FGV IIU Discussion Papers

Corona Data Analyses: Looking for Signs of Recovery in Italy and Spain

Renato G. Flôres Jr., with the assistance of Leonid Garnitskiy

FGV IIU International Intelligence Unit

DP 02/20

FGV IIU is a small think tank directly connected to the Presidency of Fundação Getulio Vargas. **FGV IIU** observes the international scene with the purpose of identifying issues, trends and partners relevant for Brazil, as well as contributes to explain Brazilian issues and characteristics to foreign audiences. The exchange of good practices, in both directions, is a subsidiary aim of the Unit.

FGV IIU Discussion Papers address new ideas or more technical issues deserving a broader discussion, that sometimes cannot wait for a more detailed, academic-like analysis. Their content can be varied and many of the ideas may later turn out into specific projects. They state viewpoints, analyses and syntheses considered relevant for the moment. The views in **FGV IIU Discussion Papers** do not engage in any way FGV or its executives and may only be attributed to the respective Authors.
FGV IIU Discussion Papers are identified sequentially, by year, at the left bottom of the front page.

FGV IIU
npii@fgv.br

Director: Renato G. Flôres Jr.
Praia de Botafogo 210, 12th floor; Rio de Janeiro, Brazil.
Voice: +55 21 3799 6220 (Licia)

1. Introduction: one thousand and one data analyses.

One of the certainties about the covid-19 pandemic is that little is known about it. With the progress of the epidemic in several countries, experiments, trials, more encompassing and less unreliable data, together with ideas and innovations from the biological and medical sciences start to generate knowledge on the specific dynamics of the virus. In FGV IIU (2020)[1] a pledge is made for more, actually plenty of, analyses of the data, of all sorts and approaches, in order to produce insights and stylised facts that may help in public policy designs to tackle the different present and future impacts of the catastrophe.

In a pioneering paper, John Tukey[2] set the seminal ideas of looking at data per se, without the goal of building up models or testing sophisticated hypotheses. The basic goal, Tukey suggested, should be to use simple techniques to open routes or paths, to be further developed through more sophisticated approaches. This gave origin to the field known as Exploratory Data Analysis (EDA), with several independent and creative evolutions since. The enormous increase in computer power, and the related algorithms for dealing with extremely large data bases, gave later origin to what came to be called the Big Data area and its corresponding analytical procedures -included by many in the Artificial Intelligence (AI) field- for examining very large quantities of data.

Unbeknownst of the principles and purposes of EDA, some believe that Big Data combined with AI's methods replaced its original procedures, oftentimes (misleadingly) called small sample approaches[3]. Reality is however more complex, and the main principles have not changed. Using Taleb´s terminology[4], either small or big data techniques valid in Mediocristan -the realm of the normal distribution and all those which somehow cannot be considered too far from it- will not work for sets that follow other, more exotic distributions, like the heavy tailed ones, and both approaches have the same need to develop strategies for these so-called odd distributions. Also, insights are sometimes deeper when extracted from smaller samples, but tested in different data sets

---

[1] *Corona-numbers & Policies: some Reflections*, 2020, FGV IIU Flash Notes, March, 25; Rio de Janeiro: FGV.
[2] The Future of Data Analysis, 1962, J. W. Tukey, *Annals of Mathematical Statistics*, 33 (1); 1-67.
[3] Misleading indeed, as AI algorithms may deal with either large or small samples, numerical vs categorical, or sparse vs crowded data, etc, etc.
[4] *The Black Swan: The Impact of the Highly Improbable*, 2007, N. N. Taleb, UK: Allen Lane.

of similar size. Briefly, rather than opposite, both methodologies should be used complementarily. Last but not least, the dubious quality of data on the epidemic, extremely dependent on the way measurements are made, puts an additional caveat on the use of more sophisticated or assumptions-demanding models.

In this paper, simple techniques, in a true EDA spirit, are used to try to find clues, from available data on the pandemics, that would signal that a given policy package is being successful and when its likely end would be. Thus, in a more ambitious attempt, some inkling on the probable length of time, till the epidemic could be considered under full control is also suggested.

This is most crucial, as more flexible confinement rules are now dearly needed, but the legitimate fear of applying them too soon heightens and sometimes biases the debate on this decision.

Section 2 explores (daily) data on new infected cases, while section 3 those related to (daily) deaths attributed to covid-19. In both, the publicly available files from *worldometers.info* have been used. Section 4 wraps up what could be extracted from the previous analyses, in terms of practical procedures, while section 5 concludes with a critical view of the evidences gathered and suggestions for additional pursuits. No panacea or magic solution is presented, rather two tracking mechanisms that could act as a supplementary aid to guess how controlled the epidemic is.

As in the old Middle East classic, at least one thousand and a hundred analyses are needed for unveiling all the relevant patterns of this pandemic and avoiding death; here is simply one of them.

2. Exploring data on infected people.

2.1. Evidences from ratios.

The number of infected people, in all its modalities, is prone to error, especially under-estimation, as it only counts people who have been tested[5]. Even the tests themselves, as known, may not be fully reliable, allowing for false negatives. Despite this, these data -

[5] See, for instance, the Flash Note cited in footnote 1.

in a daily basis- can provide signals on the evolution of the fight against the epidemic. The simplified model in the Annex provides the motivation for the calculations to be performed.

Information on the new daily cases -though still suffering from underreporting- is likely to be less imprecise and with less confounding than that on (cumulated) total infected cases. The ratio of these numbers, at a basic step of development or duration period of the epidemic, is a function of two parameters deeply related to the dynamics of the epidemic process within the given community: the average number of people infected by a person identified in the start of the period with the virus (called R_0, in the Annex), and a synthetic measure of the amount of infected people who, for a variety of reasons, may be considered outside the contagion group, at the end of the period[6] (R_1, in percentual terms).

A first problem for computing such ratios is what precisely is the duration of "a basic step of development of the epidemic"? Certainly not one day, but how long?

At present, there aren't universally agreed answers to key questions like: for how long does an infected person contribute to contagion? how long does it take to detect the presence of the virus in an infected person? do people considered cured really become immune to the virus? what about cycles for the process as a whole?

Based on the available knowledge up to now, the length of the basic step has been chosen as 14 days. During this time, *in the average*, infected people at the beginning of the period can either be cured, or hospitalised, or killed, new contagions reveal themselves, and precautionary measures have a minimum time for being effective. Given the huge uncertainty surrounding the duration of all these events, we've also used 7 days as an alternative check on the duration for the basic step.

A second point is that these ratios should change along the evolution of the epidemic, hopefully lowering, signalling how positive, or not, is the fight against the epidemic. In order to achieve some stability in the process in a given country, and to diminish somewhat the unavoidable underreporting inherent in any new cases data, two countries where the disease is quite advanced already, and a substantial number of cases is available, have been chosen: Italy and Spain.

[6] For more on both R_0 and this second parameter see the Annex. However, even in intuitive terms, it is reasonable to postulate that these two parameters influence the ratio.

Starting from a given initial date, ratios were daily computed, always using for the numerator the value of the same variable 14 (or 7) days ahead. This produces a *daily series of ratios* that, despite noisy, more easily displays the trend of the ratios.

Exhibits 1 and 2 show the calculations for Italy, taking as starting point March 6, 2020 and computing the ratios with 14 (Exhibit 1) or 7 (Exhibit 2) days ahead.

Two things are remarkable from both curves:

i) after oscillating more wildly, the values tend to stabilise -though still slowly decreasing- around nearly the same values, with those for the 14 days interval lower than the corresponding ratios 7 days ahead (same day for the denominator). Stability (roughly), in Exhibit 1, starts around March 19/20, while a little later, around March 23/24, in Exhibit 2. These two facts were taken as supporting the 14 days choice for the duration of the step;

ii) the "stability interval" comprises values lower than 1, in both cases, signalling that the policy package recently enforced by Italy is proving successful of late. However, these values must still decrease, as further discussed below.

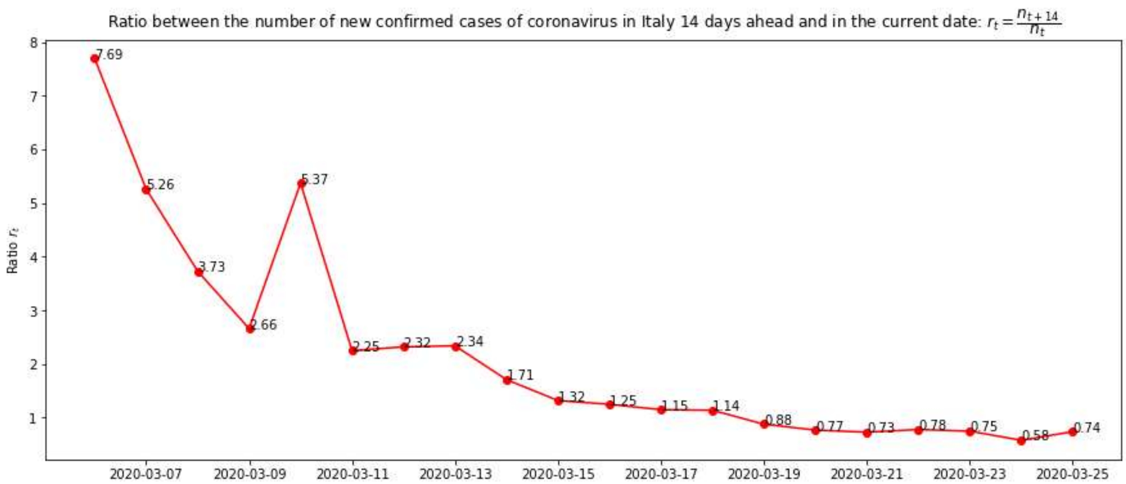

Exhibit 1: Ratio between the number of new confirmed cases of coronavirus in Italy 14 days ahead and in the shown date, starting on March 6, 2020.

Considering the five last values in Exhibit 1, they have an average of .72 and a median of .74 (a preferred location statistic).

If one accepts formula (A.7) in the Annex, this implies that the two key parameters may vary between the following boundaries (see also (A.9))

$$0 < R_0 < .74 \quad \text{and} \quad .36 < R_1 < 1 \quad ,$$

though not necessarily bad, since the (ideal) pair is (R_0, R_1) = (0, 1), much improvement is still needed. Moreover, supposing stability of the ratio around the median value, the time needed for the number of daily cases be 1/10 of that two weeks ago would be around 7 basic steps (see (A.10)), or nearly four more months from March, 25.

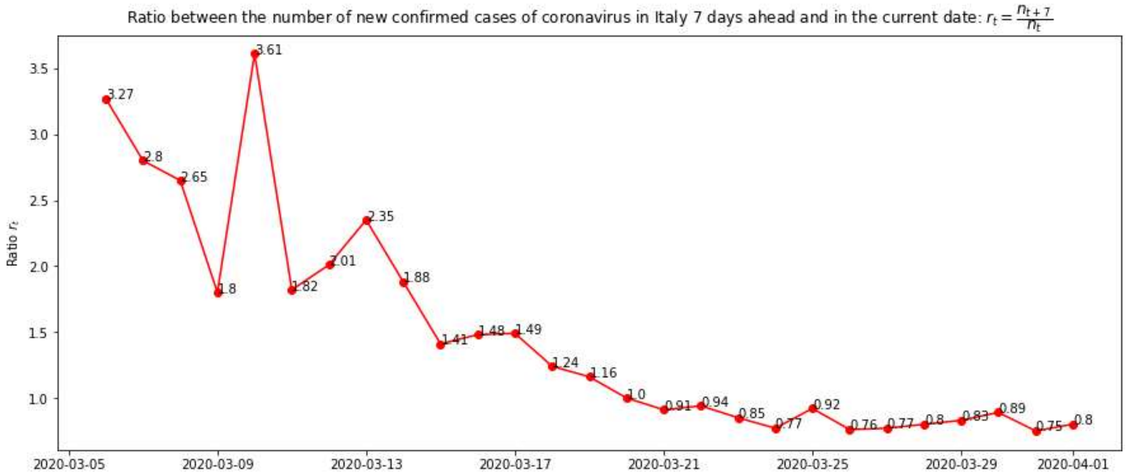

Exhibit 2: Ratio between the number of new confirmed cases of coronavirus in Italy 7 days ahead and in the shown date, starting on March 6, 2020.

Moving to Spain, Exhibits 3 and 4 display the same pattern as the one for Italy, with points i) and ii) equally applying.

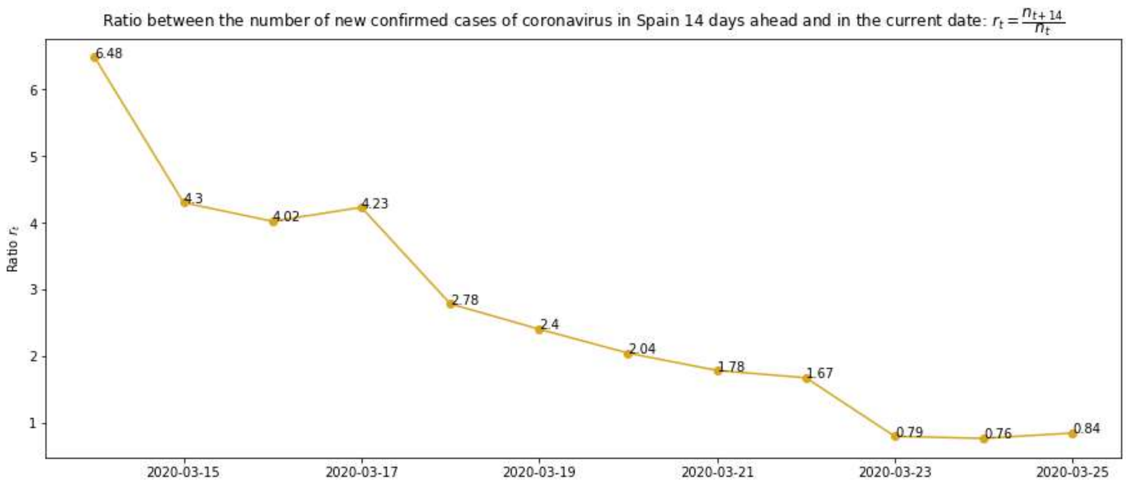

Exhibit 3: Ratio between the number of new confirmed cases of coronavirus in Spain 14 days ahead and in the shown date, starting on March 14, 2020.

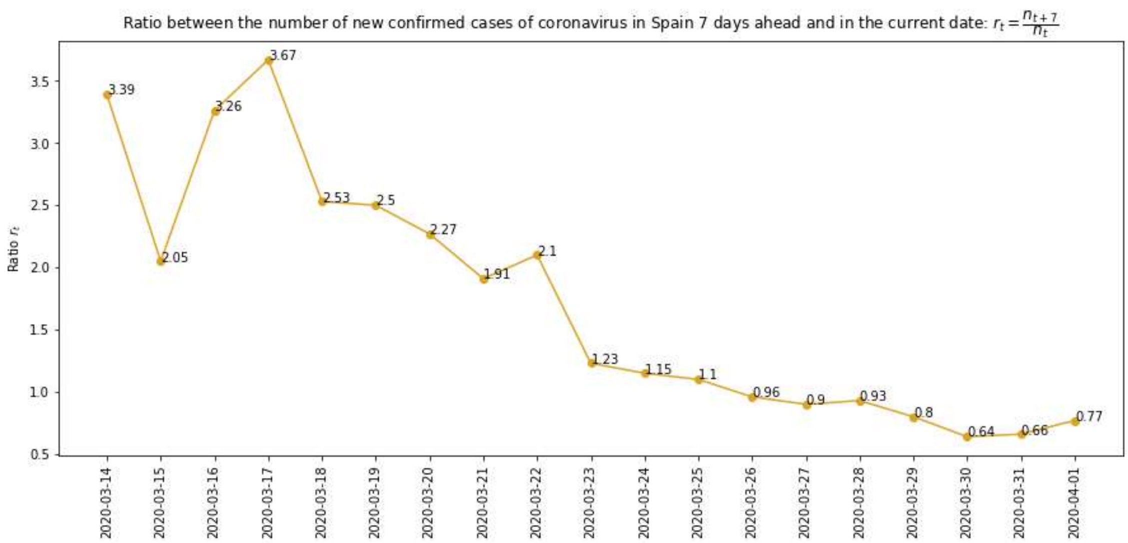

Exhibit 4: Ratio between the number of new confirmed cases of coronavirus in Spain 7 days ahead and in the shown date, starting on March 14, 2020.

Repeating the exercise conducted for Italy, now with the five last figures of Exhibit 3 (with an average of 1.17 and a median of .84), supposing stability around the median value, the time needed for the number of daily cases be 1/10 of that two weeks ago would be around 12 basic steps, or one semester from March, 25.

2.2. Sensitivity of the results.

To gain an idea of the sensitivity of the findings, computation of the ratios was pursued till the last date available[7].

Exhibit 5 complements Exhibit 1, showing seven more points. The ratios continue to oscillate around values similar to the last ones of the previous Exhibit, the new median for the last seven points being .78, quite close to the former one (.74). Somehow, the Italian situation presented no change from that "one week ago", suggesting -a bit as expected- that improvement moves rather slowly.

Exhibit 6 identically complements Exhibit 3 for Spain. The overall pattern is broadly the same as in Italy. The ratios continue to oscillate around the past trend, now -with the exception of an "outlier" in the last day/point- around values a little lower than

[7] April 15, 2020.

the previous ones. Indeed, the new median is .55, suggesting that an improvement might be at work in Spain.

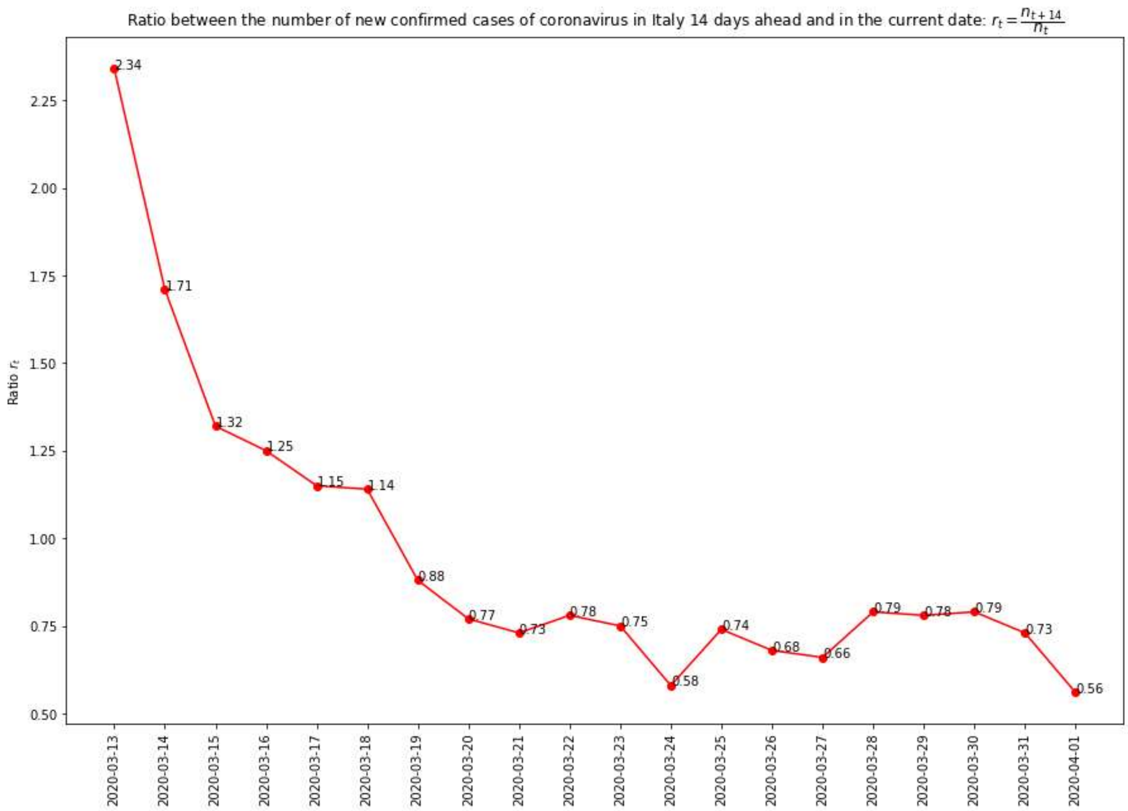

Exhibit 5: Ratio between the number of new confirmed cases of coronavirus in Italy 14 days ahead and in the shown date, during the period from March 13 to April 15, 2020 (the last point thus refers to April 1).

3. Exploring data on daily deaths.

3.1. The predictive regressions.

Data on deaths, though not without measurement problems[8], are less unreliable than those related to new cases. They play a key conceptual role, as the (discrete) first derivative of the cumulated deaths curve, FGV IIU (2020). Only when they clearly start to decrease, signalling that the cumulated curve has changed concavity, one can gain hope that the

[8] Two big problems are: confounding – the patient was infected and died, but his death was actually due to another cause; under-reporting – many countries were until recently, and quite a few until today, only reporting those deaths that had taken place in a hospital, those occurred at home, for instance, going unregistered.

enforced measures are being effective. Once again, in order to analyse reasonably mature situations, only data from Italy and Spain are used here.

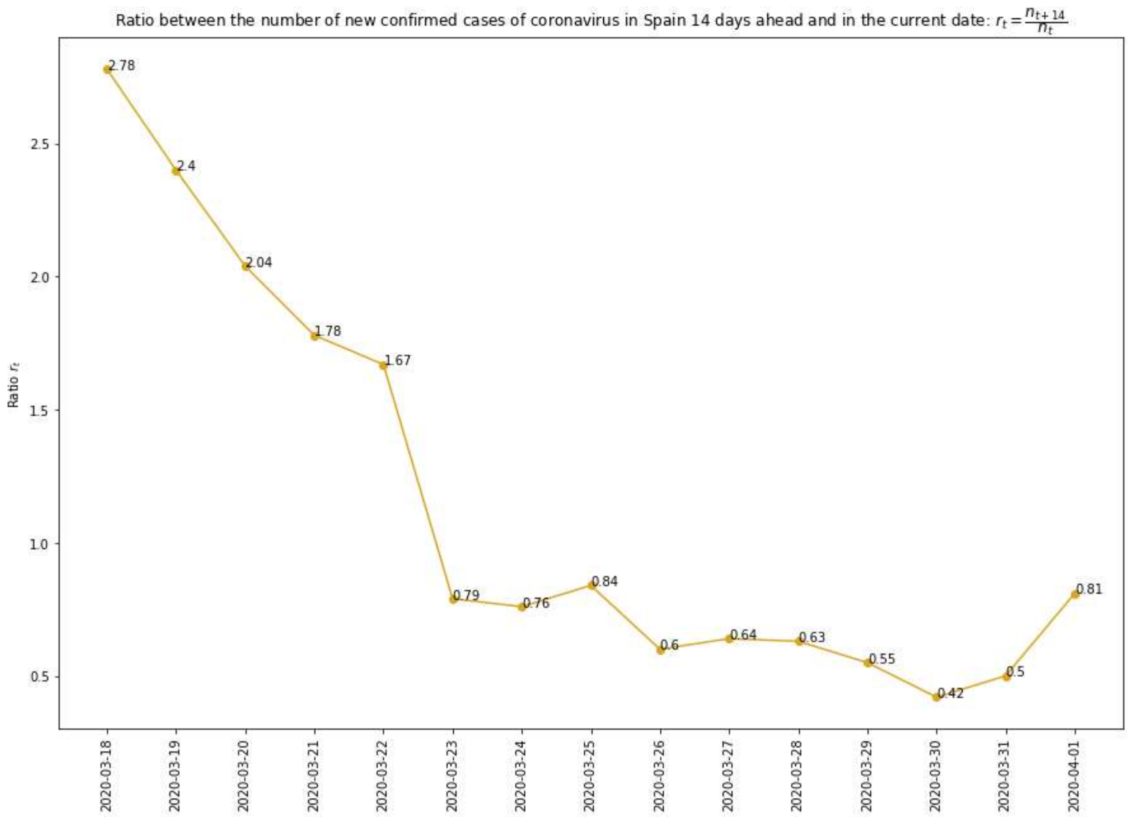

Exhibit 6: Ratio between the number of new confirmed cases of coronavirus in Spain 14 days ahead and in the shown date, during the period from March 13 to April 15, 2020 (the last point thus refers to April 1).

A first key question is: are daily deaths decreasing in these countries? The answer, in an EDA spirit, is anchored merely on visual inspection of their graph, taking in due account its random character. Inspection of the figures from the two countries -available and daily updated in the *worldometers.info* site- led to choose March 27 and April 1 as the starting points for the inversion, for Italy and Spain, respectively.

Exhibit 7 shows the values of daily deaths for the March 27 – April 8 interval, for Italy, and a least-squares regression line fitted to them. The fit is quite good, as information in Exhibit 8 shows; the fitted line can be written as

[estimated new deaths in time t] = 931 - 29 t ; (1)

t=1 (March 27), 2, …, to 13 (April 8).

Once the regression line is obtained, one can find the amount of time (days) needed for the new deaths to be zero. This is a random estimate of something lying outside the regression line interval, and consequently must be taken as a very rough idea of the right point. Exhibit 9 provides a picture of the point, but a sensible idea of its variability must be produced.

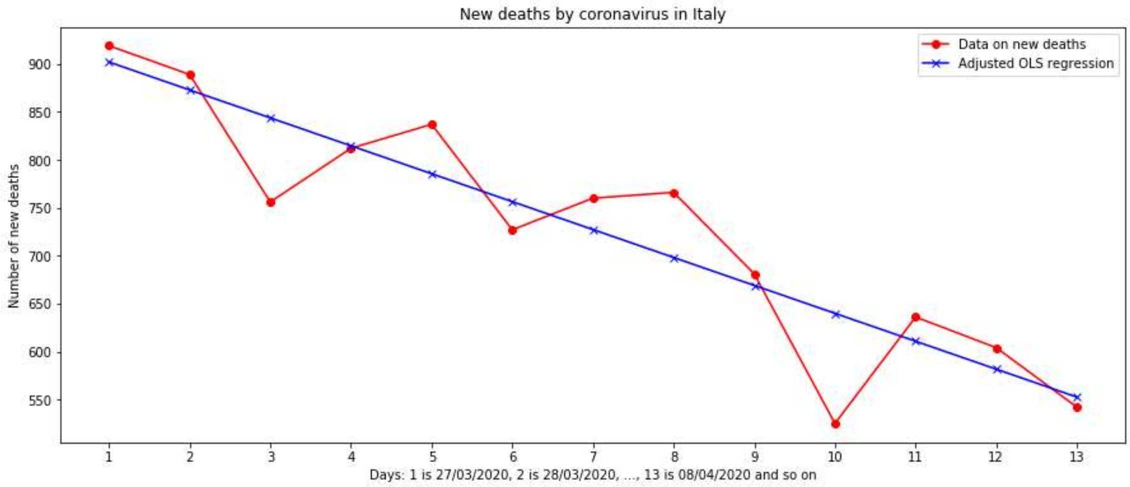

Exhibit 7: Italy - number of daily covid-19 deaths, from March 27 to April 8, 2020. An adjusted least-squares line is in blue.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        newdeaths_Italy   R-squared:                       0.828
Model:                            OLS   Adj. R-squared:                  0.813
Method:                 Least Squares   F-statistic:                     53.01
Date:                Thu, 09 Apr 2020   Prob (F-statistic):           1.58e-05
Time:                        15:49:11   Log-Likelihood:                -69.210
No. Observations:                  13   AIC:                             142.4
Df Residuals:                      11   BIC:                             143.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
date         -29.1264      4.000     -7.281      0.000     -37.931     -20.321
const        931.1154     31.753     29.324      0.000     861.228    1001.003
==============================================================================
Omnibus:                        4.962   Durbin-Watson:                   2.264
Prob(Omnibus):                  0.084   Jarque-Bera (JB):                2.522
Skew:                          -1.065   Prob(JB):                        0.283
Kurtosis:                       3.347   Cond. No.                         17.0
==============================================================================
```

Exhibit 8: Italy – estimated parameters and corresponding statistics of the least squares line shown in Exhibit 7.

There are different ways to produce an idea of how variable the estimate of the intercept in the horizontal axis is. A pessimistic alternative is to modify equation (1) according to the standard errors of the coefficients, shown in Exhibit 8. Moving the intercept two standard errors down and the angular coefficient two up, one ends with a new line[9]

$$\text{[estimated new deaths in time t]} = 867 - 21\ t \quad . \tag{2}$$

The point of zero deaths lies around 41-42 for (2), or rather May 6/7, while it was around April 26/27 for the estimated line (1). From the evidence thus gathered, comparing with the horizons established for a sensible reduction of the ratios in the previous section, this further date seems to merely give a lower bound for the crossing date.

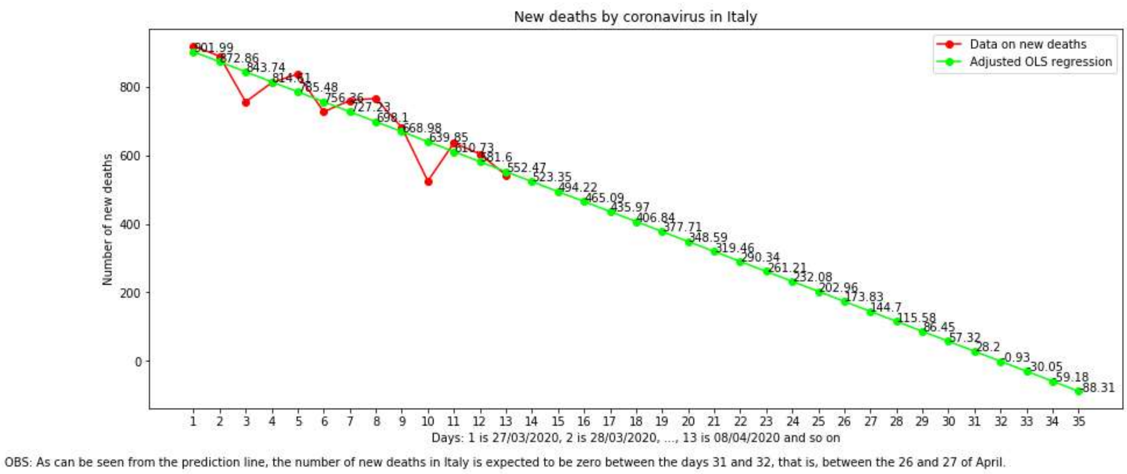

Exhibit 9: Italy – extending the adjusted least-squares line for the number of daily covid-19 deaths, till reaching the horizontal axis; crossing is between days 31 and 32.

An even more pessimistic estimate could have been obtained by employing three standard errors for correcting the two coefficients, but better than this is to continuously update the regression, keeping track of how the two estimates above evolve.

The same can be done for Spain, with Exhibits 10 and 11 showing the decreasing trend of the new deaths, starting from April 1, and the statistics of the fitted regression line. The latter is equal to

[9] This an intuitive, straightforward way of finding a "pessimistic line"; statistical considerations on the covariance between the two least squares estimators are discarded.

[estimated new deaths in time t] = 953 - 36 t ; (3)

t=1 (April 1), 2, …, to 8 (April 8).

A pessimistic version can again be obtained by changing the coefficients, in the proper direction, by two standard errors, giving way to the following straight line:

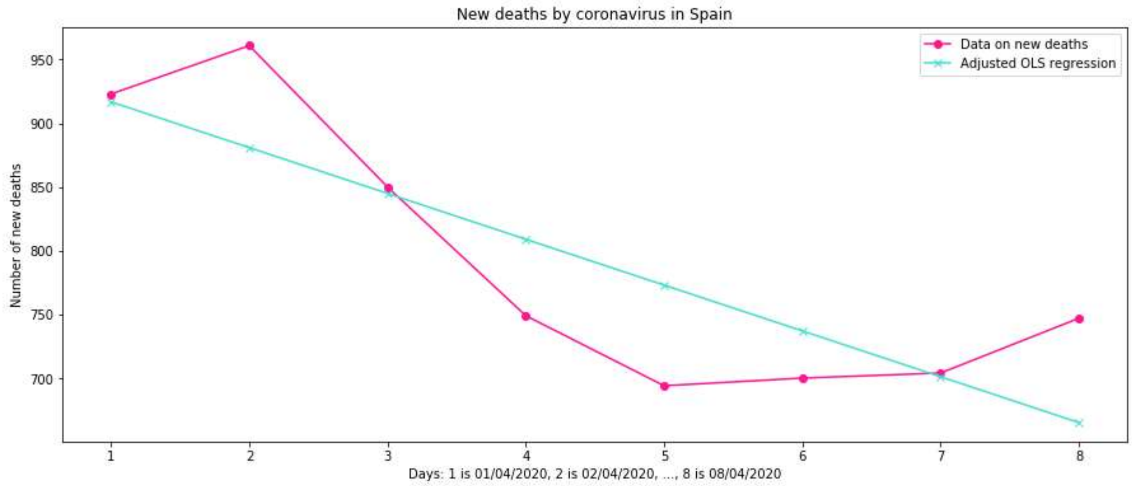

Exhibit 10: Spain - number of daily covid-19 deaths, from April 1 to April 8, 2020. An adjusted least-squares line is in green.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          newdeaths_Spain   R-squared:                       0.690
Model:                              OLS   Adj. R-squared:                  0.639
Method:                   Least Squares   F-statistic:                     13.37
Date:                  Thu, 09 Apr 2020   Prob (F-statistic):             0.0106
Time:                          18:04:20   Log-Likelihood:                -43.444
No. Observations:                     8   AIC:                             90.89
Df Residuals:                         6   BIC:                             91.05
Df Model:                             1
Covariance Type:              nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
date         -35.9762      9.841     -3.656      0.011     -60.056     -11.897
const        952.8929     49.694     19.175      0.000     831.297    1074.489
==============================================================================
Omnibus:                          0.492   Durbin-Watson:                   1.036
Prob(Omnibus):                    0.782   Jarque-Bera (JB):                0.486
Skew:                             0.225   Prob(JB):                        0.784
Kurtosis:                         1.879   Cond. No.                         11.5
==============================================================================
```

Exhibit 11: Spain – estimated parameters and corresponding statistics of the least squares line shown in Exhibit 10.

[estimated new deaths in time t] = 853 - 16 t . (4)

The point of zero deaths lies around May 23/24 for (4), while it was around April 26/27 for the estimated line (3).

Once again, in comparison with the horizons estimated from the new cases ratios, the method looks rather optimistic, providing a lower bound for the desired date.

3.2. Sensitivity of the results.

Sensitivity was again tested by enlarging the sample -actually, adding one extra week of observations- while keeping the same initial date for the regressions.

Exhibits 12 and 13 show the new results for Italy. The declining trend is maintained and the new line

[estimated new deaths in time t] = 871 - 19 t ; (5)

t=1 (March 27), 2, …, to 20 (April 15),

is rather close to the pessimistic version (2). Following the same logic as before, from the information in Exhibit 13, the new pessimistic version becomes

[estimated new deaths in time t] = 807 - 13 t ; (6)

thus, moving the crossing date to around June 16.

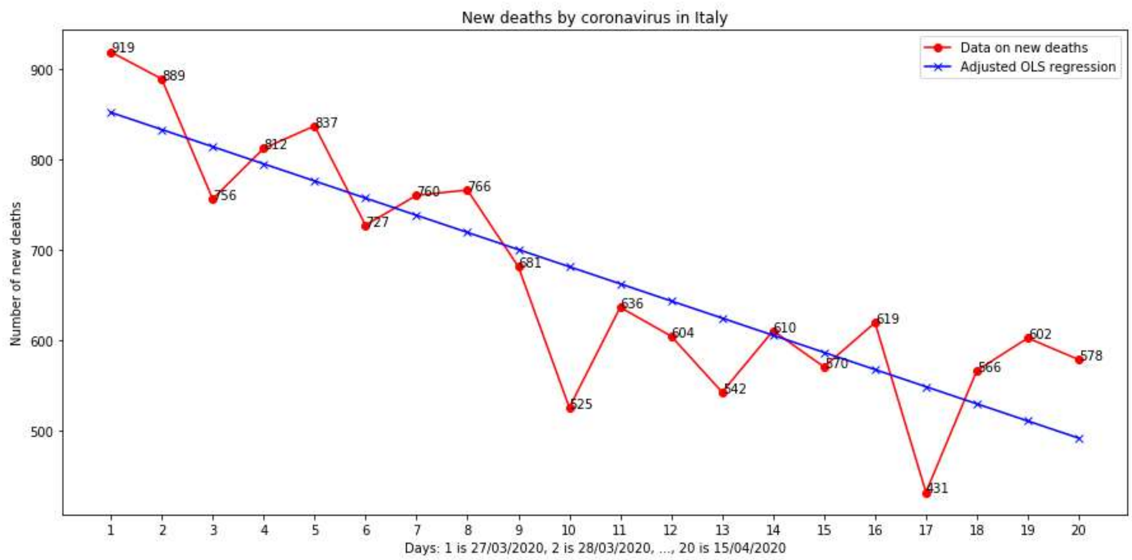

Exhibit 12: Italy - number of daily covid-19 deaths, from March 27 to April 15, 2020. An adjusted least-squares line is in blue.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        newdeaths_Italy   R-squared:                       0.735
Model:                            OLS   Adj. R-squared:                  0.721
Method:                 Least Squares   F-statistic:                     50.05
Date:                Thu, 16 Apr 2020   Prob (F-statistic):           1.35e-06
Time:                        11:38:02   Log-Likelihood:                -112.08
No. Observations:                  20   AIC:                             228.2
Df Residuals:                      18   BIC:                             230.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
date         -18.9940      2.685     -7.074      0.000     -24.635     -13.353
const        870.9368     32.162     27.079      0.000     803.366     938.508
==============================================================================
Omnibus:                        2.190   Durbin-Watson:                   1.649
Prob(Omnibus):                  0.335   Jarque-Bera (JB):                1.555
Skew:                          -0.672   Prob(JB):                        0.460
Kurtosis:                       2.759   Cond. No.                         25.0
==============================================================================
```

Exhibit 13: Italy – estimated parameters and corresponding statistics of the least squares line shown in Exhibit 12.

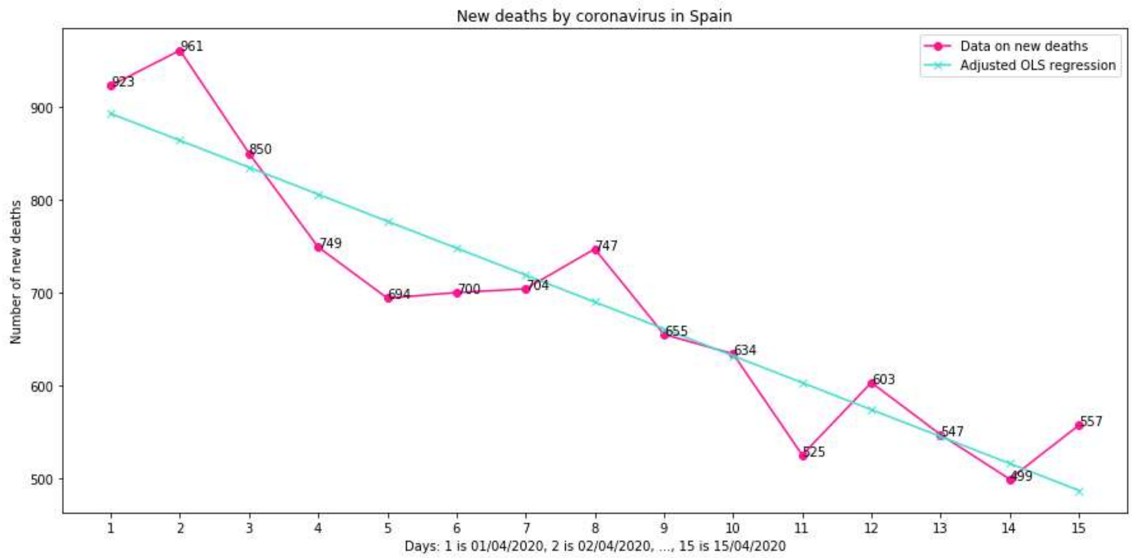

Exhibit 14: Spain - number of daily covid-19 deaths, from April 1 to 20, 2020. An adjusted least-squares line is in green.

Results for Spain are in Exhibits 14 and 15, the new regression being somewhat more conservative than the previous one

[estimated new deaths in time t] = 922 - 29 t ; (7)

t=1 (April 1), 2, …, to 20 (April 20) ,

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        newdeaths_Spain   R-squared:                       0.859
Model:                            OLS   Adj. R-squared:                  0.848
Method:                 Least Squares   F-statistic:                     79.31
Date:                Thu, 16 Apr 2020   Prob (F-statistic):           6.78e-07
Time:                        12:41:59   Log-Likelihood:                -80.183
No. Observations:                  15   AIC:                             164.4
Df Residuals:                      13   BIC:                             165.8
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
date         -29.0036      3.257     -8.906      0.000     -36.039     -21.968
const        921.8952     29.611     31.134      0.000     857.925     985.866
==============================================================================
Omnibus:                        0.140   Durbin-Watson:                   1.427
Prob(Omnibus):                  0.933   Jarque-Bera (JB):                0.353
Skew:                           0.082   Prob(JB):                        0.838
Kurtosis:                       2.266   Cond. No.                         19.3
==============================================================================
```

Exhibit 15: Spain – estimated parameters and corresponding statistics of the least squares line shown in Exhibit 14.

though still optimistic, producing a crossing around May 1/2. The new “pessimistic line”

$$[\text{estimated new deaths in time t}] = 862 - 23\,t \quad ; \qquad (8)$$

would not change much the situation, moving the crossing date to May 6/7.

Sensitivity checks, in both countries, suggest that the decreasing trend seems to be a reality: the measures are working, though significant falls take time. The regressions seem to demand at least three weeks of data, as the Spanish case suggests, moreover, a three-standard-errors pessimistic line may be used.

4. The two proposals - a first version

What could be of practical use from the analyses above? Two procedures are proposed:

i) Systematic use of the two-weeks ratio for the new cases statistics, monitoring their trend and (hopefully) decrease:

Computation could take place in a weekly basis, and the trend of the median value of the last 5 to 7 numbers should be followed. Use of this value to gauge the number of weeks/months till reaching a .10 value -as in section 2.1- should be a supplementary information for allowing changes in the implemented policy package. If the ratios start to oscillate around .20 - .10, flexibilization -with due care- imposes itself.

ii) Use of the new deaths-regression, with at least 21 observations:

Again, computations could be made weekly, and a moving window of three weeks, for instance, could be used for updating the results. The crossing points for both the basic regression and its two or three-standard errors pessimistic version, should be obtained, with the proviso they give lower bounds to the desired final outcome.

The combined use of both procedures can act as a complementary check of how steady and positively-trended are the measures being enforced. Convergence of the two "predictions", in the sense that the pessimistic lower bound in ii) becomes closer to the starting date for the .10 ratio in i), is a further welcome signal that could help in the decision for more flexibility.

Needless to say, both checks must continue to be used after a change in the policy package -usually towards a less strict mode- takes place. They would now alert whether a reversal of the positive trends occurs.

5. Conclusions.

Available data on the corona pandemic suffer from manifold problems. Poor data is however not a synonym for useless data; and analyses with the existing data bases are dearly needed.

Among the many questions related to the management of the epidemic, two are foremost given their comprehensive impact: how can one know whether a given policy package is being effective? when is there enough evidence supporting a policy change, usually towards more flexibility regarding free movement of people and business activities?

Answers, beyond the political side anchoring any such decision, need a sufficient number of evidences from various aspects and domains related to the epidemic. Data analyses must be part of it, usually as a provider of ancillary evidence.

Two procedures, based on daily country figures of new infected cases and deaths have been proposed here. Work continues on both. For the former, more checks on the ideal interval for the ratios are in progress, as well as ways to improve the simplified model. For the latter, additional tests on the ideal sample and method for the deaths-regression are being conducted.

The procedures confirm that the strict measures imposed in both countries are working. In Italy, it is expected that in about two to three months from April 15, 2020, the situation may be close to normality, in Spain a lengthier period is expected, that may take up to four months or more. The number of deaths, in both countries, can become rather small before the previous deadlines. All these forecasts can be weekly updated and refined, what is one of the appeals of the proposal.

Other approaches may also be designed. As said in the Introduction, much work and a myriad of procedures are a pre-condition for arriving at full knowledge of the joint dynamics of the covid-19 epidemic *and* its undissociated social responses.

Annex: A Simplified Model

A.I. Model description.

A simplified and easy model is developed for the expansion of the epidemic *under* a given policy package. It implies many assumptions and quite a few approximations, hopefully described below. Moreover, as oftentimes happens, there is a gap between the model and the way it is used in the text as motivation for the daily new cases-procedure, aggravated by the far from ideal quality of the data available. All this implies that the model must be considered as a first-order approximation to a dynamics (that of the ratios described below) found insightful for evaluating the progress of a given policy to handle the pandemic.

Let a community with sufficiently many people be given and x stand for the whole number of infected individuals at initial time t=0. The model moves at discrete *time steps*. Though numbered as 1, 2, 3, etc, their actual *time duration* is crucial for practical applications, as later discussed.

It is assumed that, when moving from t to t+1, there was sufficient time for:

* new people to become infected and detected as such. The average number of people

infected by someone *who had the virus in time t, including direct and higher order contagions*[10], will be denoted by R_0, as it is quite similar (though not identical) to the classical parameter in mathematical epidemic theory. R_0 can be greater, smaller or equal to 1;

* part of the infected people to leave the group. This may be due to a variety of reasons: they may have been cured, or died, or have become free of the virus (they were asymptomatic, for example, and lost their capacity to infect others), or part has been set aside -i.e., not infecting new people in the next time interval- either because are safely cared in a hospital or duly confined, without possibility of further contagion. The proportion of all these groups together, in relation to the infected population at the start of the period, will be denoted by R_1, a number the maximal (ideal) value is 1.

Starting at t=0, at time t=1 the number of infected people will be equal to the sum of two quantities:

those newly infected: xR_0 and

those remaining from the previous group of infected people: $x(1-R_1)$,

so that the new number of infected ones is: $xR_0 + x(1-R_1) = x[R_0 + (1-R_1)]$. (A.1)

Moving to time t=2 amounts to repeat the process.

The new number of infected people will be equal to the sum of two quantities:

those newly infected: $x[R_0 + (1-R_1)]R_0$ and (A.2)

those remaining from the previous group of infected people: $x[R_0 + (1-R_1)](1-R_1)$,

with the new number of infected ones being: $x[R_0 + (1-R_1)]^2$. (A.3)

It is easy to verify that, at t=n,

those newly infected will be: $x[R_0 + (1-R_1)]^{n-1}R_0$ and (A.4)

---

[10] This implies that, during the time interval, someone infected by another person who had been infected, during the same period, by a member of the group in t, is "considered" as infected by this original individual.

those remaining from the group of infected people at t=n-1:

$$x[R_0 + (1-R_1)]^{n-1}(1-R_1), \qquad (A.5)$$

so that the total number of infected ones is: $x[R_0 + (1-R_1)]^n$. (A.6)

An important consequence is that the ratios of

* the total number of infected people at time t=n by the total at t=n-1,

* the newly infected ones at time t=n by those (newly infected) at t=n-1

* the remaining infected at time t=n by those at t=n-1

are *all the same* and equal to $[R_0 + (1-R_1)]$. (A.7)

Expression (A.7) is crucial in several ways. As shown by (A.6), It generates the multiplier that will progressively modify the number of infected people. The condition for the expansion of contagion to be put to an end is that it is inferior to 1; for this, not only R_0 must be lower than 1 -as basic epidemic models tell us- *but rather its value increased by* $(1-R_1)$, which pools together several factors inherent to the joint dynamics of the epidemic and the global measures adopted to contain it.

Indeed, parameter R_1 combines factors that can be considered "positive", like the proportion of cured victims, with others negative, that of fatal victims, with others that may signal the efficacy of several measures, precautionary ones for instance, like total (and successful) confinement of people already infected.

In a dynamic context, (A.7) is not expected to remain constant, actually what is aimed at is that, thanks to changes in both parameters, it will progressively decrease, become lower than one and, eventually (and ideally), reach zero.

Expression (A.7) also tells that independently of the initial number of infected people, the ratios are solely dependent on the two parameters, what gives way to the idea of the first proposal.

A.II. Application.

To use the model with real data, a first and key decision is about the time duration of each step, a question also discussed in subsection 2.1 of the main text. We've experimented with 7 and 14 days and eventually opted for 14, which produced more sensible results. Work is still in progress regarding 21 days, both in terms of interpretation and the results produced by this duration.

One may also dispute that, along a 14 days interval, say, new infected cases pop up daily, those 'in the next day', for instance, being due to cases a few days before the one at stake. Moreover, as 'new cases' *actually stands for new people tested* -for whatever reason- and proved positive, the 'next-14-days-cases' are likely to include individuals who were not in the original (14 days ago) new cases.

This is broadly true and may indeed add a confounding factor to the 'next-14-days-cases' figure used for the numerator. Without denying this nuisance, three counterpoints are raised.

The first is that only countries where the epidemic is quite diffused and in autonomous development already have been used, as Italy and Spain, hopefully reducing the number of 'infected newcomers'. The second is that if the time duration has been well chosen, the cases coming from before the denominator's date are expected to be small to negligible. Thirdly, and perhaps surprisingly, we have calculated the ratios in a daily basis, in order to give an idea of the evolution of the several (daily) cohorts[11].

Moreover, evidence is always drawn from a time sequence of ratios, which validates behaviours like stability or decay.

Once an empirical estimate of (A.7) is obtained, say $\ddot{\alpha}$, the 45° line

$$R_0 = (-1+\ddot{\alpha}) + R_1 \quad , \tag{A.8}$$

gives, remembering that R_1 is at most 1, the possible ranges for both parameters

$$0 \leq R_0 \leq \ddot{\alpha} \quad \text{and} \quad 1-\ddot{\alpha} \leq R_1 \leq 1 \quad , \tag{A.9}$$

drawing attention why $\ddot{\alpha}$ should be small.

A reasonable small value for (A.7) seems to be .10, and the number n* of periods, *after the last numerator's date*, for achieving this is easily computed as n*=n# - 1, where

$$n\# = \ln .10 / \ln \ddot{\alpha} \quad . \tag{A.10}$$

[11] Borrowing this concept from demography.