + 0.0439*S1:04-05 - 0.0153*S1:05-05 + 0.0286*S105-25 & (12) \\ & (0.044) & (0.0083) & (0.0051) & (0.0059) \end{aligned}$$

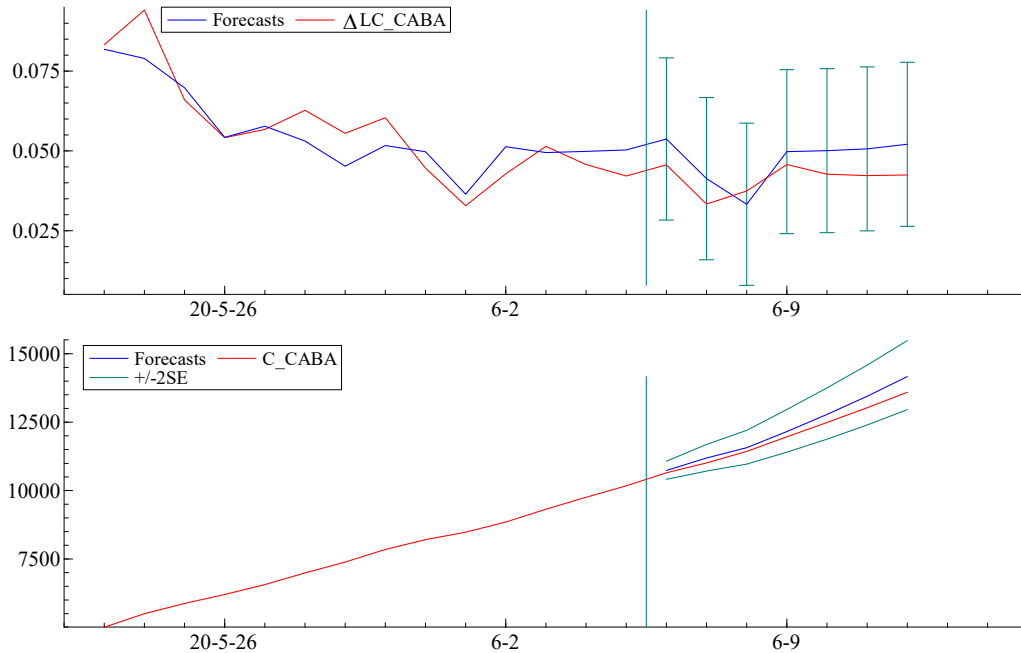
sigma = 0.0154 Adj.R^2= 0.722
no. of observations = 71 (from March 27 to June 05)

AR 1-2 test: F(2,60) = 0.34462 [0.7099]
ARCH 1-1 test: F(1,69) = 0.38095 [0.5391]
Normality test: Chi^2(2) = 2.8439 [0.2412]
Hetero test: F(8,60) = 1.7763 [0.0997]
Hetero-X test: F(9,59) = 1.5869 [0.1403]
RESET23 test: F(2,60) = 7.6782 [0.0011]**

¹³ Two additional impulse dummies were estimated for March 30 and April 24 when the intraweek effect was detected. They responded to non-normality but RESET test now rejected the null. Although it did not affect forecasting performance of the model, nonlinear forms were tried but no one improved the model.

Figure 10

7 day-ahead forecasts of daily growth rates (upper panel) and levels (lower level) of reported COVID-19 cases in CABA- June 06 to June 12, 2020.



As observed in the last figure, forecasts are very close to actual data although small but systematic biases for the last 3 observations, which could be reduced with robust forecasts if the underestimation lasts for following data. They could be associated with a downward shift in the daily growth rate, returning to the proximity of 4.5%. To sum up, our results suggest that even we consider the jump in the growth rate of cases seen in early May mostly as a result of the outbreak in the poor neighborhoods with a later return to the April rate, the lockdown relaxation explains a large part of growth rate of observed cases and the trend in mobility in June is expected to make this effect even stronger.

4.2 Forecasting reported Deaths due to COVID-19 in CABA

To obtain a forecasting model for deaths we started with an autoregressive-distributed lag model for the log of reported deaths on the log of reported cases taking into account the lags reported in other models (see Imperial College, 2020b), i.e. 21 days for the distributed lags. We also included the possibility of intraweek variations by including 7 lags in the case of reported deaths. We used *Autometrics* to select the relevant lags along with dummy saturation. Then, the model was rewritten in log differences according the estimates, as shown below in equation (13).

$$\begin{aligned}
\Delta LDeathsCABA = & -0.154*\Delta LDeathsCABA_6 - 0.175*\Delta 2LDeathsCABA_3 + 0.501*\Delta LC_CABA_16 \\
(SE) & (0.052) \qquad\qquad\qquad (0.035) \qquad\qquad\qquad (0.081) \\
& + 0.47*\Delta LC_CABA_19 - 0.0312*S1:04-23 + 0.0376*S1:05-08 \\
& (0.08) \qquad\qquad\qquad (0.011) \qquad\qquad\qquad (0.0084) \\
& - 0.0143 + 0.0185*S1:05-27 - 0.0248*dweekday7 \qquad\qquad\qquad (13) \\
& (0.009) \quad (0.0097) \qquad\qquad\qquad (0.0088)
\end{aligned}$$

sigma = 0.0238 R^2 = 0.859
no. of observations = 62 (from April 5 to June 5)

AR 1-2 test: F(2,51) = 0.67802 [0.5121]
ARCH 1-1 test: F(1,60) = 0.097826 [0.7555]
Normality test: Chi^2(2) = 0.64676 [0.7237]
Hetero test: F(12,49) = 0.34812 [0.9749]
Hetero-X test: F(18,43) = 0.69984 [0.7921]
RESET23 test: F(2,51) = 2.4747 [0.0942]

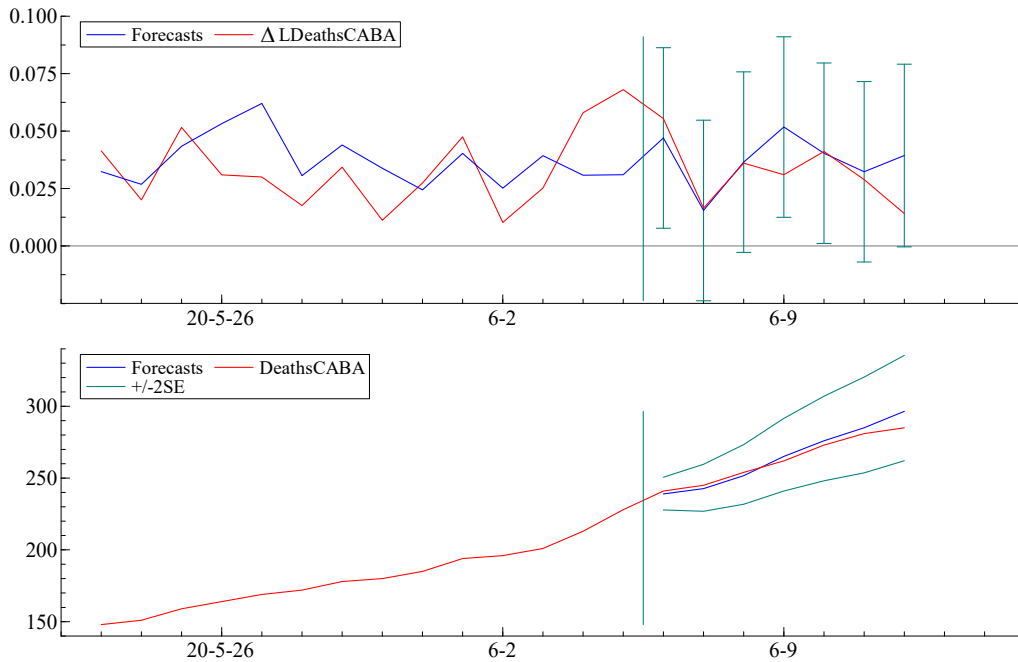
We can observe that the lag between reported cases and deaths is a bit shorter than the one assumed in Imperial College (2020b): we detect one effect after 16 days and another after 19 days. The fact that the two lagged (16 and 19 days) coefficients of the rate of growth of cases are about 0.5 shows that the rate growth of deaths is about half of that of cases 16 to 19 days before. This is consistent with the monotonic reduction in the average case fatality rate observed in the sample.

Results also show the dynamic effects of deaths and a Sunday effect, which would be associated with the intraweek behavior of reported cases. Three different steps change the constant mean growth rate of deaths, an upward shift during April that approximately reverted in May and even a lower rate at May 28.¹⁴ Figure 6 shows the forecasting performance for the estimated models where actual and 7 day ahead forecast are very close in the forecasted (second) week of June, despite the in-sample differences.

¹⁴ We can note that the shorter lag of 16 days was previously obtained when the model was estimated until May 10 (only 36 observations).

Figure 11

7 day-ahead forecasts of daily growth rates (upper panel) and levels (lower level) of reported COVID-19 deaths in CABA- June 06 to June 12, 2020



5. Summary of results and final remarks

The uncertainty surrounding the extension, occurrence and intensity of waves of the COVID-19 pandemic has implications for the way we assess the current position of a given country and, although this uncertainty strikes all countries differently depending on their own trajectory and their position in the worldwide transmission of the disease, it seems to be present in all cases. This would come as no surprise to epidemiologists that have used historical references about likely patterns that may emerge, and where second waves have been present both in the so called “Spanish Flu” pandemic of 1914-18 but also in less well known cases such as the 1957-58 pandemic (Moore et al, 2020; see also Jefferson and Heneghan, 2020). These second waves are not necessarily seasonal (winter) related and may be seen across and within countries. Waves do not seem to be generated ex ante as cycles of a richer epidemiological model but instead seem related to different shocks to the pandemic. Perhaps this seems so because such sort of analytical models have not yet been understood and agreed upon, or perhaps waves are of secondary importance today (not so in the past) because the innovation process and the vaccine should be the main concern as it acts upon current NPI that shape the dynamics of the disease.

We argue that short term forecasting of cases and deaths may become more useful than thought because of waves of infection and can be done in a complimentary way to an epidemiology model that has now become the norm of research that serves for policy guidance even when the final position of the pandemic cannot be forecast

with accuracy. We show that short term forecasts of the rate of growth of cases and deaths of COVID-19 can be done in a way that it relates to key parameters of the SIR model and its variants. We believe this is useful for both policy evaluation and dialogue concerning measures such as the relaxation or hardening of NPI and for assisting calibrated models and their simulation to use parameters that better conform to actual data. As richer data sets allow, the approach can accommodate heterogeneity across areas and groups, mobility and spatial interactions, and the performance of the health system.

We applied our estimation and forecasts of the growth rate of cases and deaths to the data of Buenos Aires City (CABA) which has an initial start-up in early March followed by a lockdown on March 20 that mitigated and reduced the contagion process and then was followed by a second wave in May initially associated with poor neighborhoods and then becoming more widespread until mid-June when our sample ends. Data shows that cases outside CABA and in the metropolitan area belonging to Buenos Aires Province are becoming now the main focus of the disease dynamics, but we do not study that process as the CABA case is enough to illustrate our modelling approach. We use saturation techniques in a time series automatic selection modelling to choose different steps in the growth rate of cases and deaths in this sub-periods with weekly forecasts and model re-estimation once shifts are detected in the data. Our results for the daily growth rate of observed cases show a rate of about 4% (down from the 20% peak at the start of the pandemic) after the introduction of the lockdown in March 20 that becomes very stable throughout April, with forecasts showing no need of re-estimation. Then our forecast performance detects a jump of the growth rate in May 6 which leads to a duplication of the growth rate to 8% but prove to be transitory and is associated with cases in poor neighborhoods and a third phase where the model and forecast stabilize at around 4.5%. This rate is consistent with a doubling time rate of 16 days, a contagion rate coefficient $\alpha = 0.07$ and a reproduction rate R_t clearly less than one assuming values of β^{-1} lower than 14 days. On the other hand, a stable growth rate of deaths, which we find to be about 70% of that of cases, or 3% although it is lower than the one implicit in Imperial College (2020b) for Argentina, it may become less comfortable as it implies an increase from about 30 deaths per day in mid-June to 84 in mid-July and to 236 in mid-August, the latter being quite problematic for the existing management of UCI, even ignoring spillovers from a similar increase in the Province of Buenos Aires. We also find evidence that observed cases have an effect on deaths between 16 and 19 days later, which is a bit lower to the 21 days assumed in Imperial College (2020b) forecasts.

Thus, one would ponder, where is the trouble with the Argentine strategy? In fact our approach shows that there is one, and it unfolded as a consequence of the visibly early but still only partial success in containing the pandemic. The transitory acceleration and sustained growth rate of observed cases after the introduction of a 12 week lockdown is, we believe, a key feature or stylized fact of the Argentine case

as represented by CABA data. In Ahumada *et al* (2020) while estimating a 75% reduction of the contagion rate of a SIR model from 0.23 as shown in the data before the lockdown, we registered an inconsistent message in the announcement of an estimated peak of the pandemic in June since this left open to question the evolution of the cases in the following weeks, as the lockdown would lose its initial effectiveness. Stable growth rate of cases, single digit deaths, a falling case fatality rate from 5% to less than 3% (as testing efforts and new cases in young out-of-peril inhabitants of poor neighborhoods drove the ratio of deaths to cases down) and ample ICU beds capacity in hospitals could lead to a “satisficing” strategy. In turn, recent performance shows how problematic this may become as the social and economic pressure for lockdown softening mounts over the weeks, with a policy response that accommodates as much as it can to these demands (that are replicated in many other countries). Instead of a monotonic reduction in the growth rate of cases and deaths, as it should be predicted by a convergence process describing in a Gompertz-type curve, stable rates even if they display non accelerating pandemic rates, may raise absolute daily cases, hospitalizations and deaths. All this may become very costly in terms of the health system capacity in the short to medium term as the Chilean experience shows.

As the sustained growth rates raised absolute daily observed and expected numbers to less and less comfortable values, a policy dialogue instance concerning the strategy towards the relaxation of the lockdown has occupied the center stage at the time of writing this version. Our results contribute to this policy debate by showing that the decline in effectiveness of the lockdown as proxied by a mobility indicator has had an important effect on the growth rate of cases. Thus, the country is facing a tradeoff between health and economic/social outcomes. An increase (a relaxation) of the current level of mobility (lockdown) at about 40% in Buenos Aires City to, say, 60% would imply an increase in the growth rate of cases of about 3.5% (from 4.5% to 8%) which maps into an increase in the growth rate of deaths of about 2.5% (from 3% to 5.5%), creating a highly stressed scenario. On the other hand, a reduction of mobility to a level of 25% which represents the level observed in late March after the lockdown announcement would imply a reduction of the growth rates of cases and deaths towards 2.5% and 1.8% which would be more consistent with controlling the evolution of the pandemic, but will come with a second wave of economic contraction. The country needs to overcome this hard choice because it is not even clear that a way back to March is implementable and a return to a constrained economic activity may trigger extensive margin (participatory) decisions by firms, particularly in the service sector with serious effects on employment levels. This seems to require implementing rapidly a new more aggressive strategy on tracing and isolation.

References

- Ahumada H, S. Espina-Mairal and F. Navajas (2020), "Modelos y pronósticos de la dinámica del COVID-19: Mirando la Argentina", mimeo, FIEL, Abril 10.
- Alon T., M. Kim, D. Lagakos and M. Vuren (2020), "How Should Policy Responses to the COVID-19 Pandemic Differ in the Developing World?", Working Paper 27273, National Bureau of Economic Research, May.
- Alvarez F., D. Argente, and F. Lippi, (2020) "A Simple Planning Problem for COVID-19 Lockdown," Working Paper 26981, National Bureau of Economic Research, April.
- Aronson J., J. Brassey and K. Mahtani (2020), "When will it be over?": An introduction to viral reproduction numbers, R_0 and R_e ", Centre for Evidence-Based Medicine, Oxford University.
- Atkeson, A. (2020) "What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios," Working Paper 26867, National Bureau of Economic Research, March
- Batista, M. (2020). "Estimation of the final size of the coronavirus epidemic by the SIR model", *MedRxiv*, Feb.28, <https://doi.org/10.1101/2020.02.16.20023606>.
- Barnett M., G. Buchak and C. Yannelis (2020), "Epidemic responses under uncertainty", Working Paper 26867, National Bureau of Economic Research, May.
- Baum, C. and M. Henry, "Socioeconomic Factors influencing the Spatial Spread of COVID-19 in the United States", May 29, <https://ssrn.com/abstract=3614877>
- Bavel et al (2020), "Using social and behavioural science to support COVID-19 pandemic response", *Nature Human Behavior*, April 30, <https://www.nature.com/articles/s41562-020-0884-z>
- Biggerstaff, M., Cauchemez, S., Reed, C. *et al.* (2014) "Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature". *BMC Infectious Diseases* 14, 480 <https://doi.org/10.1186/1471-2334-14-480>
- Bozzoli G, M. Grandes et al (2020), "Índice de propensión al riesgo en salud IPRIS", Fundación Bunge y Born, Buenos Aires, Junio.
- Charpentier A., R. Elie, M. Laurière and V. Tran (2020), "COVID-19 pandemic control: balancing detection policy and lockdown intervention under ICU sustainability", *MedRxiv*, May 22 <https://doi.org/10.1101/2020.05.13.20100842>
- Christen P., K. Hauck et al (2020), "Strengthening hospital capacity for the COVID-19 pandemic: J-IDEA pandemic hospital planner", Report 15, MRC Centre for Global Infectious Disease Analysis, Imperial College London
- Castle, J. L., Clements, M. P., & Hendry, D. F. (2015), "Robust approaches to forecasting", *International Journal of Forecasting*, 31(1), 99-112.
- Castle J., J. Doornik and D. Hendry (2020), "Short-term forecasting of the Coronavirus Pandemic 2020-07-27", University of Oxford, UK, April 27
- Castro R. et al (2020), "Simulación de escenarios basados en modelos matemáticos para Covid19 en Argentina", Coloquio del Departamento de Matemática, FCEyN, UBA. Mayo de 2020
- Chang R. and A. Velasco (2020), "Economic Policy Incentives to Preserve Lives and Livelihoods", DP14614, Center for Economic Policy Research, April.
- Delamater P. E. Street et al (2019), "Complexity of the Basic Reproduction Number (R_0)", *Emerging Infectious Diseases*, 25, 1, January, <https://doi.org/10.3201/eid2501.171901>
- Doornik, J. A. (2009), "Autometrics". In J. L. Castle & N. Shephard (eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Oxford: Oxford University Press

- Eichenbaum M., S. Rebelo, and M. Trabandt (2020), "The Macroeconomics of Epidemics," Working Paper 26882, National Bureau of Economic Research, March.
- Ferguson N et al (2020), "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand", Imperial College, March 16. <https://spiral.imperial.ac.uk:8443/handle/10044/1/77482>
- Fernández-Villaverde, J. and C. Jones (2020), "Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities," Working Paper 27128, NBER, May.
- Galiani S. and I. Stambulsky (2020), "La industria hotelera podría ser la puerta para salir de la cuarentena sin que se disparen los contagios", Infobae, <https://www.infobae.com/opinion/2020/05/10/la-industria-hotelera-podria-ser-la-puerta-para-salir-de-la-cuarentena-sin-que-se-disparen-los-contagios/>
- Garriga, C., R. Manuelli and S. Sanghi (2020), "Optimal Management of an Epidemic: Lockdown, Vaccine and Value of Life", April 27, latest version: https://sidsanghi.github.io/website/OptimalManagement_COVID_GMS.pdf
- Garfield M., S. Ridley et al (2001), "Seasonal variation in admission rates to intensive care units", *Anaesthesia*, 56, pp. 1136-40
- Gonzalez-Eiras, M. and D. Niepelt, "On the Optimal "Lockdown" During an Epidemic", CEPR Discussion Paper No. DP14612. April. Available at SSRN: <https://ssrn.com/abstract=3594242>
- Harris J. (2020), "The Coronavirus Epidemic Curve is Already Flattening in New York City", MIT, March 30, NBER Working Paper 26917.
- Hendry, D. F. (1999). "An Econometric Analysis of US Food Expenditure, 1931-1989", Chapter 17, in Magnus, JR and Morgon, MS (eds.), *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, Wiley & Sons.
- Hendry, D. F. (2000), "On detectable and non-detectable structural change", *Structural change and economic dynamics*, 11(1-2), 45-65.
- Hendry D. (2020), "A Short History of Macro-econometric Modelling", Nuffield College Oxford, January 20, https://www.nuffield.ox.ac.uk/economics/Papers/2020/2020W01_MacroHist18.pdf
- Hendry, D. F., and J. A. Doornik. *Empirical model discovery and theory evaluation: automatic selection methods in econometrics*, MIT Press, 2014.
- Hethcote H. (2000), "The Mathematics of Infectious Diseases", *SIAM Review*, 42, 4, pp. 599-653.
- Hsiang, S., Allen, D., Annan-Phan, S. et al. (2020), "The effect of large-scale anti-contagion policies on the COVID-19 pandemic", *Nature*, <https://doi.org/10.1038/s41586-020-2404-8>
- IHME (2020), "Forecasting the impact of the first wave of the COVID-19 2 pandemic on hospital demand and deaths for the USA and 3 European Economic Area countries", *MedRxiv*, doi: <https://doi.org/10.1101/2020.04.21.20074732>
- Imperial College London (2020a), "Short-term forecasts of COVID-19 deaths in multiple countries", MRC Centre for Global Infectious Disease <https://mrc-ide.github.io/covid19-short-term-forecasts/index.html>.
- Imperial College (2020b), "Situation Report for COVID-19: Argentina, 2020-06-09", downloadable from <https://mrc-ide.github.io/covid19-short-term-forecasts/index.html>.

- Jefferson T. and A. Heneghan (2020), "COVID-19 Epidemic 'Waves'", Centre for Evidence-Based Medicine, University of Oxford, April 30, <https://www.cebm.net/covid-19/covid-19-epidemic-waves/>
- Kasy M. and A. Teytelboym (2020), "Adaptive Targeted Infectious Disease Testing", May 12, *Oxford Review of Economic Policy*, <https://doi.org/10.1093/oxrep/graa018>
- Kermack, W.O., and A.G. McKendrick. (1927, 1991). "A contribution to the mathematical theory of epidemics - I." *Bulletin of Mathematical Biology* (reprinted) 53 (1-2):33-55.
- Lee Y., B. Lei and B. Mallick (2020), "Estimation of COVID-19 spread curves integrating global data and borrowing information", *MedRxiv*, May 19 <https://doi.org/10.1101/2020.04.23.20077065>
- Martinez E., D. Aragon and A. Nunes (2020), "Short-term forecasting of daily COVID-19 cases in Brazil by using the Holt's model", *Journal of the Brazilian Society of Tropical Medicine*, Vol.:53, doi: 10.1590/0037-8682-0283-2020, May 20.
- Moore K., M. Lipsitch, J. Barry and M. Osterholm (2020), *The Future of the COVID-19 Pandemic: Lessons Learned from Pandemic Influenza*, Part 1, The CIDRAP viewpoint, Center for Infectious Disease Research and Policy, University of Minnesota.
- NCID (2020), "Period of Infectivity to Inform Strategies for De-isolation for COVID-19 Patients", Academy of Medicine, Singapore, May 23. [https://www.ams.edu.sg/view-pdf.aspx?file=media%5c5556_fi_331.pdf&ofile=Period+of+Infectivity+Position+Statement+\(final\)+23-5-20+\(logos\).pdf](https://www.ams.edu.sg/view-pdf.aspx?file=media%5c5556_fi_331.pdf&ofile=Period+of+Infectivity+Position+Statement+(final)+23-5-20+(logos).pdf)
- Okell LC, R Verity, O Watson *et al.* (2020) "Have deaths from COVID-19 in Europe plateaued due to herd immunity?". *Lancet*; June 11. [http://dx.doi.org/10.1016/S0140-6736\(20\)31357-X](http://dx.doi.org/10.1016/S0140-6736(20)31357-X).
- Panadeiros M. (2020), "Coronavirus en CABA ¿No tan mal como parece?", mimeo, Seminario FIEL, Mayo 24.
- Pereira I. et al (2020), "Forecasting Covid-19 dynamics in Brazil: a data driven approach", *MedRxiv* doi: <https://doi.org/10.1101/2020.05.11.20098392>. May 18
- Saez M, A. Tobias, D. Varga and M.A. Barcelo (2020), "Effectiveness of the measures to flatten the epidemic curve of COVID-19. The case of Spain", *Science of The Total Environment*, 727, <https://doi.org/10.1016/j.scitotenv.2020.138761> April
- Sanchez-Villegas P, Codina AD, (2020) "Modelos predictivos de la ´ epidemia de COVID-19 en Espana con curvas de Gompertz", *Gaceta Sanitaria* <https://doi.org/10.1016/j.gaceta.2020.05.005>
- Stock J. (2020), Stock, James H, "Data Gaps and the Policy Response to the Novel Coronavirus," Working Paper 26902, National Bureau of Economic Research March
- Sturzenegger F. (2020), "COVID-19: vidas vs. economía, reflexiones y estimaciones", Academia Nacional de Ciencias Económicas, Junio 5.
- Urbiztondo S. (2020), "Covid-19: Una actualización de la evolución de la pandemia en Argentina, y de los sensores para la flexibilización del aislamiento social", *Indicadores de Coyuntura*, FIEL, Junio.
- Utsunomiya YT, Utsunomiya ATH, et al (2020), "Growth Rate and Acceleration Analysis of the COVID-19 Pandemic Reveals the Effect of Public Health Measures in Real Time". *Frontiers in Medicine*, 7:247. doi: 10.3389/fmed.2020.00247
- Yates C. (2020), "How to model a pandemic", *The Conversation*, March 25.
- Wallace-Wells (2020), "What the Coronavirus Models Can't See", *New York Magazine*, May 2.

West R., S. Michie, J. Rubin and R. Amlot (2020), "Applying principles of behaviour change to reduce SARS-CoV-2 transmission", *Nature Human Behavior*, May 6, <https://www.nature.com/articles/s41562-020-0887-9>

Wölfel, R., V.M. Corman, W. Guggemos et al. (2020), "Virological assessment of hospitalized cases of coronavirus disease 2019", *MedRxiv*, <https://doi.org/10.1101/2020.03.05.20030502> , March 8.

Appendix: Data sources and issues

For reference and comparisons (eg Spain) country-level data for reported (coNPIrmed) cases, deaths and recovered are taken from Johns Hopkins University¹⁵, and are subject to the database's updates and revisions. For Argentina we employ the National Ministry of Health's (MSAL) Daily Press Releases¹⁶ as our main data source, which reports the aggregate, national as well as the provinces. The decision to use this data at a national level instead of Johns Hopkins' arises due to significant lags in reporting in the case of the latter, especially at the beginning of the pandemic episode. Thus, data for CoNPIrmed Cases, Deaths and Recovered is based on the MSAL daily reports. The only exception for this is the recovered people data previous to 23/03/2020, where Johns Hopkins data is used. It must be noted that province-level data is based on a jurisdictional criterion and not on a detection or treatment site one. In other words, confirmed cases are assigned to the province of residence, rather than contagion site.

From these daily press reports, we also extract data on critical cases and PCR tests when available. Detailed data on CABA is also provided by daily reports from the City Health Department¹⁷: cases, deaths and recovered are also separated between residents of CABA and non-residents. Provincial-level data on PCR testing, critical cases and detailed accounts of poor neighborhoods is also provided, although with a limited time span so far. Buenos Aires Province detailed data regarding testing, critical cases and so on, is not available to date in time series format.

We must note that national-level data are not fully consistent with provincial-level data, and constant omissions, errors and vague inter-provincial categorizations in daily reports mean some noise must be accounted for, especially regarding provincial-level time series on deaths where no accumulated data is published by the National Ministry of Health to date. We also use national-level data on PCR testing from Our World in Data¹⁸ for countries where available.

Issues regarding deaths reporting

We deliberately make use of data on death counts from the National Ministry of Health (even though disaggregated time series are not published) because we find the series published by the European Centre for Disease Prevention and Control (ECDC)¹⁹ -which is the source used by Imperial College (2020) to be noisy and inconsistent with data reported by local sources and by Johns Hopkins, at least for the case of Argentina.

This being said, even "first hand sources" such as the National Ministry of Health present several issues regarding consistent reporting on death tolls. A daily updated "Situation Room"²⁰ reports aggregate national data including a nationwide time series of cumulative deaths according to "reporting date". This series is significantly and consistently different from published records on daily press reports. Although methodological notes are not available and no specification is made regarding what "reporting date" truly means, figures suggest the series might be *backwardly-corrected* (and daily updated) according to the

¹⁵ <https://github.com/CSSEGISandData/COVID-19>

¹⁶ <https://www.argentina.gob.ar/coronavirus/informe-diario>

¹⁷ <https://www.buenosaires.gob.ar/coronavirus/noticias/actualizacion-de-los-casos-de-coronavirus-en-la-ciudad-buenos-aires>

¹⁸ <https://ourworldindata.org/coronavirus-testing>

¹⁹ <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>

²⁰ <https://www.argentina.gob.ar/salud/coronavirus-COVID-19/sala-situacion>

“true” dates of reported deaths attributed to COVID-19. This reveals that the reporting of daily deaths is subject to serious lags: delays involving classification and reporting of these deaths result in differences between both time series of up to 500% in levels and first differences, and up to 2.5 relative ratios in terms of daily growth rates. The magnitude of these discrepancies raises warnings regarding the use of daily deaths data to produce short-term forecasts in deaths-driven models such as Imperial College (2020) unless a well informed criteria allows for adjustments.

For data on cases, instead, this lag effect is much smaller (in several orders of magnitude) than for deaths data, at least regarding the Ministry of Health official reports to date. Time series on daily cases based on a symptom’s start criterion (rather than a reporting day criterion) are frequently published, but the very nature of their required *backwards-correction* renders them practically useless for short-term forecasting, as several days and usually weeks must pass before data becomes stable to constant correction and updating.