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

Hildegart Ahumada   Fernando Navajas

(jointly with Santos Espina Mairal)

Seminario del Instituto de Economía Aplicada, ANCE

Julio 31, 2020

Paper available at http://dx.doi.org/10.2139/ssrn.3633500

Motivation

- Basically methodological, seeking to combine data-driven and model-driven approach to assess evolution (Ahumada *et al*, April 10, 2020)
- Most likely hypothesis: “cicles/waves/stages” of unknown emergence, magnitude, duration due to uncertainty along with mitigation and social response (best shown in Moore *et al*, April 30 2020) until a vaccine arrives.
- Short term forecast of new cases (Castle *et al*, March 19, 2020) with a stochastic trend (like a “weather forecast” model). A dynamic forecasting approach with no reference to epidemiological model parameters.
  - A rival forecasting model would come from a non-linear estimation of an epidemiological model (eg Batista, January 2020)
- Alternative to non-linear estimation: Take-off/Flat-out estimation with a “linearized” static (OLS, Poisson) estimation of the contagion rate of a SIR model (Harris, March 30 2020). Not for forecasting but linked to epidemiological model parameters
- Contribution: We propose a short term forecast strategy of cases and deaths but related to parameters of a SIRD model.

Motivation

- Critical dimensions to distinguish: Short-Term, Data-Driven, Model-Related, Parameter/Distribution Shifts.
  - Uncertainty of process suggests short term/data driven, policy dialogue suggests model related. Estimation by saturation techniques to accommodate shifts essential (Hendry, 2000; 2020)
- Some epidemiological models (Imperial College, 2020) do perform short-term forecasts for many countries, with errors. Some overestimation of deaths for Argentina, so far.
- Medium range models (IHME, 2020) or Gompertz-(logistic) curve models forecasts do not survive to shifts and need re-estimation. IHME on Argentina: scary, 28k deaths 1/11
- Dialogue with economic+epidemilogical models (Alvarez *et al*, 2020; Garriga *et al*, 2020; Acemoglu *et al*, 2020; Gonzalez Eiras and Niepelt, 2020) as they are source of several insights and effects, we add to parameters choices from observed data.

COVID-19: The CIDRAP Viewpoint

April 30th, 2020

Part 1: The Future of the COVID-19 Pandemic: Lessons Learned from Pandemic Influenza

Kristine A. Moore, MD, MPH
Marc Lipsitch, DPhil
John M. Barry, MA
Michael T. Osterholm, PhD, MPH

Dr. Moore is medical director of the Center for Infectious Disease Research and Policy (CIDRAP). Dr. Lipsitch is the director of the Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard TH Chan School of Public Health. John Barry is a professor at Tulane University School of Public Health and Tropical Medicine. Dr. Osterholm is director of CIDRAP, University of Minnesota Regents Professor, and McKnight Presidential Endowed Chair in Public Health.

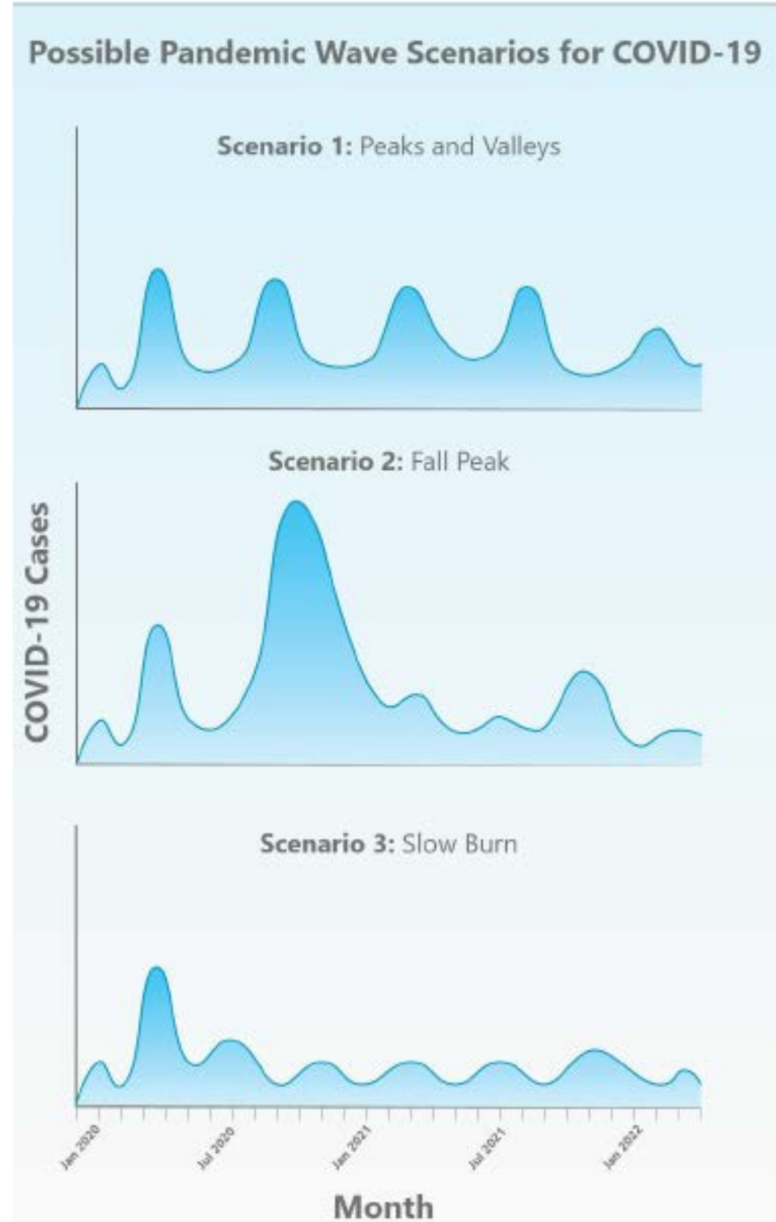
CIDRAP, founded in 2001, is a global leader in addressing public health preparedness and emerging infectious disease response. Part of the Academic Health Center at the University of Minnesota, CIDRAP works to prevent illness and death from targeted infectious disease threats through research and the translation of scientific information into real-world, practical applications, policies, and solutions. For more information, visit: www.cidrap.umn.edu.

COVID-19 Viewpoint reports are made possible with support from the University of Minnesota Office of the Vice President for Research and the Bentson Foundation.

COVID-19: The CIDRAP Viewpoint working group:
Michael T. Osterholm, PhD, MPH, CIDRAP director
Kristine A. Moore, MD, MPH, CIDRAP medical director
Julie Ostrowsky, MSc, CIDRAP research associate
James Seifert, JD, MS, MPH, CIDRAP program manager
Angela Ulrich, PhD, MPH, CIDRAP research associate
Alison Kraigsley, PhD, MS, CIDRAP research associate
Maya Peters, MPH, CIDRAP program analyst
Jim Wappes, CIDRAP editorial director

Editing: Jim Wappes; Report Layout and Design, Cover Design: Hannah Winesburg

CIDRAP
Center for Infectious Disease Research and Policy
University of Minnesota



simple SIRD model

- Kermack and McKendrick (1927); Heathcote (2000)
- 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 $N_t=S_t+I_t+R_t+S_t$ and C=I+R+D.
- Equations illustrate the transitions 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**.
- $\dot{S}_t = -\alpha I_t S_t / N$ (1)
- $\dot{I}_t = \alpha I_t S_t / N - \beta I_t - \gamma I_t$ (2)
- $\dot{R}_t = \beta I_t$ (3)
- $\dot{D}_t = \gamma I_t$ (4)

Getting (α, β, γ)

- Parameters (α, β, γ) used to compute or simulate the evolution of the variables from given initial conditions, from epidemiological studies.
- In correspondence are associated values of the initial ($R_0=\alpha/\beta$) and effective (R_t) reproduction numbers
- But many different forms to estimate R (Aronson *et al*, 2020; Biggerstaff *et al*, 2014; Delamater *et al*, 2020) make cross comparisons tricky. Own evolution of a given form preferable reference.
- 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.

Econometric estimation

- The first one is to use non-linear square methods, as done in Batista (2020) and Castle *et al* (2020).
- 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 the 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) as a result of lockdowns.
- This requires a sufficient number of observations. Beyond that point, given the structural break, the estimated α or R_t is adjusted to the data but in a different stage to be defined.
- If a new stage or wave of the contagious process were to occur, a new testing will have to be performed (expost) once enough data is available to adjust the parameters.

"Linearized" estimation of α

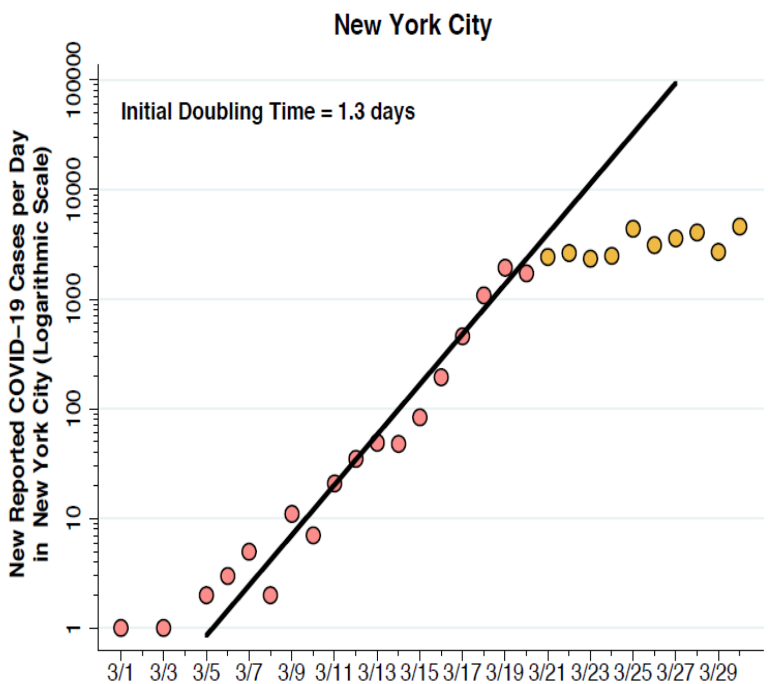
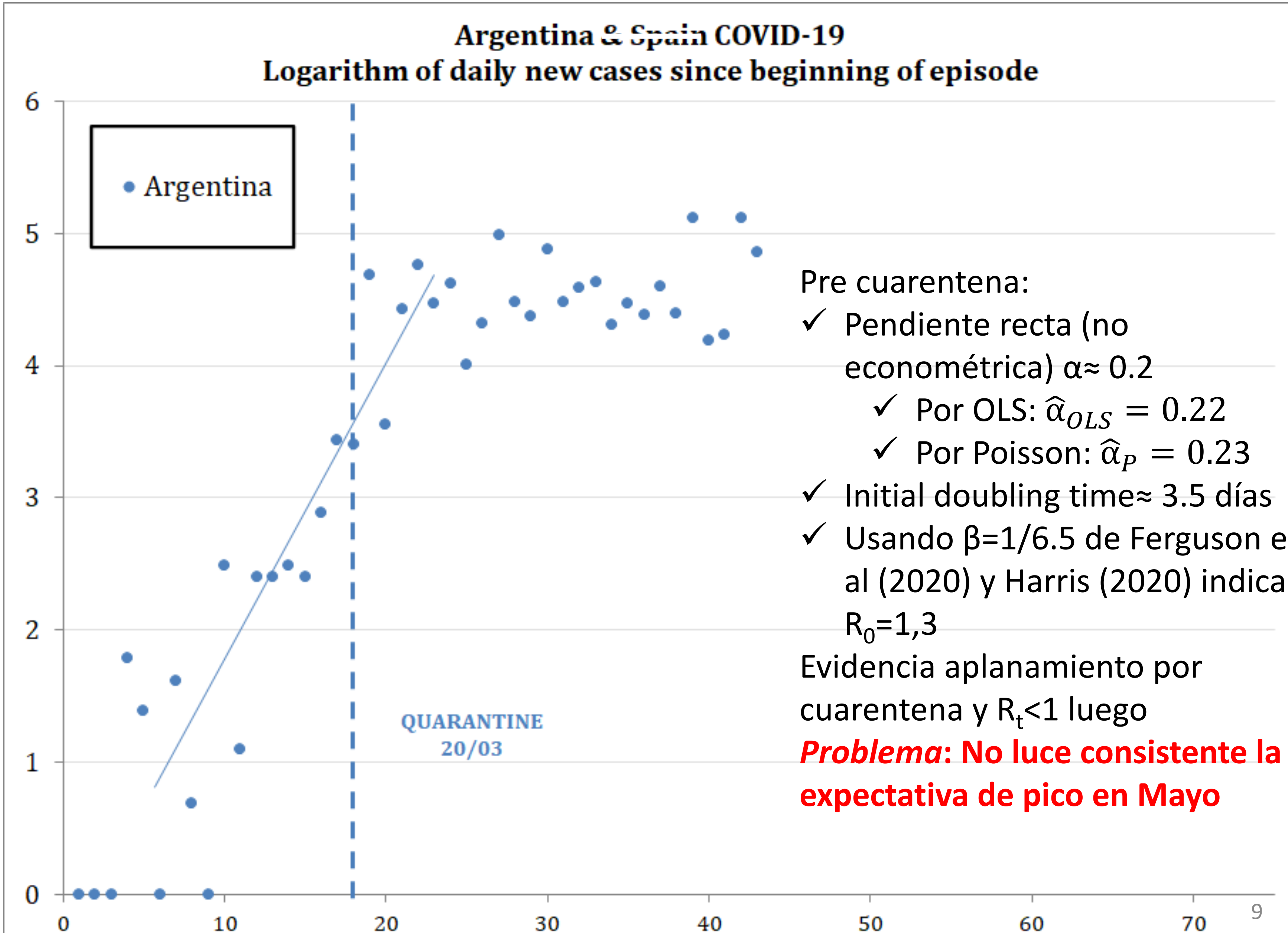


Figure 1. New Reported COVID-19 Cases per Day in New York City (Logarithmic Scale). Best-fit trend line (Poisson regression) and estimated doubling time based upon observations through 3/20/2020, shown as pink data points. Observations from 3/21/2020 onward shown as mango data points.

- C_t observed cases
- $C_t=I_t+R_t+D_t$
- $\dot{C}=\alpha I = \alpha C \ (at \text{ "0"})$
- $C_t = I_0 \exp(\alpha t)$
- $\dot{C} = \alpha . I_0 \exp(\alpha t)$
- $log \dot{C}_t = log I_0 + \log(\alpha) + \alpha t$
- "α" is estimated from a linear regression of log C

Argentina up to April 15 in Ahumada *et al* (2020)



Pre cuarentena:

- ✓ Pendiente recta (no econométrica) α≈ 0.2
  - ✓ Por OLS: $\hat{\alpha}_{OLS} = 0.22$
  - ✓ Por Poisson: $\hat{\alpha}_{P} = 0.23$
- ✓ Initial doubling time≈ 3.5 días
- ✓ Usando β=1/6.5 de Ferguson et al (2020) y Harris (2020) indica R_0=1,3

Evidencia aplanamiento por cuarentena y R_t<1 luego

***Problema*: No luce consistente la expectativa de pico en Mayo**

Short term forecast of observed cases

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

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

- **A short term (eg weekly) forecast of $\Delta \boldsymbol{log C_t}$ 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 doubling time value of cases, given by approximation by $\log(2)\,/\log(1 + \Delta log C_t)$

Short term forecast of R_t

- An implicit value of the forecasted effective reproduction rate $R_t = (\alpha/\beta)\,(S_t/N)$ can be obtained from (5) using the forecast of $\Delta \log C_t$ as an input and using values for (β) taken from epidemiological studies

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

- 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 are based on Atkeson, 2020 which estimate is based on Wang et al, 2020).

Short term forecast of deaths

- From equation (4) and using (5) we can derive the following relation between the rate of growth of observed deaths and observed cases,

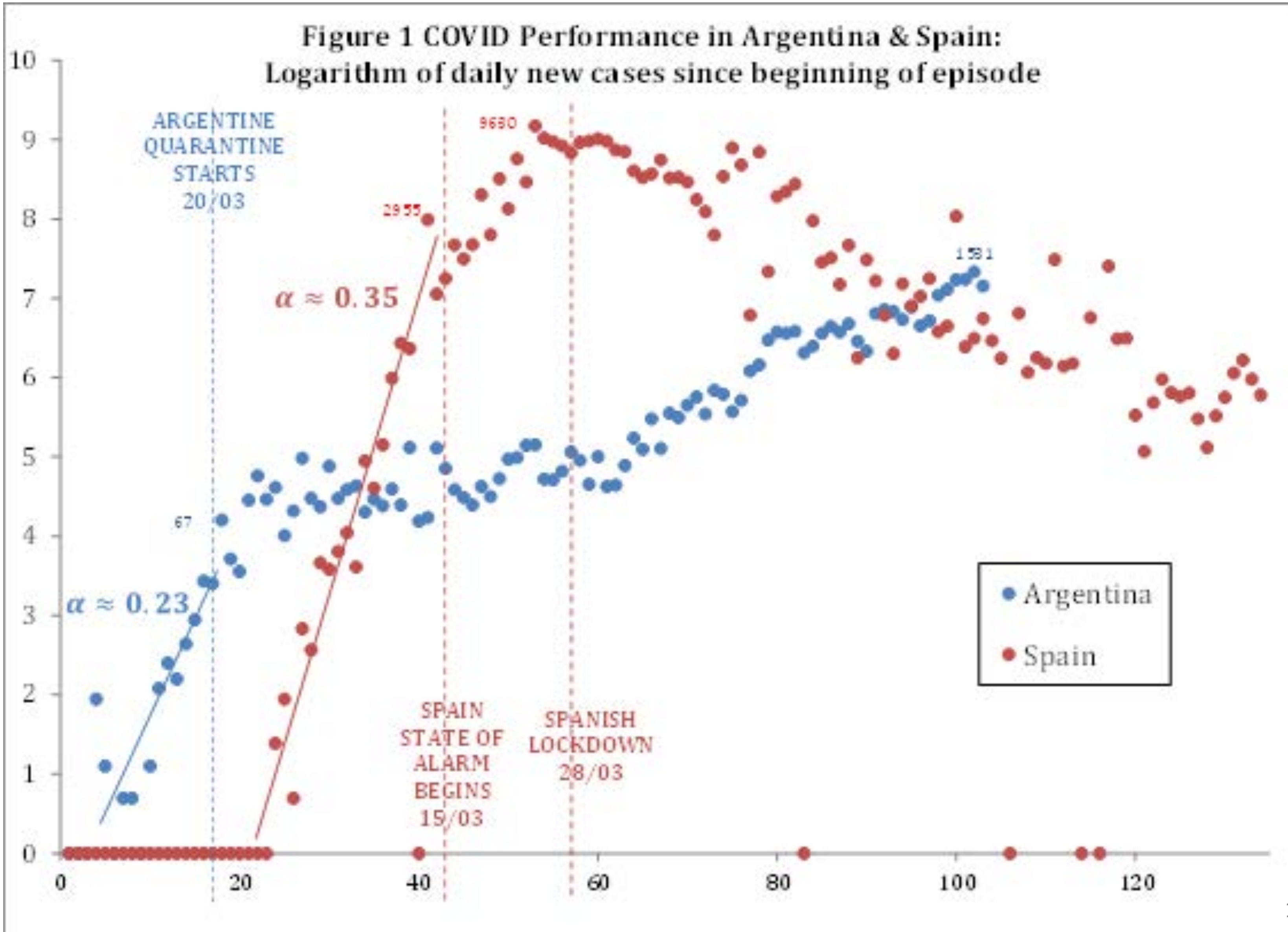
$$\Delta log D_t \cong \frac{\dot{D}_t}{D_t} = \frac{\gamma}{\alpha} \frac{C_t}{D_t} \frac{N}{S_t} \ \ \Delta log C_t \qquad (7)$$

- Lags should be expected but here are assumed away due to the simple model
- Non-linear effects of I_t on D_t to capture likely congestion problems in the health system (eg Alvarez *et al*, 2020) have been well documented at the dramatic startups of Italy and Spain.
- Estimates of γ can be obtained from (7) given an estimated relation between D and C, and the estimate of α

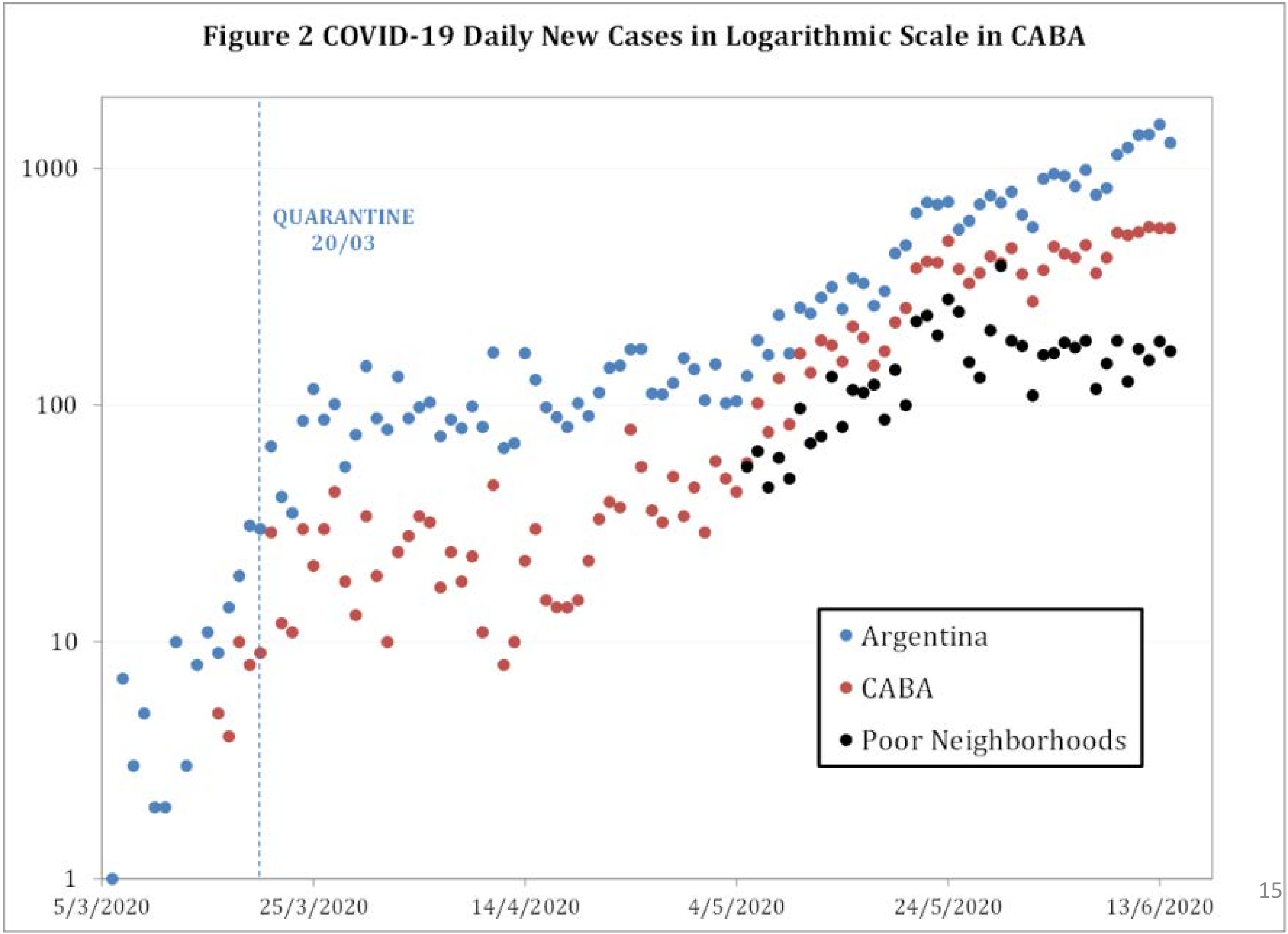
Mobility to capture the effect of NPI

- The effectiveness of NPI such as lockdowns are captured by a quadratic expression $(1-\theta L)^2$ which corresponds to a quadratic matching model specification (Alvarez et al, 2020) where L is the degree of the lockdown and θ an unknown parameter capturing effectiveness, with $L \leq \bar{\bar{L}}$ denoting an upper bound to the lockdown.
- This term enters equation (2) to affect the value of the infection rate α and is easily introduced in the RHS of equation (5) or in the denominator of equation (6).
- For empirical purposes it can be approximated by a mobility indicator M (https://www.google.com/covid19/mobility/).
- 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.

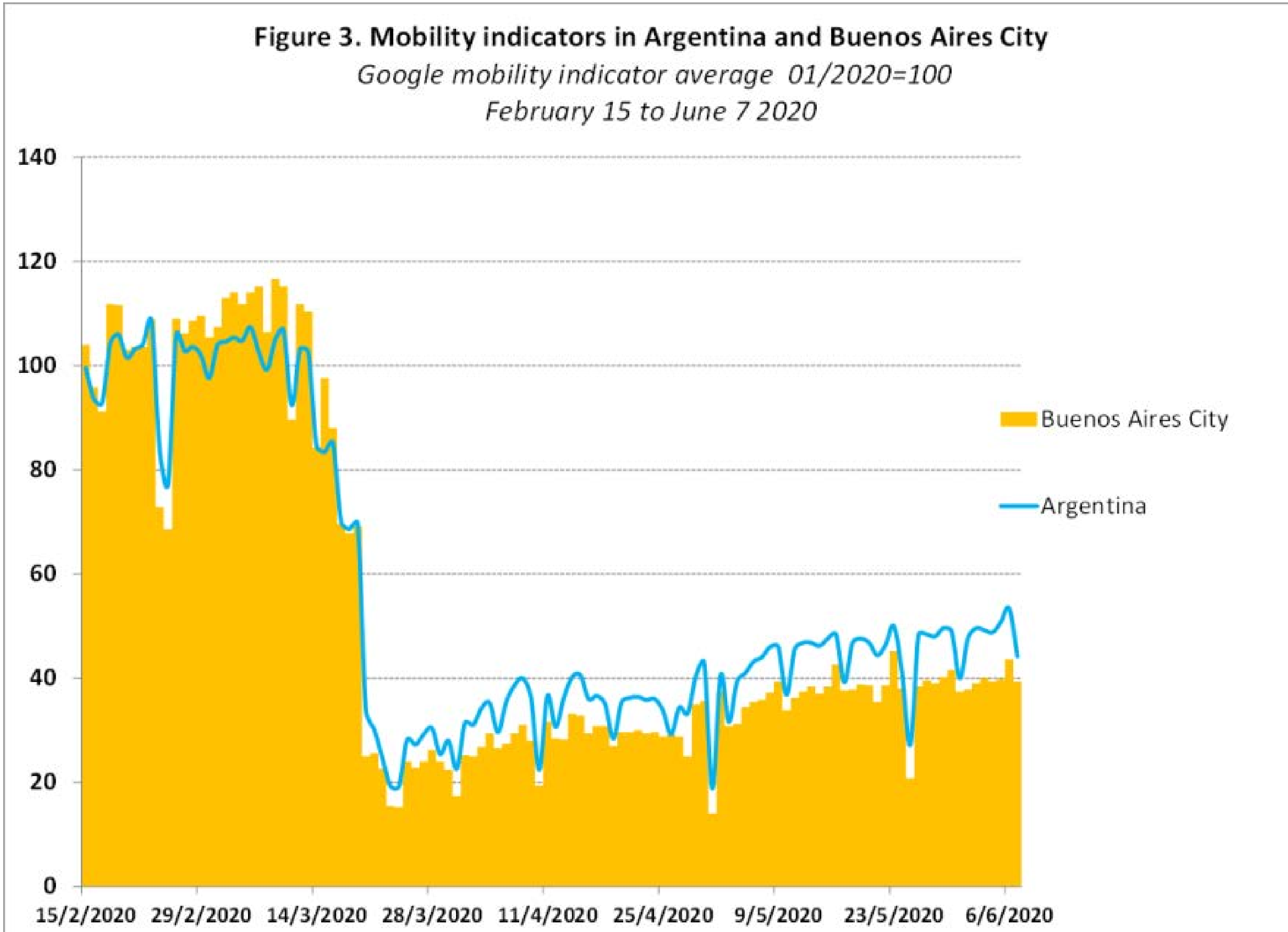
Performance representation for Argentina



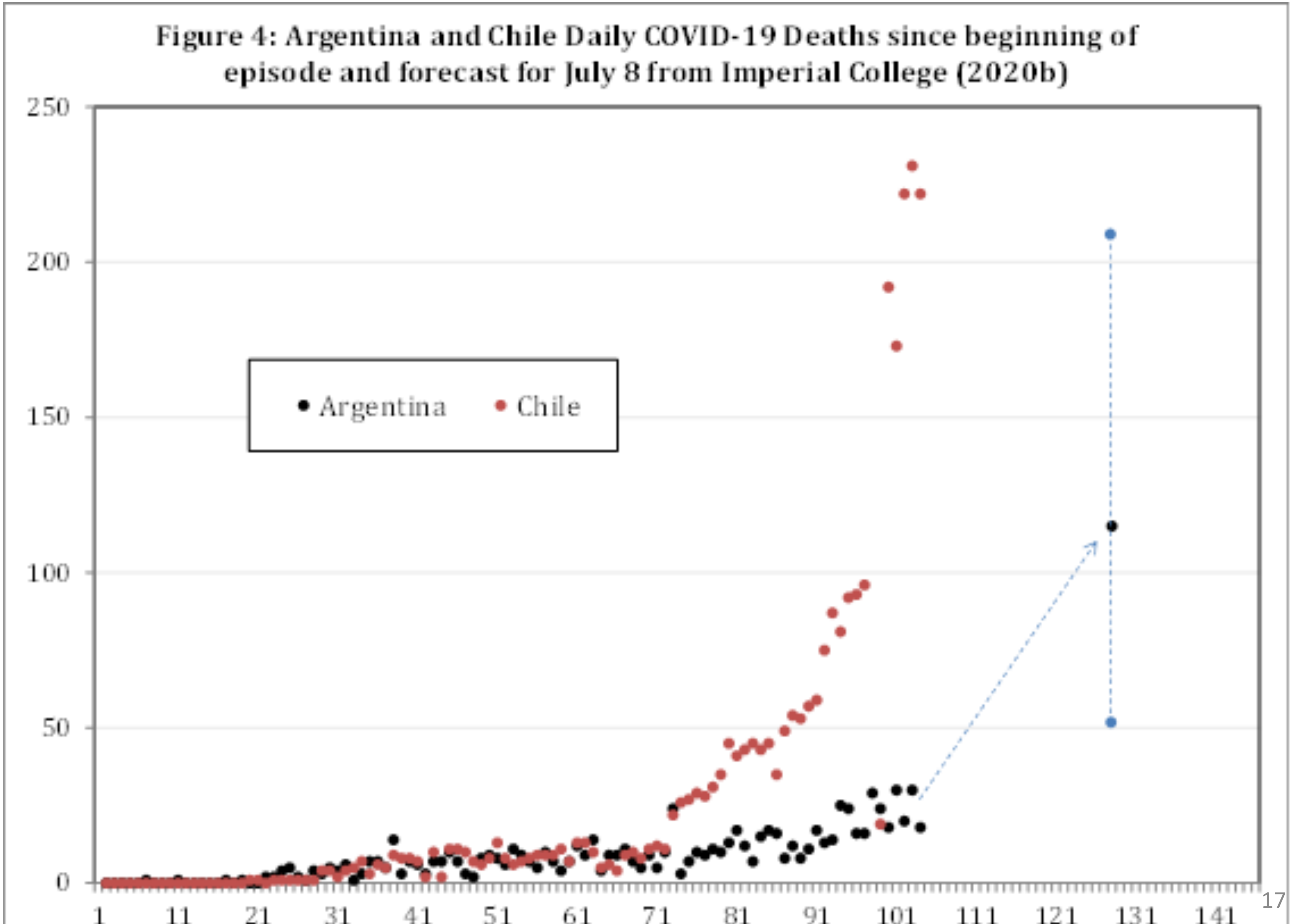
CABA as an illustration



Mobility in CABA



Death performance

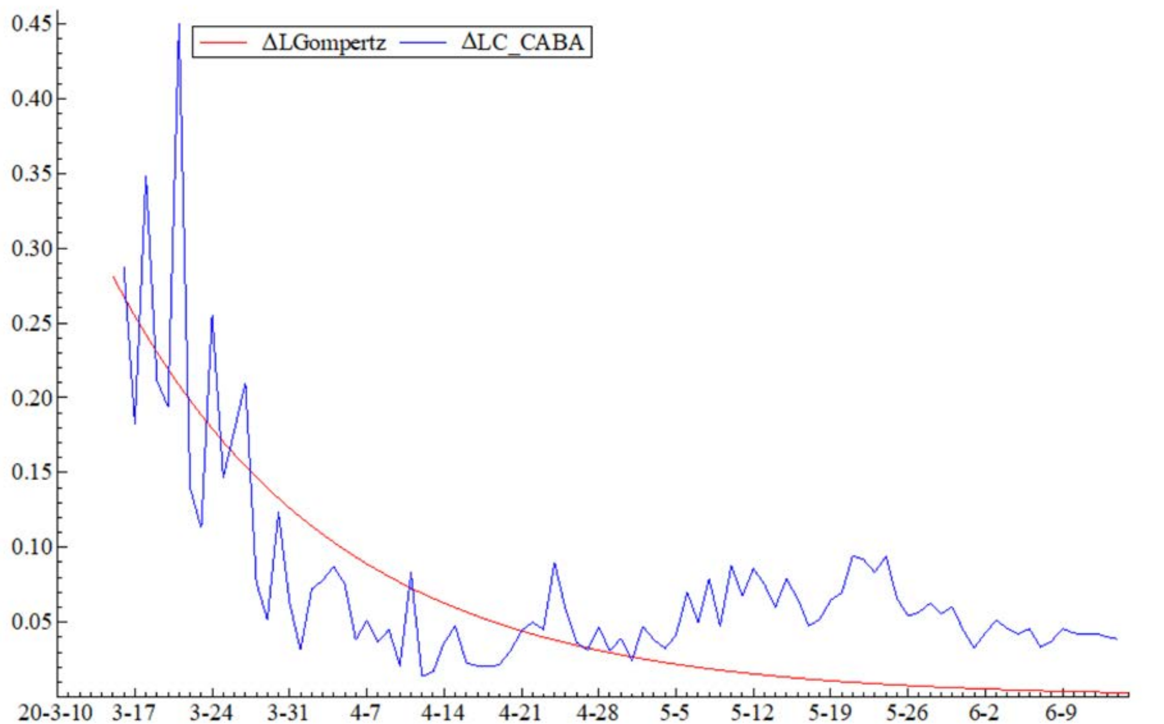


CABA: Gompertz rate vs. actual rate

- This figure shows what would be a single peaked process with one significant wave with a monotonically decreasing growth rate compared with actual CABA data

Figure 5

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



Short-term forecasts of reported cases and deaths in CABA

- We focus on the model ability to forecast these 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.
- An thus establish the connection between short-term forecasts with parameters and indicators of a SIRD model.
- Short run (a week-ahead) forecasts can be done by estimating simple statistical dynamic models but that allow updating and/or rapid break detection.
- This is necessary to follow the disease evolution due to different policy interventions as the degree of lockdowns and their effectiveness
- And sudden shocks which could derive in a contagion process acceleration as the observed in the poor-neighborhoods of the city.

Forecasts approach: dummy saturation and robustification

- To deal with policy interventions and sudden shocks we applied ***step saturation*** and ***impulse saturation*** to take into account shifts and outliers in the models.
- Impulse saturation (of the form 0,0,0, . . . ,1, . . . 0) and step saturation (1,1,1, . . . ,0, 0, 0) are part of an econometric approach that searches for the presence of these dummies for every observation of a given sample. Data themselves are informative about the dummy type and location. 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. See also Ahumada (2018)
- They are essential for our purpose to detect changes in the contagion dynamics that can be informative about transitions between stages.
- Also robust forecasts are meantime needed to rapidly adjust our ex-ante forecasts (see Castle et. al, 2015).

The lockdown effect on CABA reported cases in April

$$\underset{(SE)}{\Delta LC\_CABA} = \underset{(0.061)}{-0.173*\Delta LC\_CABA\_2} + \underset{(0.0052)}{0.0466} + \underset{(0.013)}{0.13*S{:}03\text{-}27} + \underset{(0.0093)}{0.0438*S{:}04\text{-}05}$$

$$+ \underset{(0.024)}{0.266*I{:}03\text{-}21} + \underset{(0.016)}{0.0568*DI{:}03\text{-}24} \qquad (10)$$

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]

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

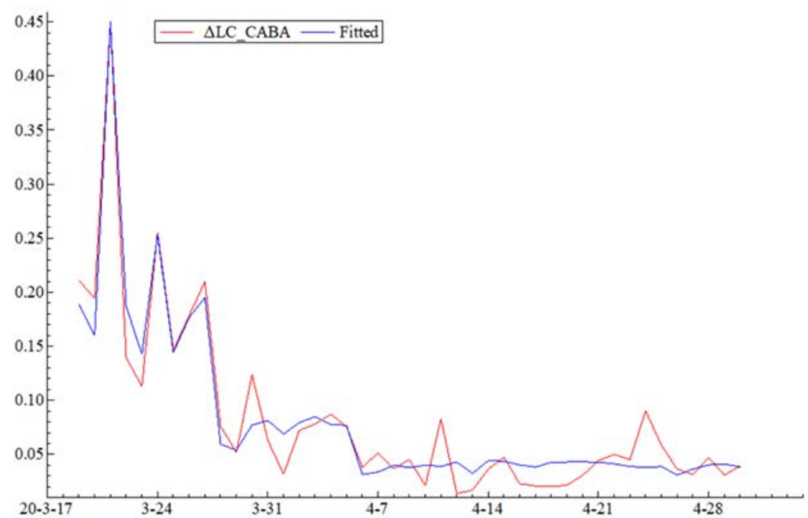
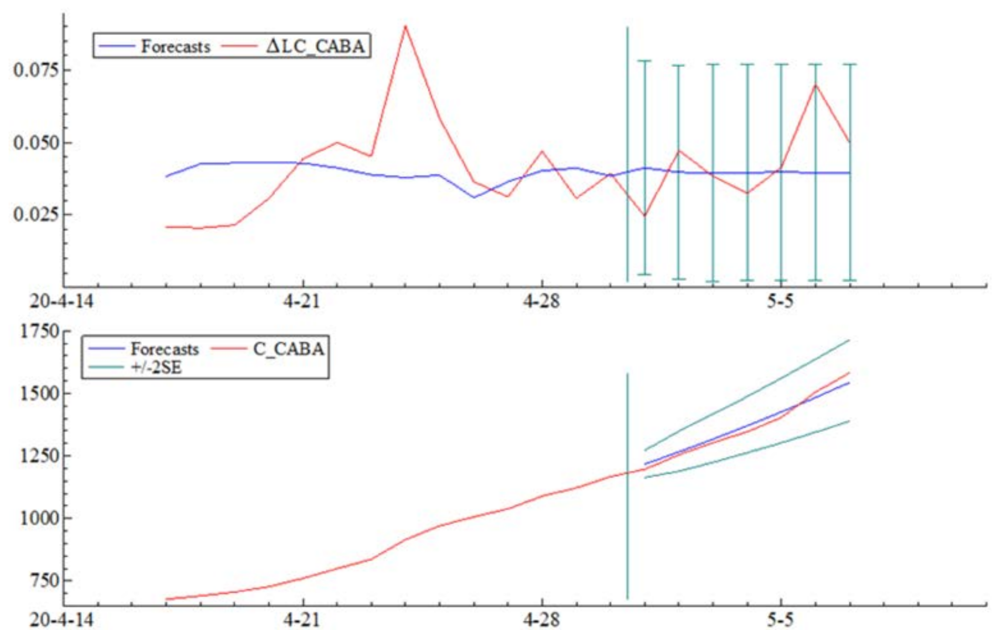


Figure 7
7 day-ahead forecasts of daily growth rates (upper panel) and levels (lower level) of reported COVID-19 cases in CABA- May 1 to May 7, 2020



Second wave: May performance after the outbreak in poor neighborhoods

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

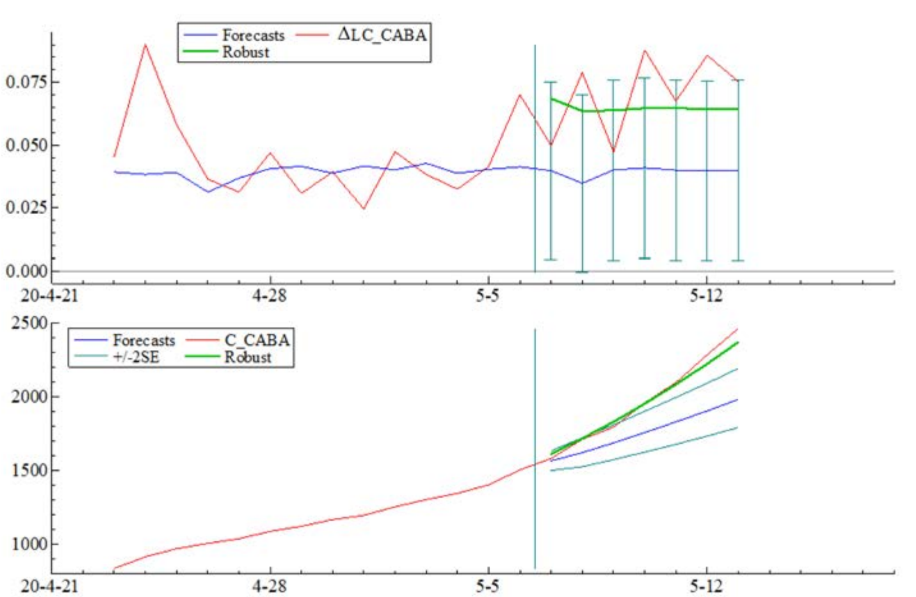
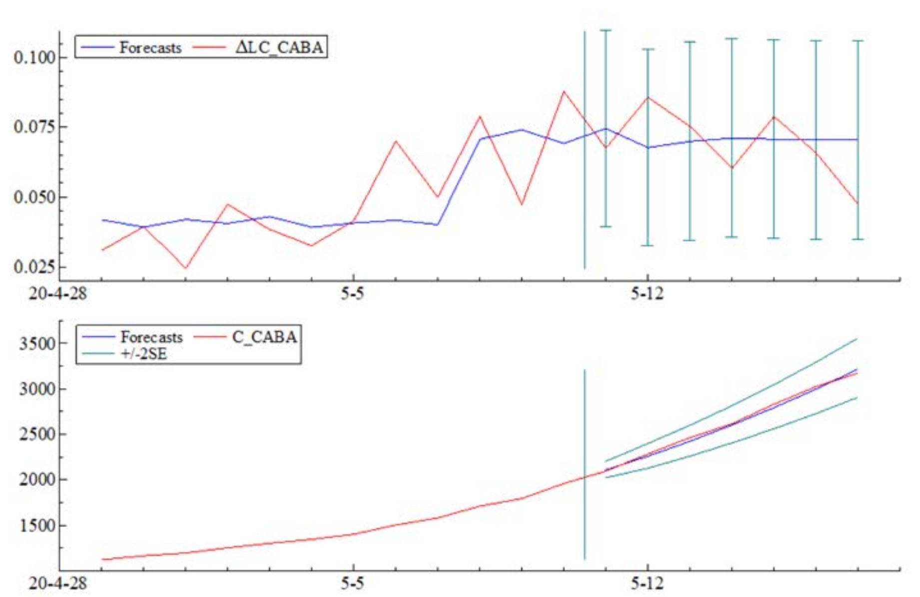


Figure 9

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



Testing Mobility effects

$$\Delta LC\_CABA = \underset{(0.07)}{0.152}*\Delta LC\_CABA\_3 + \underset{(0.012)}{0.0703}*LMobilCABAadj\_8 - \underset{(0.0042)}{0.014}*dweekday1+7$$

(SE)

$$- \underset{(0.044)}{0.214} + \underset{(0.0083)}{0.0439}*S1{:}04\text{-}05 - \underset{(0.0051)}{0.0153}*S1{:}05\text{-}05 + \underset{(0.0059)}{0.0286}*S105\text{-}25 \qquad (12)$$

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]**

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 our sample.

Forecasting COVID-19 deaths in CABA

$$\Delta LDeathsCABA = -0.154^{*}\Delta LDeathsCABA\_6 - 0.175^{*}\Delta 2LDeathsCABA\_3 + 0.501^{*}\Delta LC\_CABA\_16$$
$$(SE) \quad (0.052) \quad (0.035) \quad (0.081)$$

$$+\ 0.47^{*}\Delta LC\_CABA\_19 - 0.0312^{*}S1{:}04\text{-}23 + 0.0376^{*}S1{:}05\text{-}08$$
$$(0.08) \quad (0.011) \quad (0.0084)$$

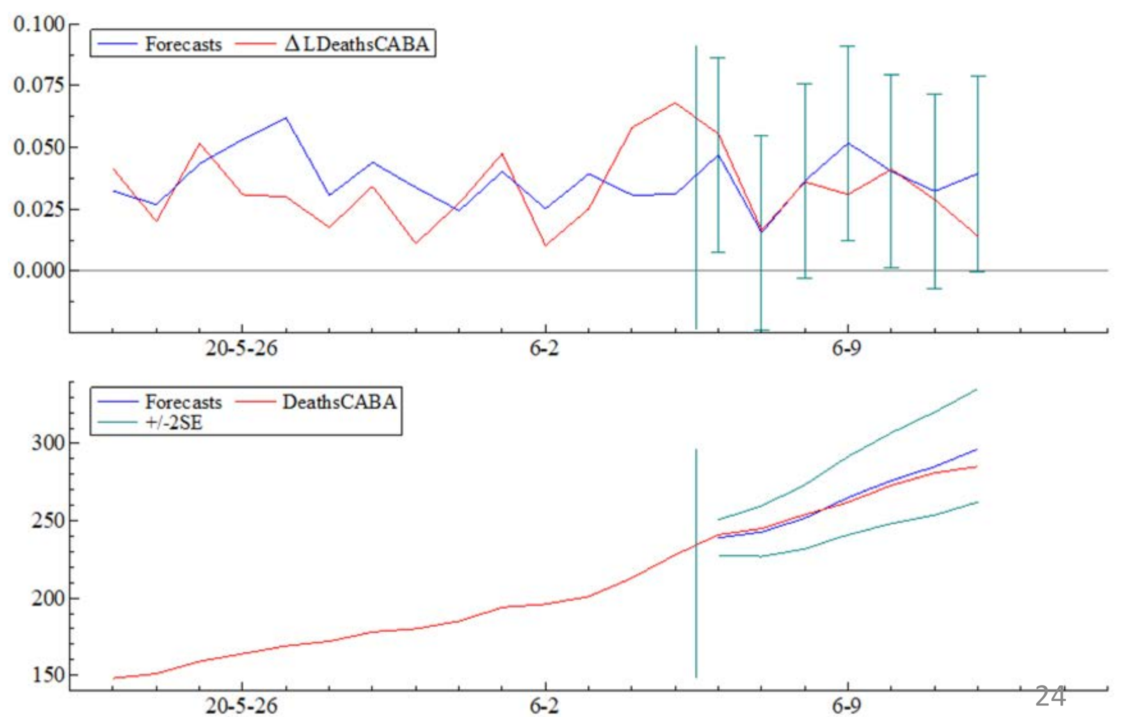
$$-\ 0.0143 + 0.0185^{*}S1{:}05\text{-}27 - 0.0248^{*}dweekday7 \qquad (13)$$
$$(0.009) \quad (0.0097) \quad (0.0088)$$

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

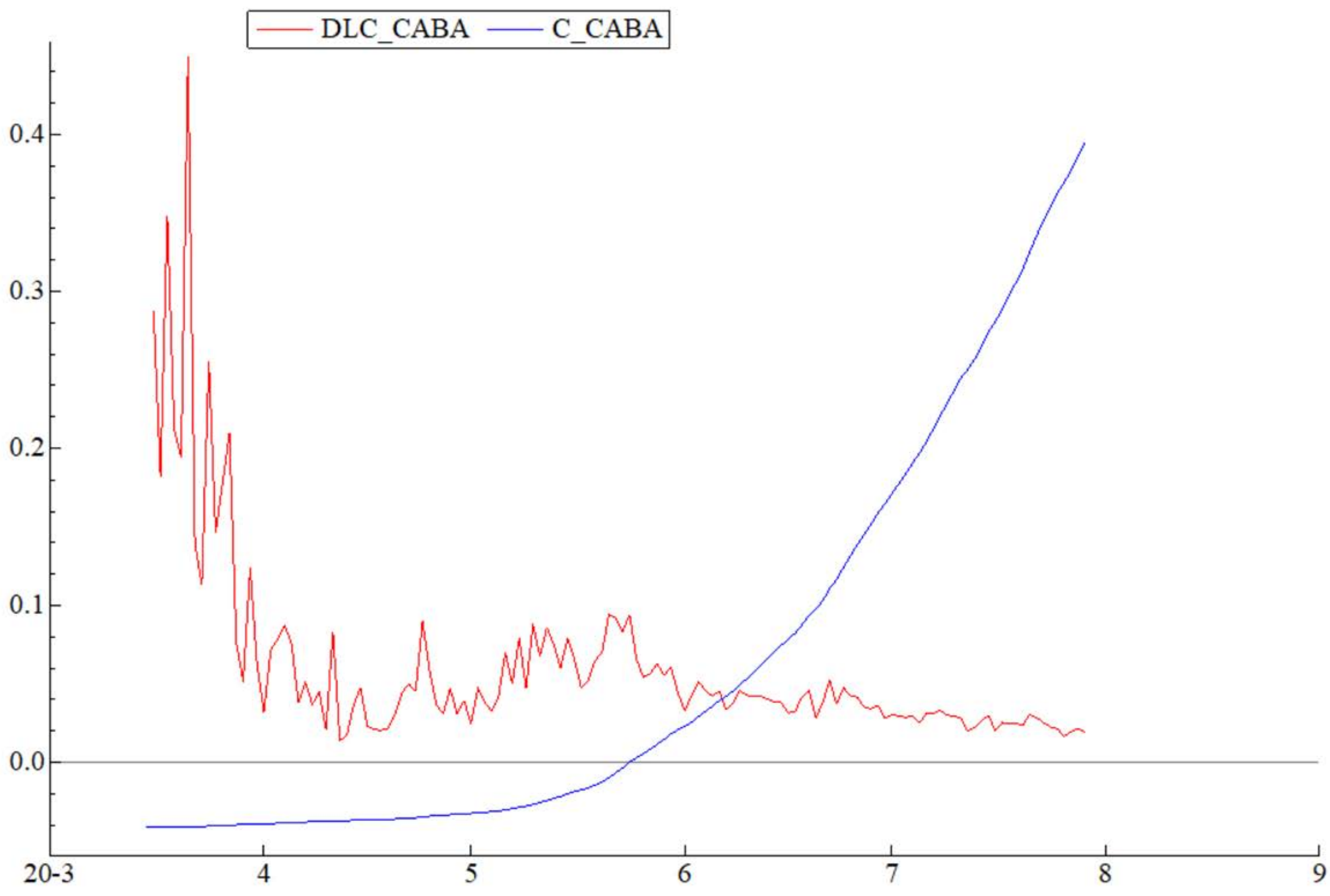
AR 1-2 test:	F(2,51) =	0.67802	[0.5121]
ARCH 1-1 test:	F(1,60) =	0.097826	[0.7555]
Normality test:	Chi^2(2) =	0.64676	[0.7237]
Hetero test:	F(12,49) =	0.34812	[0.9749]
Hetero-X test:	F(18,43) =	0.69984	[0.7921]
RESET23 test:	F(2,51) =	2.4747	[0.0942]

Figure 11

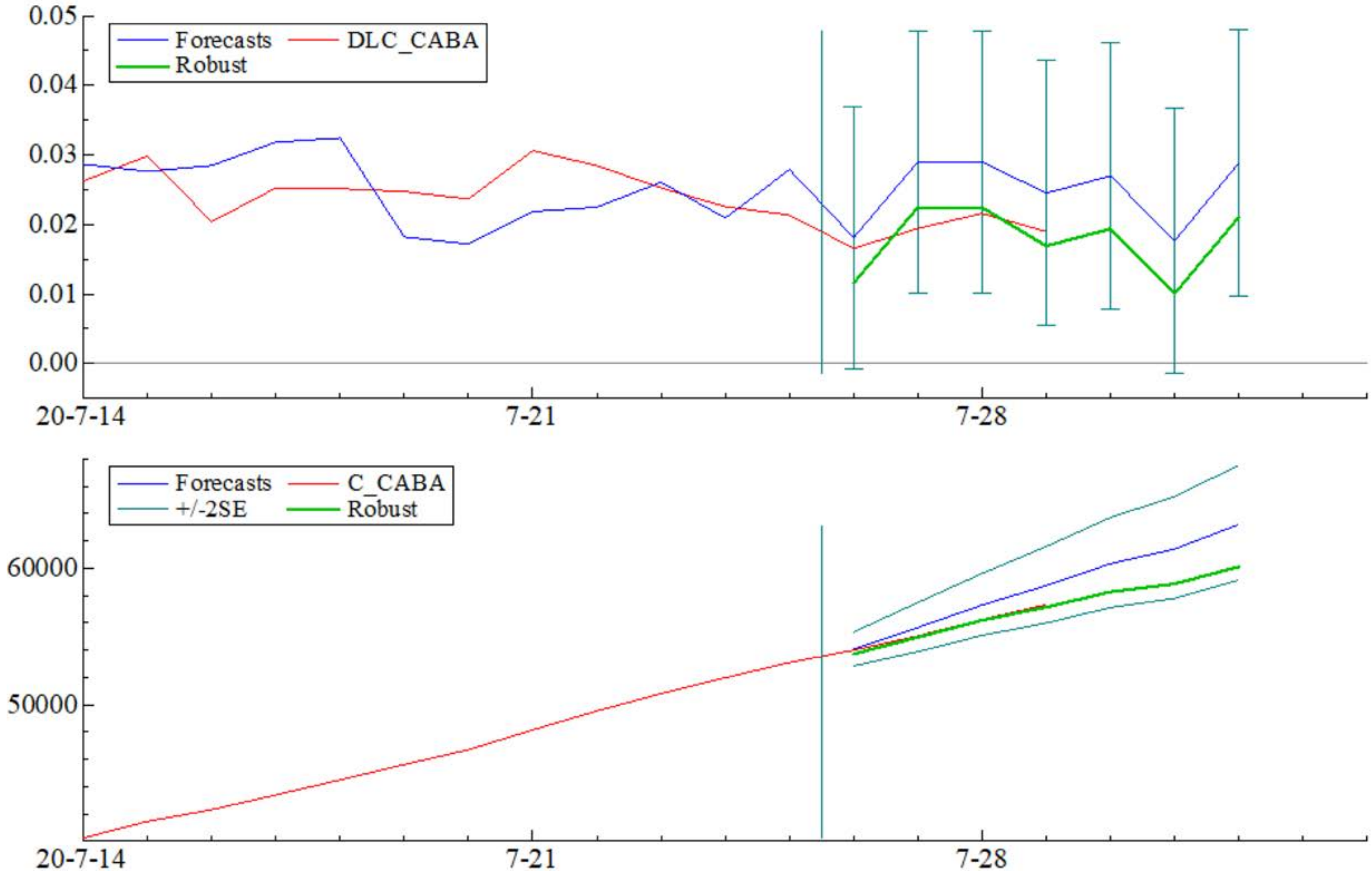
7 day-ahead forecasts of daily growth rates (upper panel) and levels (lower level) of reported COVID-19 deaths in CABA- June 06 to June 12, 2020



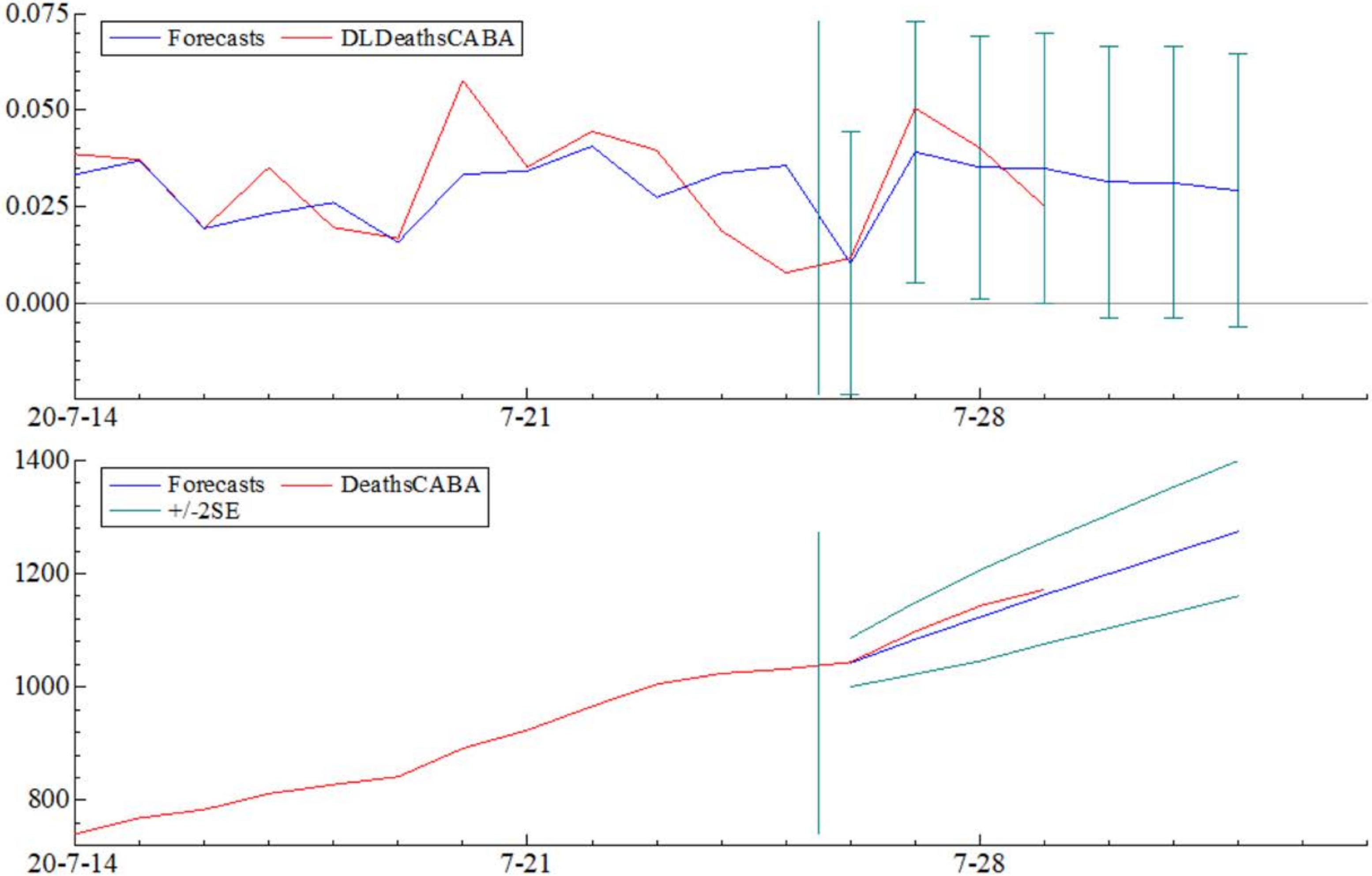
How close to a plateau of cases in CABA?



Update: Latest forecasts of cases



Update: Latest forecasts of deaths



Conclusions

- Short term forecasting of cases and deaths may become more useful than thought because of uncertain waves of infection given that 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.
- 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.
- Rol of testing/tracing/isolation strategy seems also important to assess in the future
- An empirical illustration to CABA shows that the process is indeed uncertain, subject to shifts and requires short term monitoring.
- Two empirical results are important in our application to CABA. First, we find that mobility has an impact on reported cases with an 8-days lag and a semi-elasticity of 0.07. Second we find a lag between reported cases and deaths of 16 to 19 cases.