The Lancet

COVID-19 infection process in Italy and Spain: Are the data talking?

--Manuscript Draft--

Manuscript Number:	THELANCET-D-20-05421
Article Type:	FT to NT Article (INTERNAL USE ONLY)
Keywords:	COVID-19; contagion pattern; ARIMA method; forecasting; Italy; Spain
Corresponding Author:	María de la Paloma Taltavull de La Paz, Ph.D Universitat d'Alacant Alicante, Alicante SPAIN
First Author:	Paloma Monllor Taltavull, PhD
Order of Authors:	Paloma Monllor Taltavull, PhD
	Zhenyu Su, PhD Candidate
	Laura Gabrielli, PhD
	Ana Montoro, PhD Candidate
	María de la Paloma Taltavull de La Paz, Ph.D
Manuscript Region of Origin:	SPAIN
Abstract:	Background: COVID-19 has spread successfully worldwide in a matter of weeks. After the example of China, all the affected countries are taking hard-confinement measures to control the infection and to gain some time to diminish the big amount of cases that arrive to hospital. Although the measures in China reduced the percentages of new cases, this is not seen in other countries that have taken similar measures, such as Italy and Spain. Now we are in the middle of a battle trying to prevent the healthcare system from collapsing while it effectively responds to the needs of patients who are infected and require hospitalization. Methods Using China as a mirror of what could happen in our countries and with the data available, we calculated a model that forecasts when are the peak of the curve going to happen. We aim to review the patterns of spread of the virus in the three countries and their regions, looking for similarities that reflect the existence of a common pattern in this expansive virulence and the effects of the intervention of the authorities with drastic isolation measures, to contain the outbreak. Findings A model based on ARMA methodology and including Chinese disease pattern as a proxy predicts a reduction of the speed of contagious during the last week of March in Italy and smooth reduction until the peak in the second week of April. In the case of Spain, the infections still will follow growing fast until the end of March . Interpretation : We present the following predictions to inform political leaders since they have the responsibility to maintain the national health systems away from collapsing. We are confident this data could help them into decision taking and place the capitals (from hospital beds to human resources) into the right place.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

manuscript

COVID-19 infection process in Italy and Spain: are data talking?

Paloma Monllor [a,b], , Zhenyu Su [a], Laura Gabrieli [c] Ana Montoro [d] and Paloma Taltavull [ae]

[a] International Economy Institute, University of Alicante. Campus Sant Vicente del Raspeig 03080 Alicante, Spain

[b] Department of Physiology, Faculty of Medicine University of Valencia, Avda Blasco Ibáñez 15, Valencia, Spain

[c] Architecture School, University of Venice. Isola di San Servolo, 30133 Venezia VE, Italy

[d] Emergency department, Complejo Hospitalario Universitario de Albacete. Hermanos Falco, 37, 02006 Albacete

[e] Department of Urban Studies and Planning, Massachusetts Institute of Technology (MIT), 77 Massachusetts Av., Cambridge, Massachusetts 02139, US

Abstract

Background: COVID-19 has spread successfully worldwide in a matter of weeks. After the example of China, all the affected countries are taking hard-confinement measures to control the infection and to gain some time to diminish the big amount of cases that arrive to hospital. Although the measures in China reduced the percentages of new cases, this is not seen in other countries that have taken similar measures, such as Italy and Spain. Now we are in the middle of a battle trying to prevent the healthcare system from collapsing while it effectively responds to the needs of patients who are infected and require hospitalization.

Methods Using China as a mirror of what could happen in our countries and with the data available, we calculated a model that forecasts when are the peak of the curve going to happen. We aim to review the patterns of spread of the virus in the three countries and their regions, looking for similarities that reflect the existence of a common pattern in this expansive virulence and the effects of the intervention of the authorities with drastic isolation measures, to contain the outbreak.

Findings A model based on ARMA methodology and including Chinese disease pattern as a proxy predicts a reduction of the speed of contagious during the last week of March in Italy and smooth reduction until the peak in the second week of April. In the case of Spain, the infections still will follow growing fast until the end of March**.**

Interpretation: We present the following predictions to inform political leaders since they have the responsibility to maintain the national health systems away from collapsing. We are confident this data could help them into decision taking and place the capitals (from hospital beds to human resources) into the right place.

Introduction

The ongoing outbreak of viral pneumonia caused by SARS-CoV-2, globally known as COVID-19 is spreading fast and it is itself a stress test to the global public health, research and medical communities. The recent outburst in Italy and Spain along with the control of the disease in China has made the WHO to stablish Europe has the epicenter of the pandemic. A recent study shows that the spread of COVID-19 is unstoppable and has infected more than 100 000 people in 100 countries[1]. In fact, COVID-19 is nowadays considered as a pandemic threat by WHO[2].

Person-to-person transmission via droplets, contaminated surfaces or hands has been demonstrated[3] (*) even among asymptomatic carriers to close contacts at mainland China, where the outbreak started in January 2020 and it required of severe population control measures to manage it[4, 5]. One month later, the same happened in Italy with the origin in some imported cases from China. And two weeks later, similar trends emerged in Spain and the rest of Europe as well as in the US since the beginning due to Asian travelers back and, in the second wave, from European travelers.

The infection pattern followed in China can be seen in supplementary figure 1. The statistical information starts on Jan, 20th and just three days later, Wuhan city was closed in a set of strong measures of isolation which surprise worldwide. The evolution shows two phases: the first one with a significant and apparent exponential expansion, and the second one when a diminishing speed on infection follows until the new numbers are almost zero. From the application of the first measures, twelve days longer to reach the peak of contagious (first phase), and from then to the zero growth took another 26 days. The virus outbreak in China concentrated in Wuhan and Hubei province with a quite few dispersions through the rest of the regions, officially as the result of the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

isolation measures. These measures would have had the effect of changing the trend of the infection series but couldn't avoid the expansion in the first phase at daily rates of 47·3% on average until February 11. The decreasing rates started on Feb 12th and reached a daily rate of 0·32% at the beginning of March.

The first significant transmission outbreak in Europe was through Italy. From a few cases imported from China, the virus spread after few days following similar daily patterns: 40% of new infections on average day-to-day during the first period and with an average daily growth rate of 24% when the Government applied the harder containing measures March, 8th (supplementary figure 2). In Spain, the infection started as transmitted by leisure and business travelers from Italy. At the beginning of the first phase, the disease seems to be under control but it spreads from March 4 and accelerates from March 8 as a result of the expansion in Madrid. The territorial virus outbreak pattern of Spanish is similar to the other two countries, with the main focus concentrated in Madrid and other minor outbreaks in the Basque Country and Catalonia (supplementary figure 3).

Figure 1 around here

The three processes have, as similarity, that the transmission of the disease focuses in one region, where the majority of cases are collected (Wuhan in China, Lombardy in the case of Italy or Madrid in Spain) from which is spread out to the bordering regions although at a substantially lower rate. Assuming (for comparison purposes) that the day where main measures have taken place is the same date across the three countries, Figure 1 represents the potential 'common' evolution of the contagion in the three cities. The similarity in growth patterns is unique. Such similarity is taken as the base hypothesis in this paper through considering that the future evolution of the infection process would be similar in the European countries if the effect of the containing measures has got the same positive impact to that in China. To approach the time pattern Figure 2 represents the growth rates of the ordered series depicted in Figure 1 above. In this case, we stablished as day 0 whenever the county had around 200 infected people. From that moment we calculated the daily growth rates for each country. As it can be seen, the evolution of infection speed (measured in %) seems to be close among countries, and its evolution in shows a similar path in both Italy and Spain compared to China. In the first days, the data volatility is quite high since the total amount of cases are small. Therefore, a variation is easy to see when the total number of cases is smaller than when it is bigger. This can be seen after day 7 in Figure 2, where all the countries show clearer trends as infection numbers starts to be large. The speed of infection remains in a growth rate around 20% daily since day ten until day 22nd. Italy and Spain show an accelerated speed than China in this period.

Figure 2 around here

Methods and Modelling predictions

We obtained information on cases with confirmed COVID19 infection and diagnosis in China, Italy and Spain based on official reports from Health Minister in the case of Spain[6], from the Italian Department of Civil Protection[7] and from the National Health Commission of the People's Republic of China[7]. The data were collected in real time each morning, and it may be updated in the afternoon as new cases became publicly available. The latest update to the dataset was on March, 26th 2020. Specifically, we collected the dates of accumulated infected people each day.

The analysis of the data that has been carried out in this paper is purely statistical, based on a well-known signal extraction technic by using ARMA models[9] (Box-Jenkins methodology). It is done in several steps. First, a univariate analysis of the series of infected persons is carried out to identify the autoregressive pattern that shows them. The results reveal common pattern of disease spread in all three cases in all phases of the contagion. Second, the levels of acceleration are analyzed and a model is estimated in order to forecast the future infection path in Italy and Spain. Third, the contagion process between provinces is analyzed, looking for the pattern that could explain the reasons for the rapid expansion in some provinces and the slower one in others.

This article estimates ARMA models for three countries: China, Italy and Spain, at an aggregate level and each of their regions. The reason for choosing China is evident since it is the first country that has experienced the pandemic and has applied drastic measures for its control, constituting the unique evidence of the process to be taken into account. Secondly, Italy is selected because it was the first European country where the disease spread virulently. Its appearance occurred one month later than China, and it is spreading as a result of the contagion of few people travelling between the two countries. The Italian economic center was the first to be affected (Lombardy with Milan followed by Venice) as a clear reflection of the fact that a more significant part of global exchanges of people takes place for business or leisure trips. Spain was the third country to react to the pandemic, strongly affected as a result of its relationship with Italy. Spain warned of the effect and applied

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

drastic measures starting on March 14. This research predicts how long the contagious will last and the level of contagion reached in the latter two countries.

This paper's central hypothesis is that as COVID19 is a new virus, there is no immunity against it so it spreads out with no restriction among the population and it does not depend on the healthcare system (which is not prepared for its hardness and doesn't count with preventive elements such as a vaccine). If the hypothesis is correct, the spill-over effect in the infection process should show a similar pattern in all countries with equivalent live and health levels. We focus on how the first contagious spreads out, to define statistically the potential systematic pattern shown and to forecast when and where the process could end after the measures applied by the governments.

First, with empirical evidence the model estimates new infections at the aggregated level for every country. With this step we are asking the model to tell us the "story" of what has happened in the three countries under study from day 0 to day 45 (in China), until day 20 (in Italy and in Spain). The estimated parameters are a measure of how the contagious is expanding across the territory and they are shown in table 1 (called AR and MA) when they are significant. If left in blank, the coefficient shows no significant result and we left intentionally empty the cell for explanation purposes. AR (1) and AR (2) parameters show the speed of infection, positive in all three countries. Specifically, the model estimates the new infections as the consequence of the previously number of infected people and the trend of the preceding days. Moreover, there is a key difference between AR (1) and AR (2): AR (1) reflects the endogenous rise of the infection in the population while AR (2) shows the rise of the infection affected by external infected people.

On the other hand, The MA parameters reflect the influence on the viral spread of unexpected and non-observable reasons, like individuals moving across regions not being registered (in the case of positive impact) or the effect of measures applied, like isolation (in the case of negative value). The two patterns allow distinguishing between the direct infection (with the AR estimated parameters) and the indirect effect on the contagious (MA parameters).

In China, after the first case, during the first phase of infection, we can see AR (1) values ranges from 1.1 to 1.3 and they do not diminish as days are added in the expanding-windows models. This is suggesting a strong spread process, dependent only the number of infected in the country, which is exactly what happened. It is important to note that China does not show any AR (2) parameter significant, suggesting that no exogenous infection happened in this country. This makes sense, since the current outbreak of COVID-19 has its origin in China.

In Italy during the first phase, the parameter associated with exogenous infected (AR (2)) measures a direct effect from not local infections, suggesting that the first cases came from abroad. The value is also explosive capturing the rapid expansion, and on days 15 and 20 Italy shows endogenous expansion from the virus, meaning that is the virus dwells within the country. Curiously, this is exactly the same path followed by Spain, where in the first days the infection spreads by exogenous factors and as the series goes on in time, endogenous expansion gains strength and explains the current outbreak.

Focusing now on MA parameters, we can see a positive value of MA (1) in China in 15 days since the first data available, meaning that unknown components is enhancing the infection. The same happens in Spain the first 10 days, when restrictive measures have not been taken yet. Nevertheless, MA (1) turns negative in Italy and Spain on day 15 of the outbreak, possibly capturing how the application of restriction measures are contributing to reduce the infection. The same happens with MA (2) parameter, it becomes negative as the series goes on, reflecting external factors that are helping to reduce infection. The effect of those unknown components is to reduce the spread of the endogenous infection (AR (1)).

In all three cases, there is an extra moving average (MA (3)) component with positive parameter which acts in the opposite direction of MA (1) and MA (2), representing the power of some factors increasing infection from the past. The parameter value of this third MA component is very close during the last period in Italy and Spain, and a bit smaller for China. The interpretation of those results is that the negative MAs capture the effect of the initiatives taken to deal with the crisis: isolation and restriction seem to have had a sudden and rapid impact in all three countries, faster in Spain, with the time pattern changing in just three days. In Italy, it takes a bit more time to affect until the 20 first days. In China, the broader contention effects seem to happen since the 30th days (not before) as all previous periods register time pattern with a positive impact in the infection spread.

With the current data, the ARMA model supports that the time-path of China is being followed by both Spain and Italy, and this can also be seen with in Figure 2* with a less complicated model, measuring the daily growth rates. If the first part of the outbreak shows these similarities among countries, our hypothesis is that the contagion

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

patterns would have a common time path among the three countries. We assume that the time-pattern found is reliable and therefore, we think that the event of infection growth can be predicted.

With all the above, a new model is defined for forecasting purposes. The model also tests the hypothesis that Spanish and Italian's contagious processes follow the path shown in China and adjust it to their own ARMA pattern, that is, representing a common way of the virus to expand itself adapted to each reality. The results of forecasting for Italy and Spain are shown in figure 3. We also performed a panel prediction by regions in the case of Italy and by autonomous communities in Spain using the Chinese process as an instrument variable in the autoregressive model to forecast infections (data not shown).

Figure 3 around here

Italy was the European country which first exhibits the significant infection process and the first in which the authorities reacted by applying strong measures of isolation to content the virus expansion. There is a similar time pattern than those happened in the first moments in China (data no shown) which support the hypothesis of common global expansion pattern. Besides, Italy shows strongly significant differences by region in terms of number of infected people, just like China. Specifically, Lombardy, Veneto, Piemonte and Emilia-Romagna carry a 43·3%, 8·6%, 8·1% and 13·4% respectively of the total number of infected in the country at March, 26th. ARMA model predicts a reduction of the speed of contagious during the last week of March in Italy and smooth reduction until the peak in the second week of April, with a control on new infection happening in most regions including Piamonte, smaller speed in Veneto. The model does not predict a diminishing in new Covid-19 cases in Lombardy and Emilia-Romagna, with the current conditions ongoing, suggesting that stronger measures should be applied.

Spain is in the earlier stages of the contagion process. Results reflect a precise association with the Chinese pattern during the first phase of the infections as Spanish process seems not to reach the turning point experienced in China at the moment to write this paper. Just like in the case of Italy and China, in Spain there are regions that show a stronger relationship with the global infection pattern. In this case, Madrid and Catalonia have the 30·5% and 20·6% respectively of all the infected. In the case of Spain, the model predicts that infections still will follow growing fast until the end of March since that will start to reduce the speed of contagious. Note that the forecasting has been calculated by taking into account the unfinished first phase of the contagious process. Regarding the prediction until the actual data, the maximum level of contagious in Spain will overpass the Italian due to the fast growth in the Spanish number of cases. The precision of this comparison creates an uncertainty in the Spanish case about the moment when the infection rate reaches the turning point (observed in Italy but not still in Spain). In this issue, the alternative public measures and their success are essential to accelerate the diminishing infection speed and the consequent peak.

Discussion

This article shows a similar behavior of the infection in all the studied countries, and based on this fact, we used the ARMA model to predict the future behavior of the ongoing COVID-19 outbreak in Italy and Spain. The results show that with the ongoing restrictive measures, we estimate around 141 and 142 thousand cases for Italy in the maximum peak of the curve and the numbers for Spain will be around 152 thousand. This prediction is based on research tools to give an idea of what might be the damage to the healthcare system or the load burden in the future weeks.

Fortunately, not all the infected cases will require urgent medical assistance, such as intensive care unit, and this is the key point to maintain the hardest restrictions to population's mobility. With the current data, around 5·7 – 5·8% of total infected requires intensive care in Italy and Spain. To avoid healthcare system to collapse is central in this epidemic, as other authors state. The fact that much more is known in a week about the complications added to viral infection and its management can help save lives. Regardless of the numbers being handled today, perhaps this prediction can help the authorities to make better decisions on how to manage the available resources, because it gives an idea of what could happen in the future.

We acknowledge the existing underdiagnosis that affects the data provided by authorities. In Spain, as our knowledge goes, COVID detection tests are only made in people admitted into the hospitals and their close contacts. Therefore, the real number of cases could be higher than the ones we estimate in this study. Also, this prediction is made with ongoing confinement measures in Italy and Spain, thus these predictions could vary if people were allowed to move freely and face the risk of new exposure to the virus or the measures become more restrictive.

Finally, this estimation has been obtained in developed countries that applied efficient contention measures in a very short time, thus we fear the hardness of the outbreak in less developed countries with less resources or in the case that restriction measures are not taken. There will be a constant increase in the number of cases in

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

the following weeks and all the healthcare systems must be ready to take the lead, as it has already happened in China, Italy and now Spain. We encourage to all the authorities to take strong measures to minimize the effect of the outbreak in their countries, such as social distancing, forbid people's movement and promote basic preventive measures such as hand washing.

Contributors: PM, ZY, LG were responsible for data recollecting, PM did the data analysis and statistical models. PM and PL wrote the manuscript. AM and PM introduced the medical perspective in the manuscript.

Declaration of interest: We declare no competing interests.

Acknowledgments: None.

References:

1. Callaway E. Time to use the p-word? Coronavirus enter dangerous new phase. Nature 2020; 579: 12.

2. WHO Director-General's opening remarks at the media briefing on COVID-19. (https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020). Accessed 12 March 2020

3. Paules, Catharine I., Hilary D. Marston, and Anthony S. Fauci. "Coronavirus infections—more than just the common cold." Jama 323.8 (2020): 707-708.

4. Bai Y, Yao L, Wei T, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA. Published online February 21, 2020. doi:10.1001/jama.2020.2565

5. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020; published online Jan 29. DOI:10.1056/NEJMoa2001316.

6. Italian Ministerio della Salute, Covid-19 Bollettino, Protezione Civile, available at http://www.salute.gov.it/portale/news/

7. Spanish Ministry of Sanidad, available at https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/

8. National Health Commission of the People's Republic of China (中华人民共和国国家卫生健康委员会), available at http://www.nhc.gov.cn/ and http://en.nhc.gov.cn/DailyBriefing.html (English version).

9. Box, G. E., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control Holden-Day. San Francisco, 498.

10. Remuzzi, A., & Remuzzi, G.COVID-19 and italy: What next? The Lancet, doi:10.1016/S0140-6736(20)30627-9

Declaration of interest: the authors declare no conflict of interest

Funding source: no funding source

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

table 1

Table1: Time autoregressive pattern in the infection process of COVID-19. Three countries compared

	SPAIN			ITALY			CHINA		
Sample	10 Days	15 Days	20 Days	10 Days	15 Days	20 Days	15 days	30 days	45 days
AR (1)	··	1.101±0.361	1.009±0.400	··	1.573±0.366	1.176±0.274	1.068±0.439	1.285±0.237	1,506±0,345
AR (2)	0.918±0.433	··	··	1.4±0.254	··	··	··	··	··
MA (1)	1.604±0.563	-5.90±0.201	-1.970±0.739	··	-1.095±0.197	-0.695±0.075	0.678±0.278	-0.616±0.244	-0.992±0.3
MA (2)	··	-0.492±0.242	1.262±0.409	-0.981±0.219	1.151±0.203	-0.68±0.188	··	-0.645±0.266	··
MA (3)	··	0.835±0.144	-2.027±0.590	··	··	0.874±0.076	··	0.970±0.115	0.531±0.18

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

List of Figures

Figures

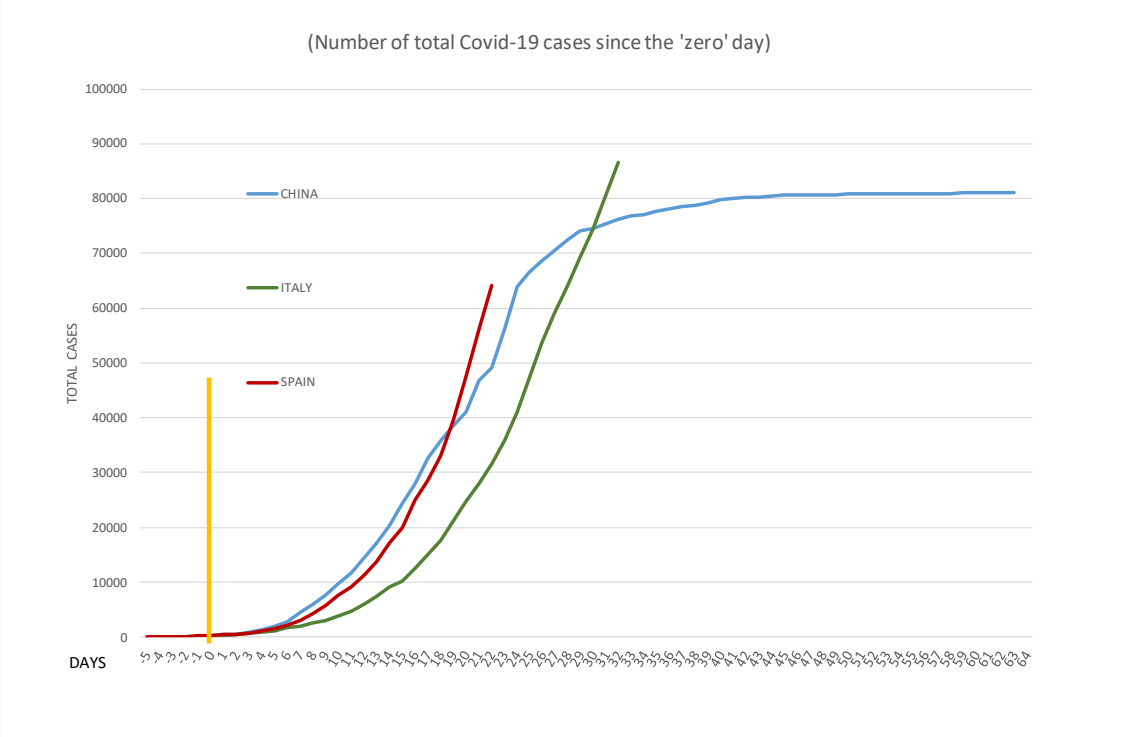


Figure 1. Contagion patterns in COVID-19. Three countries

This Figure represents the COVID-19 infection process in China, Spain and Italy starting on the same moment, indicated by a yellow line. This moment is the day in which each country has account for around 200 infected individuals. The similarity in growth patterns is relevant. Such similarity is taken as the base hypothesis in this paper through considering that the future evolution of the infection process would be similar in the European countries if the effect of the containing measures has got the same positive impact to that in China

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

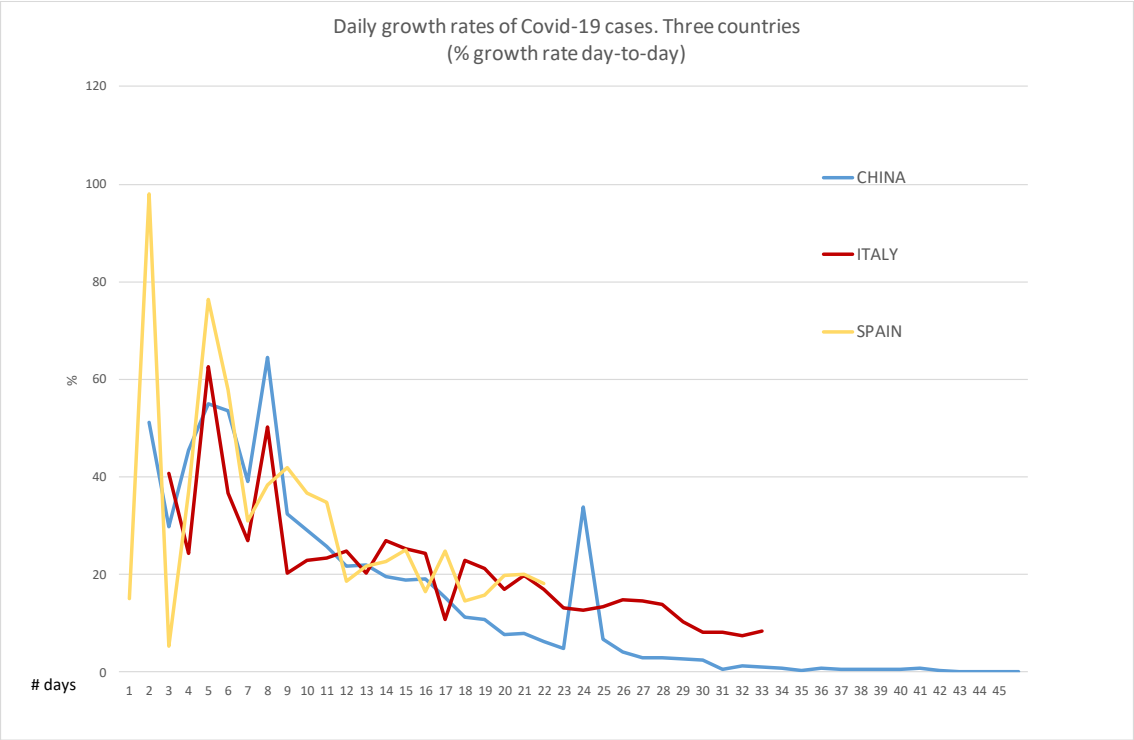


Figure 2: Diary growth rates of the infection in percentages for China, Spain and Italy.

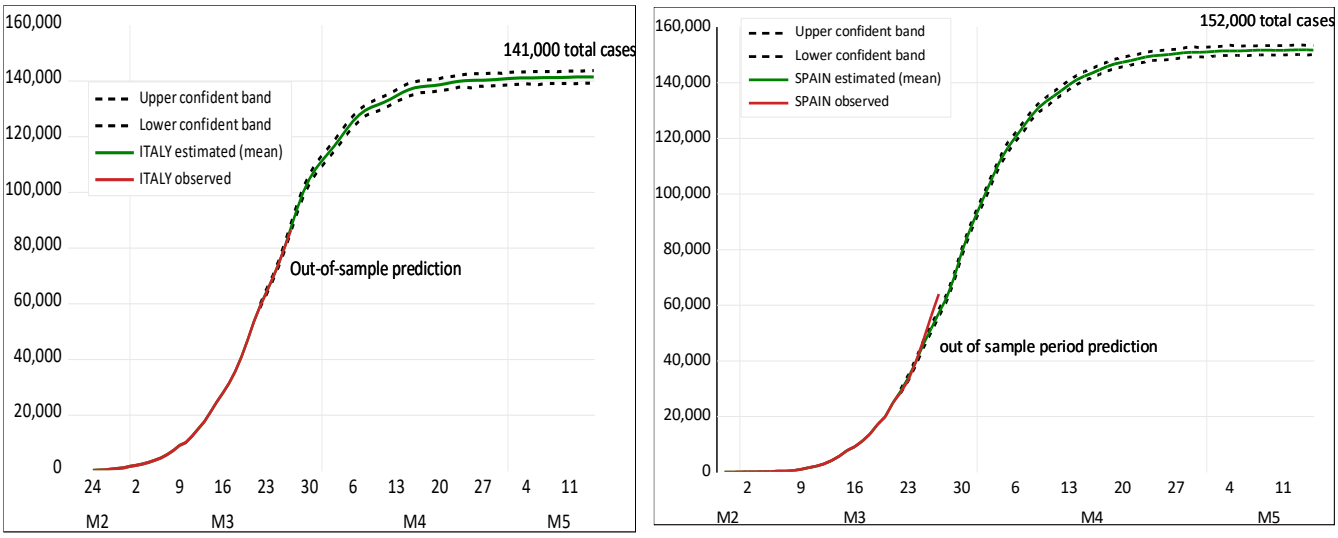


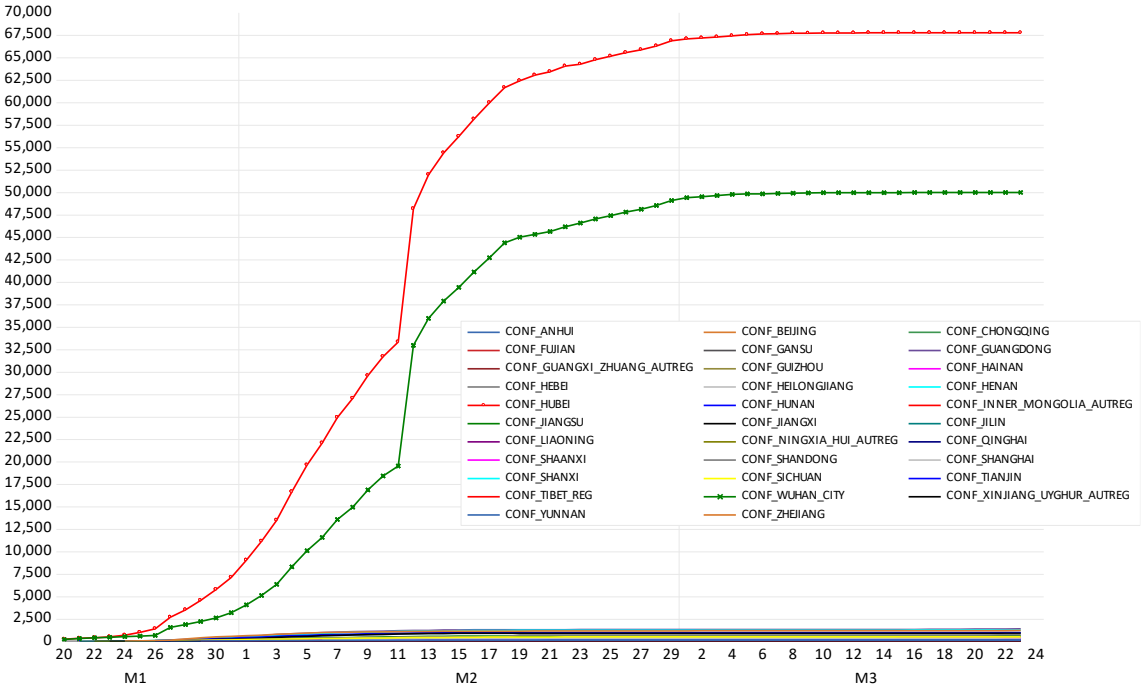
Figure 3: Forecasting for Italy and Spain according to the ARMA model

This figure represents the time pattern of COVID-19 contagion in Italy and Spain. The red lines are the observed (real) data of total cases while the green line is the estimated path of the future contagion. The black dot lines represent the confident band. The coincidence of both represents the estimation precision and has been calculated on a five-days base. The estimation has been made using stochastic method based on a ARMA model.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

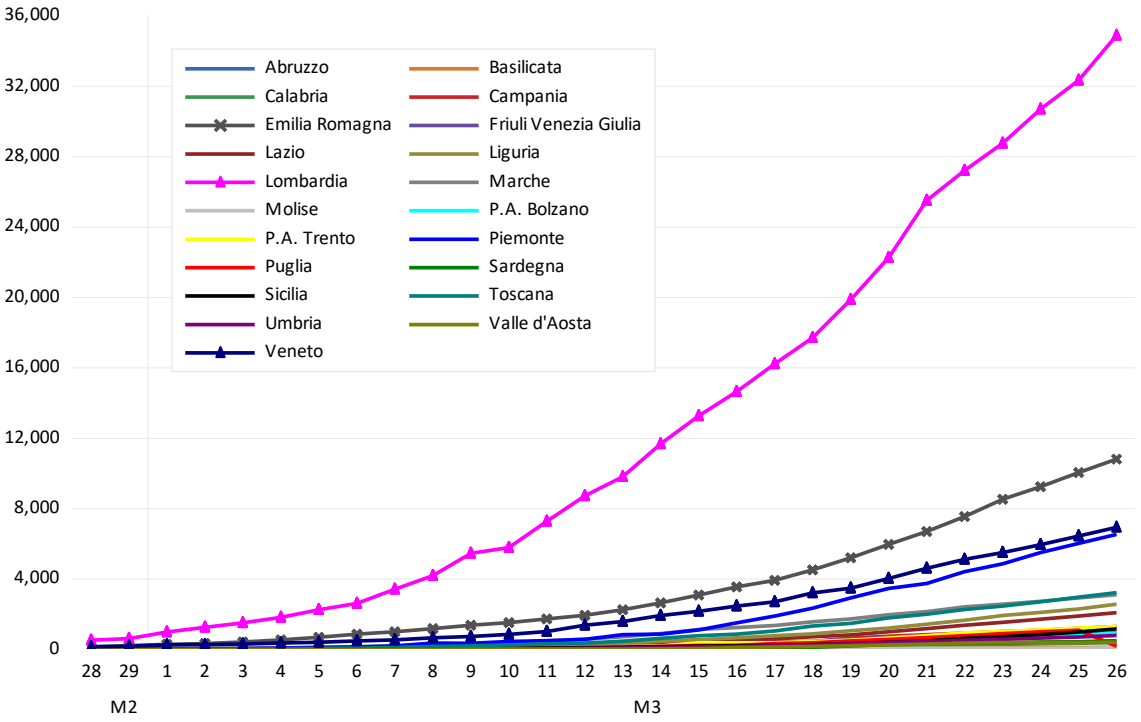
SUPLEMENTARY FIGURES

Suplementary Figure-1. Covid19 cases evolution in China: Jan 20th to March 22th



Source: 中华人民共和国国家卫生健康委员会, National Health Commission of the People's Republic of China, http://www.nhc.gov.cn/ , English version web: http://en.nhc.gov.cn/DailyBriefing.html

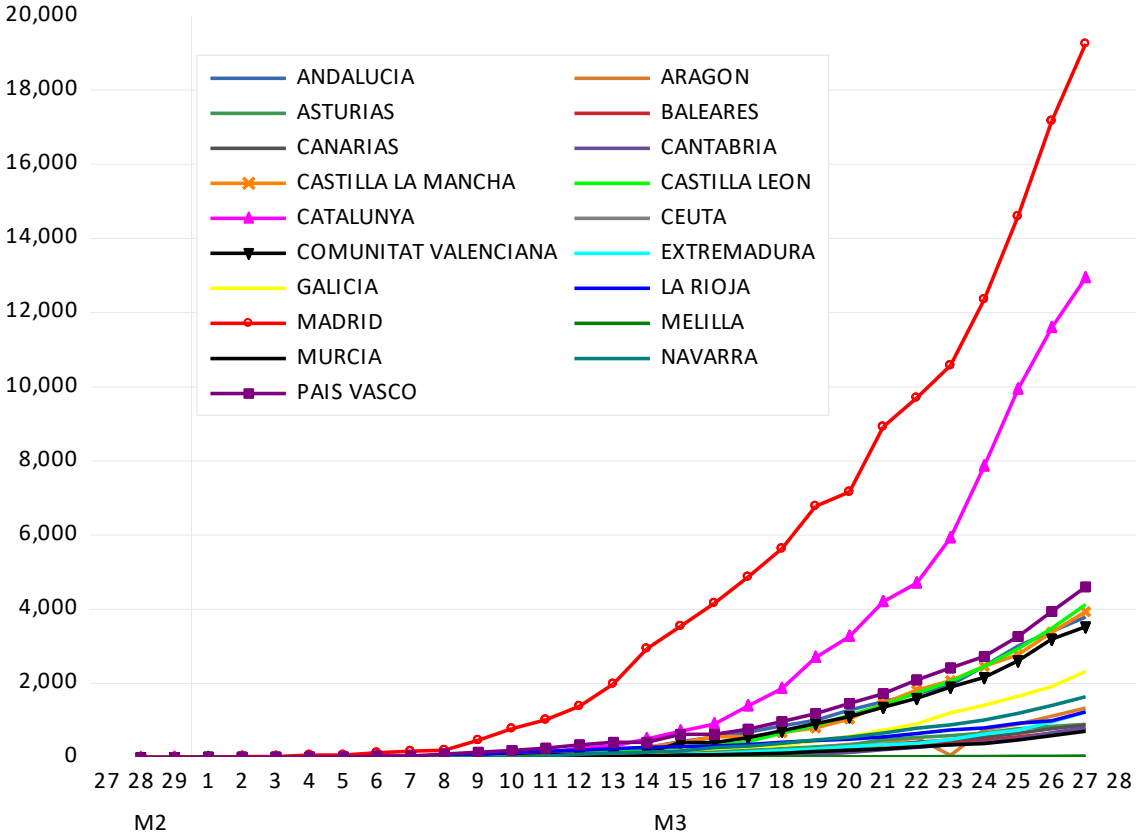
Suplementary - Figure 2. Covid19 infection pattern in Italy: Feb 28th to March 26th



Source: Bollettino Covid-19, Protezione Civile, www.salute.gov.it/portale/news. Several days

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

Suplementary Figure 3. Covid19 infection pattern in Spain: March, 3rd to 26th



Source: Ministerio de Sanidad, https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/
Several days

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

Figure 1

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

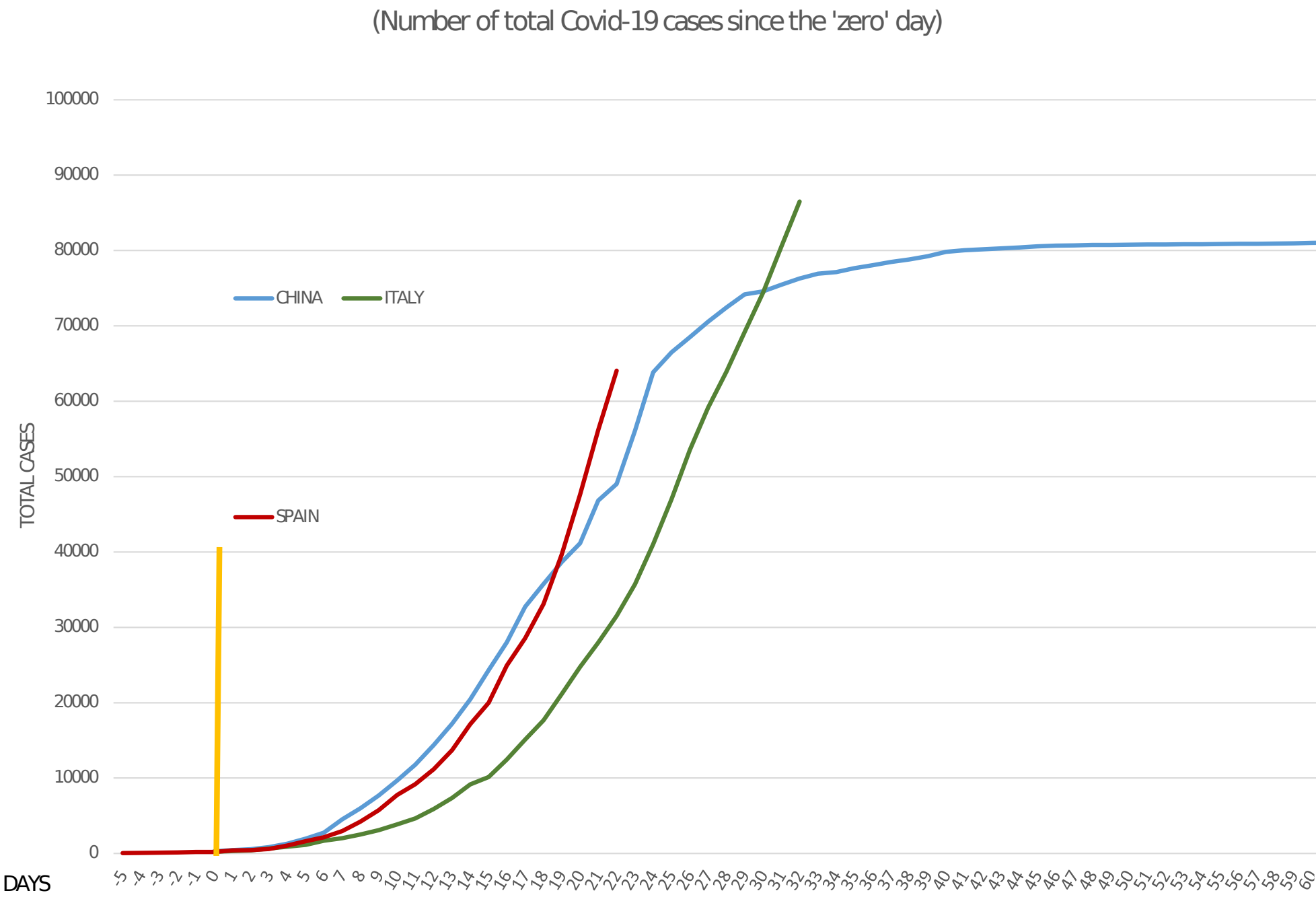


Figure 2

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

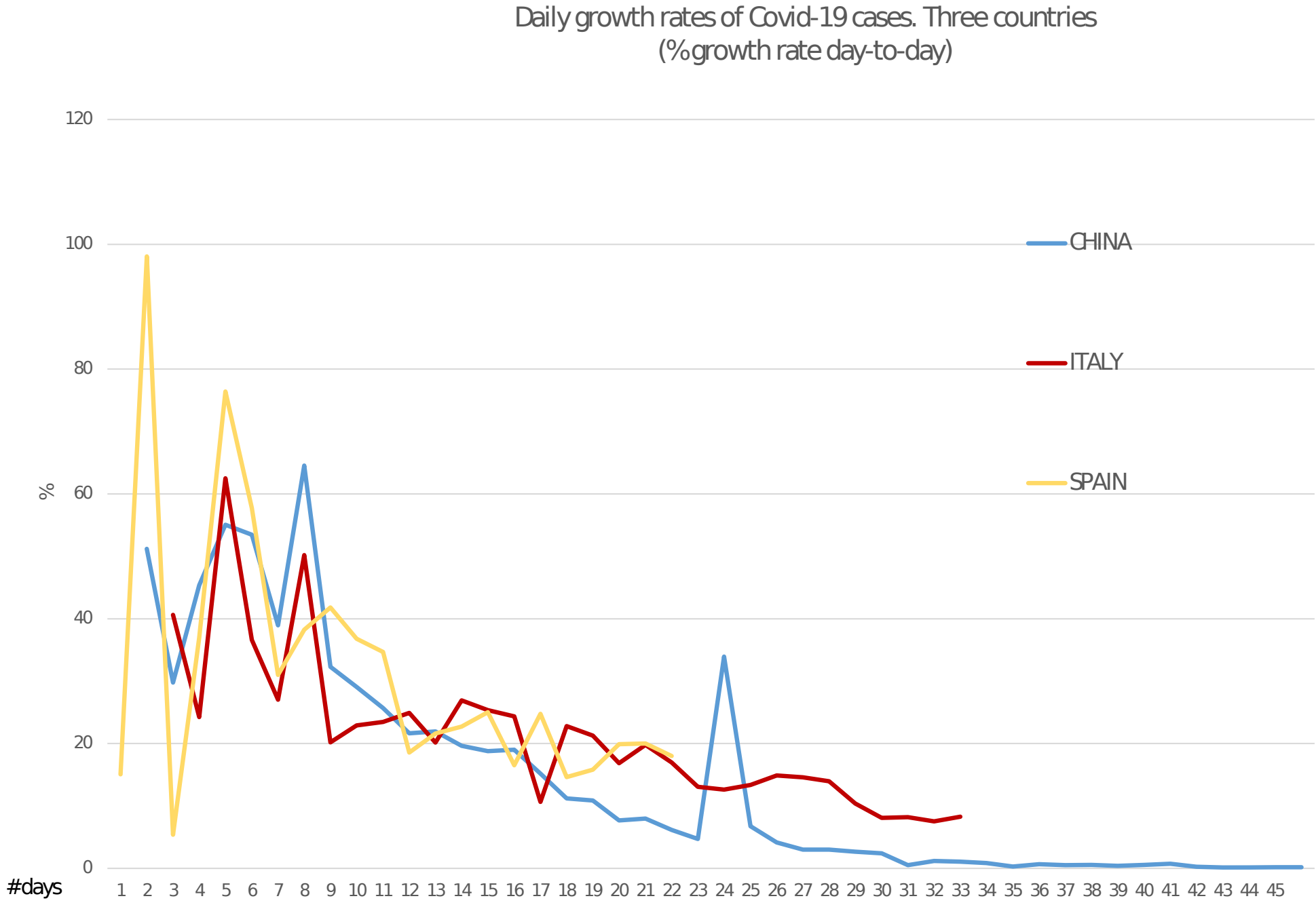
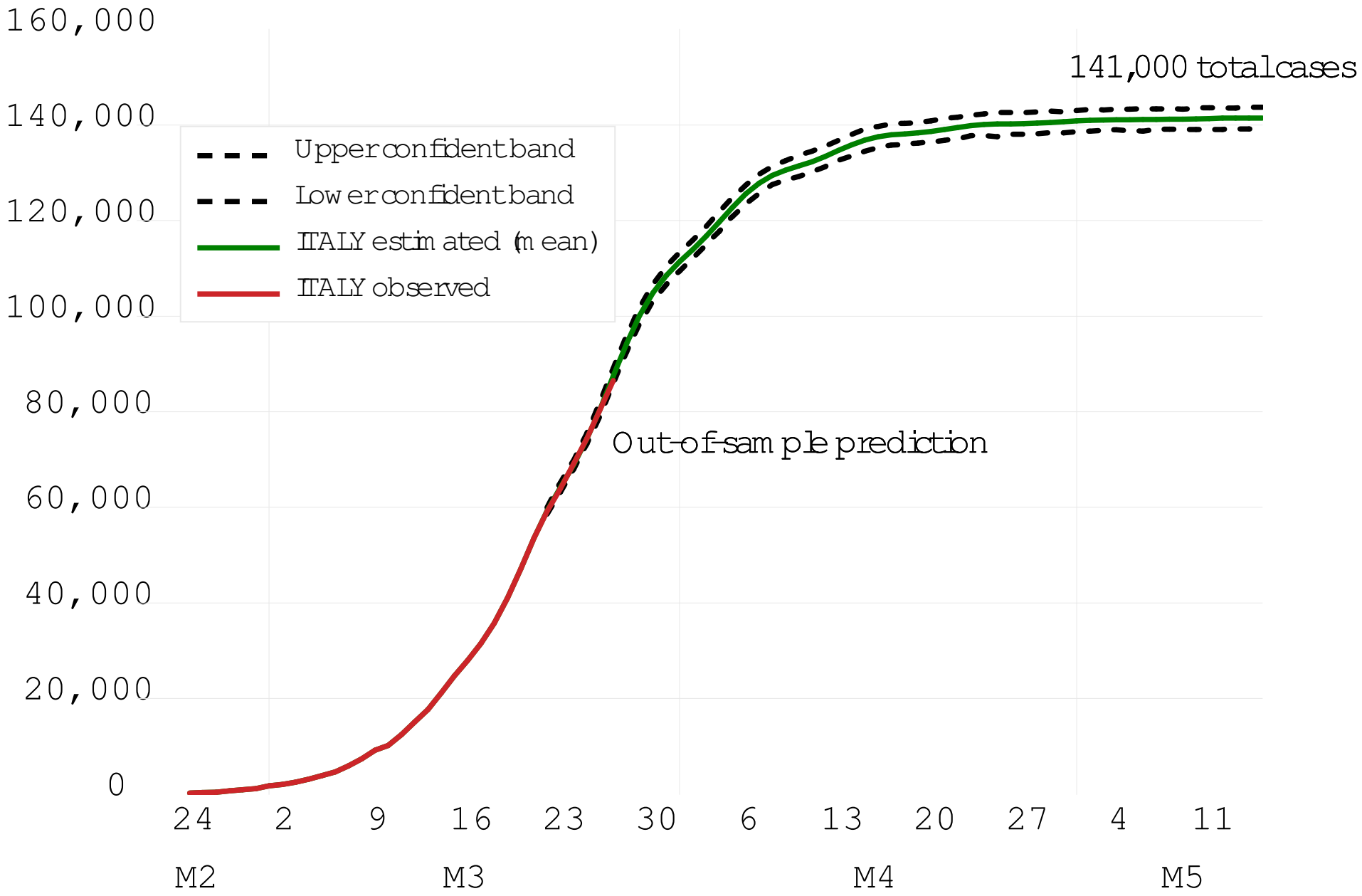
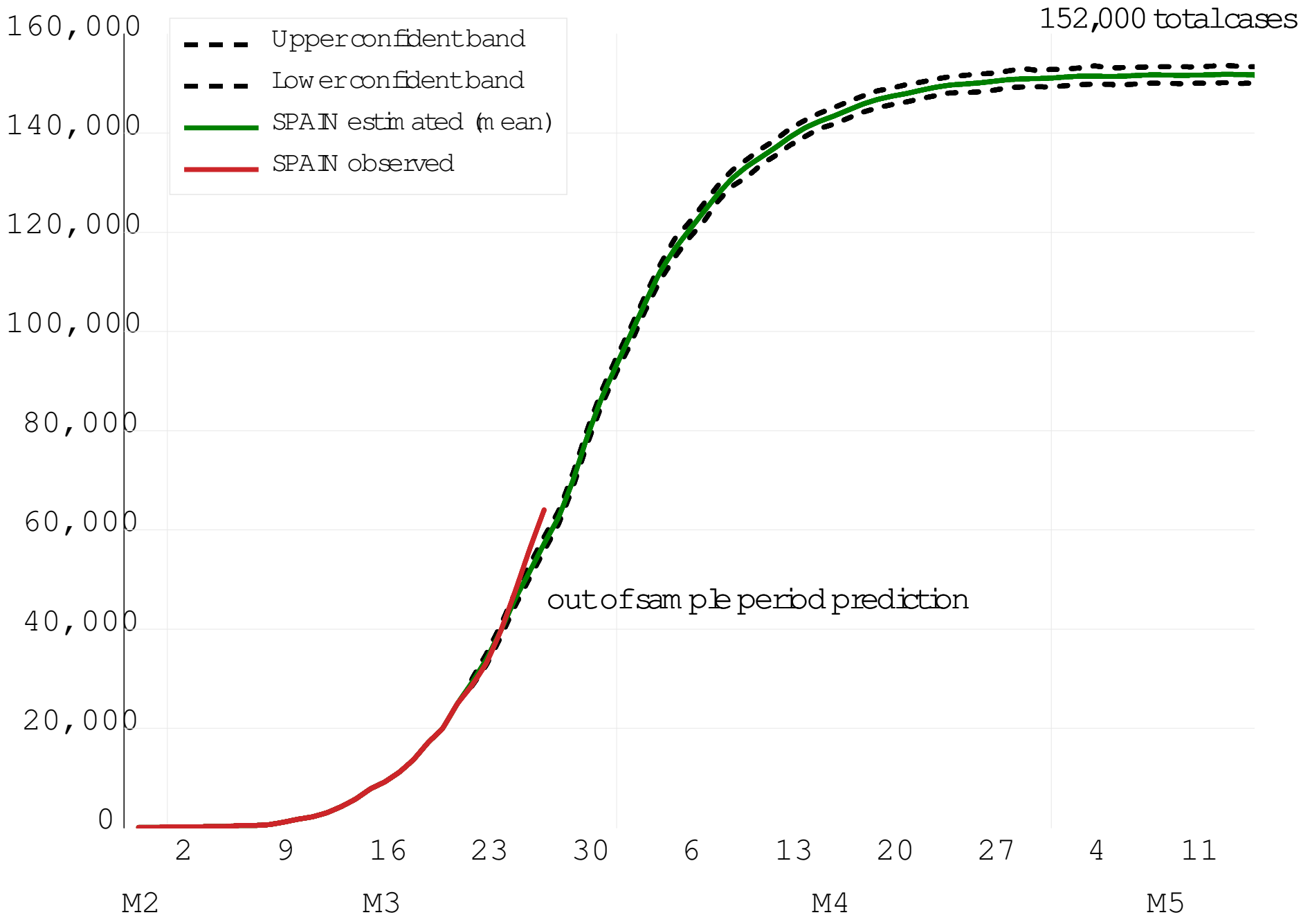


Figure 3, panel 1



This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

Figure 3, panel 2



This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

Suplementary Figure 3

Click here to access/download
Supplementary Material
supplementary_figure3_spain.eps

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

Suplementary Figure 1

Click here to access/download
Supplementary Material
supplementary_figure1.eps

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150

Suplementary Figure 2

Click here to access/download
Supplementary Material
supplementary_figure2_italy.eps

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3566150