



THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■



Synthetic Forecasting Method for COVID-19: Applications to UK and Turkey

Orkun Saka

SRC Special Paper No 19

April 2020



Systemic Risk Centre

Special Paper Series

Abstract

Synthetic control method (SCM) is generally used for producing counterfactual scenarios in policy evaluations. In the light and urgency of the recent developments regarding the COVID-19 virus and the related human and economic losses, I propose a synthetic forecasting method (SFM) in order to estimate the future path of virus-related outcomes in a country. The method takes advantage of the fact that the virus entered into different countries in slightly different points in time. An application to the cumulative COVID-19-related death toll (as of 13 April 2020) shows that the total numbers may increase by 50% in a week in UK and may double in two weeks in Turkey.

Keywords: Synthetic Forecasting Method (SFM), COVID-19.

This paper is published as part of the Systemic Risk Centre's Special Paper Series. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/R009724/1].

Orkun Saka, University of Sussex, London School of Economics and Systemic Risk Centre,
London School of Economics and Political Science

Published by
Systemic Risk Centre
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

© Orkun Saka, submitted 2020

Synthetic Forecasting Method for COVID-19: Applications to UK and Turkey*

Orkun Saka[†]
University of Sussex, LSE & SRC

Preliminary Draft: 14 April 2020

Abstract

Synthetic control method (SCM) is generally used for producing counterfactual scenarios in policy evaluations. In the light and urgency of the recent developments regarding the COVID-19 virus and the related human and economic losses, I propose a synthetic forecasting method (SFM) in order to estimate the future path of virus-related outcomes in a country. The method takes advantage of the fact that the virus entered into different countries in slightly different points in time. An application to the cumulative COVID-19-related death toll (as of 13 April 2020) shows that the total numbers may increase by 50% in a week in UK and may double in two weeks in Turkey.

Keywords: Synthetic Forecasting Method (SFM), COVID-19.

*This is a work in progress and subject to errors given the timescale in which the paper was written. All interpretations, errors, and omissions are those of the author.

[†]ORKUN SAKA is a Lecturer (Assistant Professor) at the University of Sussex, a Visiting Fellow at the London School of Economics and Political Science (LSE) and a Research Associate at the Systemic Risk Centre (SRC). Email: o.saka@sussex.ac.uk.

“... we should all understand — everyone, everywhere — that we are at different points in the same story; that in this pandemic we share the same timeline: some are a little further ahead, some a little further back.” (Paolo Giordano - Financial Times, 27 March, 2020)

1. Introduction

COVID-19 virus that first originated from Wuhan city of China quickly found its way to the Western hemisphere wreaking havoc on people’s lives and massively depressing already-fragile economies. National governments responded in many different ways ranging from closing workplaces and introducing full lockdowns to intervening in the economy via unprecedented monetary and fiscal measures (Hale, Petherick, Phillips, and Webster, 2020).

Due to the rapidly increasing number of cases and deaths in many countries, it has become vital to be able to predict the future evolution of COVID-19-related outcomes. Although such forecasting ability is crucial to be able to prepare for what’s coming in each country, the massive uncertainties surrounding this new virus (such as its still unknown infection and fatality rates) restricts our ability of foreseeing the path forward.

In this paper, I introduce a new method, namely Synthetic Forecasting Method (SFM), that takes advantage of the fact that COVID-19 virus entered into different countries in different points in time. Hence, assuming that each country is located in a different point of an otherwise similar timeline, I effectively use the information derived from the “leader” countries to be able to forecast the future paths of the “follower” countries that were infected by the virus at later stages of the pandemic. In order to produce a forecast for a *follower* country, I first match it with a synthetic (i.e., weighted) combination of *leader* countries that is best able to replicate its current outcome path. This part is done via the classic synthetic control method (SCM) often employed in evaluation studies of past policies. Once the best match and the resulting weights are calculated, I project these weights on the same countries for the following period in order to obtain a synthetic forecast.

The main advantage of this new forecasting method is the fact that it is model-free to a large extent and does not assume a certain functional form or make raw extrapolations from the past data. Instead, it makes intuitive use of the already-realised outcomes of the *leader* units in order to forecast an “average” outcome for the *follower* unit.

In the next section, I introduce SFM and apply it on the COVID-19-related death rates in two countries, namely United Kingdom and Turkey. In Section 3, I briefly discuss several areas where it could prove useful. The last section concludes.

2. Method & Application

2.1. Synthetic Forecasting Method (SFM)

Synthetic methods have been in use since the seminal paper by [Abadie and Gardeazabal \(2003\)](#). Overall, SCM helps researchers evaluate the effects of an idiosyncratic shock (treatment) introduced at a specific time point to a single unit in a panel. It achieves this by creating a counterfactual outcome for the shocked unit by optimally combining the outcomes for non-shocked units. Optimality here refers to choosing such weights that minimize the pre-shock distance in outcomes between the shocked and the weighted average of the non-shocked units. These weights derived from the pre-shock term are carried forward to post-shock period in order to represent the counterfactual outcome for the shocked unit. For the sake of brevity and time, interested readers can refer to the related literature in order to understand the process of matching between shocked and non-shocked units.¹

Despite its widespread usage in comparative case-studies that analyse past outcomes, synthetic approach has rarely been adopted to a forecasting setting. To the best of my knowledge, previous literature has not yet delved into this potentially useful line of research in a way that I will describe below.² On the other hand, there could be several advantages in employing a synthetic approach in forecasting:

1. Given the unknown characteristics of a new shock (in this case the unknown infection and fatality rates of a new virus, namely COVID-19), SFM can help researchers quickly learn from the information gained from other countries which received the same shock at an earlier time-point.

2. Instead of assuming a certain functional form or trend-fitting (see [Linton, 2020](#)), SFM can generate more realistic forecasts by directly basing its estimates on the realised outcomes of other comparable units.

3. SFM is flexible in terms of generating different forecasts based on different pre-conditions in the shocked unit. For instance, one could forecast the evolution of the COVID-19-related death toll in a country by explicitly assuming a certain policy stance (lenient or

¹See, among many others, [Abadie, Diamond, and Hainmueller \(2010\)](#) as well as [Abadie, Diamond, and Hainmueller \(2015\)](#).

²Two exceptions are [Meyersson \(2018\)](#) who kindly shared with me his work on synthetically forecasting Turkish macro outcomes after 2018 currency crisis and [Klößner and Pfeifer \(2018\)](#) who use the lagged versions of the outcome variable to create a donor pool in order to extrapolate from the past outcomes to the future. The SFM method I propose below differs from the former in the way that it chooses its shock: while [Meyersson \(2018\)](#) chooses past currency crises in history to match with that of recent Turkey, I consistently use the exact same shock (i.e., COVID-19) only entering into different countries at similar but slightly different time points, thus reducing the chance of time-dependent heterogeneity across comparable country cases. SFM also diverges from the latter as its fundamental tenet is to replicate from realised outcomes rather than extrapolating from past observations.

strict) and only including such countries in the control group.

In a standard synthetic control method, there is usually a country (i.e., unit) that receives the treatment at a certain point in *calendar* time. A weighted average of the untreated countries is then matched with the treated country *before* that time point and then compared to the realised outcomes of the treated country *after* that time point in order to infer the causal impact of the treatment. The Synthetic Forecast Method instead does not work with the usual calendar time. It assumes a starting point with the introduction of a shock and tracks all countries alongside this new timeline.

For example, Figure 1 shows the trends in COVID-19-related death toll since the day cumulative number reached a total of 10 for the first time in a country, according to the data extracted from the European Centre for Disease Prevention and Control (ECDC). As as of 13 April 2020, it is clearly visible that different countries are in different stages of the same pandemic and there are many countries showing similarities and sharing common trends in different parts of this new timeline.

Some countries in Figure 1 are more likely to be *leaders* and some are *followers* seeing their numbers increasing only recently. Thus, the fundamental logic of the SFM builds on this very point that the future evolution of the outcomes in the follower countries can be forecasted by replicating their current outcomes with the early experiences of the leader countries. Once a weighted average of the leaders is chosen, these weights can be carried forward on the same timeline in order to predict what is likely to prevail in a follower country in the immediate future.

An important point to note here is the fact once a country is chosen, the choice of forecasting horizon depends very much on *how many* countries are leading the chosen country in this timeline and by *how much*. Therefore, SFM becomes more difficult to apply in countries that have contracted the virus earlier than others (such as China or Italy); however these countries can still usefully act as a control group for the countries infected by the virus at later stages.

In general, if the shock defining the new timeline is introduced only with small leads/lags in different countries, this restricts the forecasting horizon for an average country in the sample. However, the advantage of smaller leads/lags is the potential reduction in time-dependent heterogeneity across different countries. If they had received the shocks in very remote points in the calendar time, this might bring into question the comparability of the cases that could be used as a control unit in a synthetic match.

2.2. Two applications: UK and Turkey

United Kingdom is one of those countries that received the virus at a relatively early stage, making it more of a candidate to be a *leader* country in SFM setting rather than a *follower*. Hence, in terms of forecasting power, we cannot expect to come up with a long horizon as there will be very few countries more experienced than UK. Figure 2 illustrates where United Kingdom currently stands with respect to its *leader* countries (i.e., China, Italy, Iran and Spain) with a horizon choice of 8 days. If one was to extend this horizon by one more day, this would lead Spain to drop from the figure and thus critically reducing the number of potential cases that one could use in a synthetic match.

Another degree of freedom comes from the threshold value of total deaths that determines where the timeline starts from. A lower threshold would naturally extend the length of the timeline over which one can perform a synthetic match; however it is unlikely to change the set of *leader* countries since this is somewhat predetermined by the (calendar) day on which the virus arrived at the country. Choosing a longer timeline on the other hand may increase the noise in the synthetic forecast by making the longer series of the follower harder to match with the set of series of a few leader countries. In Figure 2, this threshold is set as 500 deaths.

Given the timeline in Figure 2, I compute the optimal weights that minimize the distance between the natural logarithm of the current UK series and a weighted average of those of the leader countries over the initial 18 days until 13th April.³ Projecting the same country weights on to the next 8 days in the future and using the realised outcome values of the same set of countries produces the synthetic forecast for UK which is illustrated in Figure 3. As can be clearly seen, the forecast closely tracks the UK series over the initial period and worryingly predicts a 50% increase in the total death toll in less than a week. Countries with the largest weights in this synthetic forecast are Spain (56%) and Italy (40%), which is not surprising at all given the similarity across country trajectories pictured in Figure 2.

A potential robustness check of the results could be undertaken by performing the same exercise by excluding one country at a time from the pool of *leader* countries. Figure 4 presents the results when I re-compute the synthetic forecasts several times for different pools. It is clear that results generally point to the same direction with similar trends, except the case without Spain, which is sensible as this country enters the general synthetic forecast with a large weight. However, even the exclusion of Spain does not change the synthetic forecast to a large extent and its main prediction (50% increase at the end of the next 8 days) seems to remain intact.

³This is performed through the usual optimization technique employed in the standard synthetic control method. See [Abadie et al. \(2015\)](#).

An additional check on the performance of synthetic forecasts could be undertaken by restricting the timeline over which the synthetic match is determined. In this way, one could compare the out-of-sample forecasts generated by the method to the realised values for the same country and evaluate how much they may deviate from reality. In order to track the performance of the synthetic forecasts in April, I include in the optimization window only those observations experienced in March (i.e., the first 5 days). Thus, I can compare the forecasts generated only with the information available in March to see how accurately they may predict the real experience in April.

Figure 5 illustrates that, despite the extremely short estimation window (5 days), forecasts generated by SFM is able to closely predict the real number of cumulative COVID-19-related deaths in UK since the beginning of April. Notice that the forecasts slightly overshoot the real values from time to time but they re-converge towards the end of the 13-day horizon. This finding is somewhat reassuring in terms of the forecasting power of the model in the medium-run.

The fact that COVID-19 virus entered into UK at an early stage restricts our forecasting horizon as discussed above. To evaluate how SFM may perform in a longer horizon (2 weeks), I pick a country, namely Turkey, that received the virus at a relatively later stage. This can help me forecast further ahead as Turkey is likely to have more *leader* countries that are more advanced in the COVID-19 timeline than itself. Figure 6 shows the results when I estimate the synthetic forecasts by starting the timeline from the first day on which the death toll in Turkey reached a total of 50.⁴ As visually confirmed, the model is fitted to the data quite well during the estimation window (initial 19 days till 13th April) and predicts that the death toll may double in the next 2 weeks.

China plays a huge role in Turkey’s synthetic forecast with a weighting above 95%. Thus, Figure 7 evaluates the robustness of these forecasts to the one-by-one exclusion of a *leader* country in the comparison group. The forecasts have similar directions; however they rise more rapidly if one was to drop China from the synthetic matching process. They also seem to get noisier at the long end of the horizon. This observation is not so surprising given the small number of comparable countries to establish the matching and longer forecasting window. Alternatively, this analysis can help policymakers qualitatively assess the future path of Turkey by incorporating their subjective assessments of which countries Turkey is mostly likely to resemble to given its current policy path and level of government interventions. Since the country’s stance has so far been strict, keeping the countries that had more

⁴Lower threshold in this case is necessary to retain a similar number of days (19 in this case) in the estimation window since the total number of deaths in Turkey are currently much lower than that of United Kingdom.

interventionary approaches (i.e., China) in the control group could generate better forecasts, for instance.⁵

Finally, Figure 8 shows the prediction performance of the method during April in Turkey similar to what I did with UK numbers in Figure 5. Likewise, the forecasts seem to do well on average prediction and direction of the real trends despite the fact that the model is fitted with only 6 observations in March.

3. Discussion

Synthetic Forecasting Method (SFM) could be a useful tool especially for “live forecasting” the immediate horizon during crisis events of a new nature as in pandemics. It can be useful not only in terms of predicting the health-related outcomes but also estimating secondary variables, such as demand in hospital beds or domestic violence due to lockdowns, which may be directly or indirectly impacted by the original shock.

Other than pandemics, one could use the method for the purpose of predicting the outcomes of a rare event. For instance, Meyersson (2018) applies a similar method in the context of 2018 currency crisis in Turkey and forecasts the Turkish macro outcomes by synthetically matching this period with periods in the past when other countries experienced similar currency crises. The same logic could be valid in predicting the outcomes of natural disasters such as earthquakes or rare political shocks such as membership in the European Union.

A caveat of the method in the context of pandemics is that forecasts are bound to be very short-term -and thus potentially less helpful- for countries that have started their experience earlier than others. Such short-horizons may not be particularly useful for the policymakers if they need more time in advance to prepare their response. However, it can still be useful at least for some of the countries that are new to the pandemic and thus can learn from the early experiences of others. SFM could then provide a systematic way to perform such comparative analyses in a quick and robust fashion.

Another potential concern is the reliability of the statistics employed in synthetic forecasts. For instance, there have been several reports of some countries under-reporting their COVID-19-related case and death statistics.⁶ If that is the case in many countries and the extent of such reporting biases changes over time, it would be unavoidable to expect noisier estimates from any forecasting method including SFM.

⁵The Covid-19 government response tracker compiled by Hale et al. (2020) shows that Turkish government has so far been quite strict in its management of the pandemic with an average intervention index of 95% for the period covered in my synthetic estimations above.

⁶See, for instance, the following report in CNN: <https://edition.cnn.com/2020/04/07/uk/coronavirus-uk-deaths-intl-gbr/index.html>

Lastly, a more general caveat -as it is the case in all synthetic models- is the fact that SFM can be sensitive to model specification such as number of countries in the comparison group, length/duration of the estimation period or the inclusion/omission of the potential covariates.⁷ Even though the restrictive context of a pandemic setting is likely to exacerbate some of these problems, the flexible nature of the synthetic method also provides opportunities for coming up with various robustness checks as I have done in the previous section of this paper.

4. Conclusion

In this paper, I propose a synthetic forecasting method (SFM) in order to estimate the future path of virus-related outcomes in a country in the context of a new pandemic. The method takes advantage of the fact that the recent COVID-19 virus entered into different countries in slightly different points in time. An application to the cumulative virus-related death toll (as of 13 April 2020) shows that the total numbers may increase by 50% in a week in UK and may double in two weeks in Turkey. The method carries great potential and is open to applications in many different contexts where the outcomes of rare events could be forecasted.

⁷See [Ferman, Pinto, and Possebom \(forthcoming\)](#) for a list of the potential empirical concerns with synthetic control models.

References

- Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association* 105, 493–505.
- Abadie, A., Diamond, A., Hainmueller, J., 2015. Comparative politics and the synthetic control method. *American Journal of Political Science* 59, 495–510.
- Abadie, A., Gardeazabal, J., 2003. The economic costs of conflict: A case study of the Basque Country. *American Economic Review* 93, 113–132.
- Ferman, B., Pinto, C., Possebom, V., forthcoming. Cherry picking with synthetic controls. *Journal of Policy Analysis and Management* .
- Hale, T., Petherick, A., Phillips, T., Webster, S., 2020. Variation in government responses to COVID-19. *Blavatnik School Working Paper*, No. 2020/031 .
- Klößner, S., Pfeifer, G., 2018. Outside the box: Using synthetic control methods as a forecasting technique. *Applied Economics Letters* 25, 615–618.
- Linton, O., 2020. When will the Covid-19 pandemic peak? *Systemic Risk Centre Special Paper*, No. 18 .
- Meyersson, E., 2018. Turkey — matching future economic developments to history. *Goldman Sachs CEEMEA Economics Analyst* .

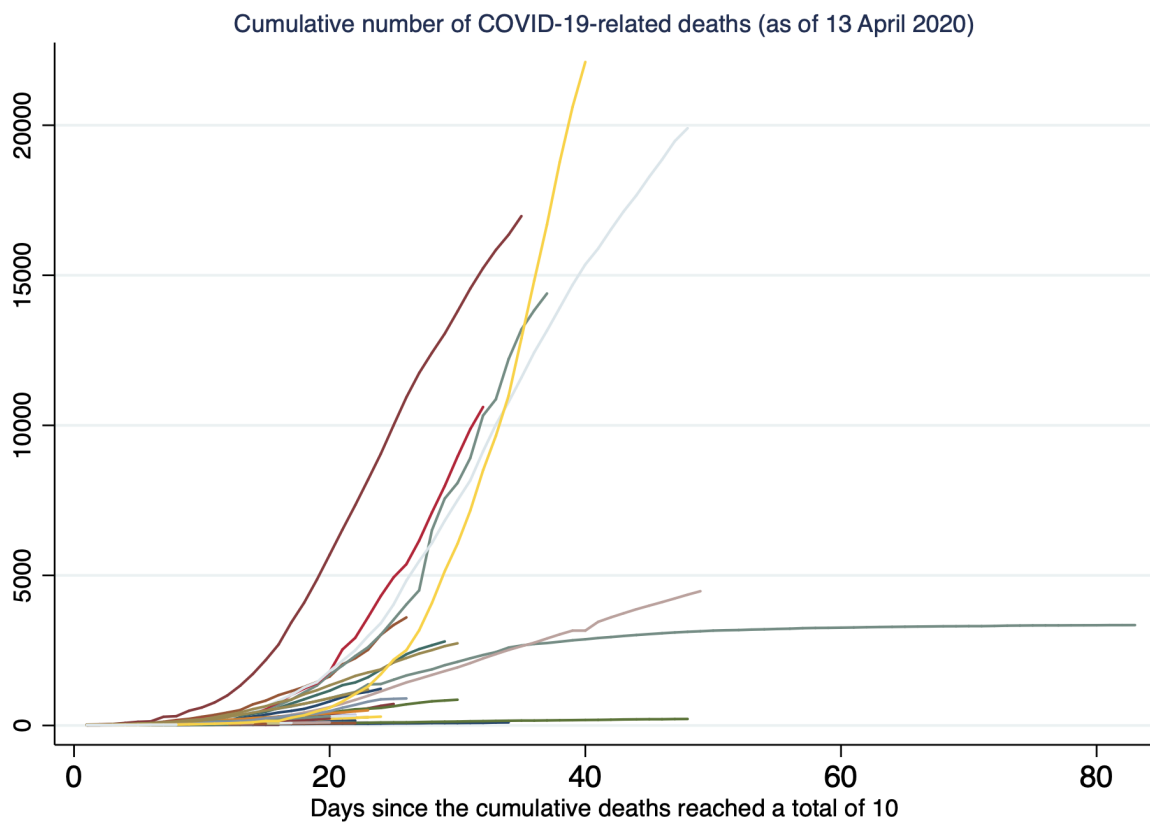


Fig. 1. **Evolution of cumulative COVID-19-related deaths since the first day the total number reached 10 in a country.** Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).

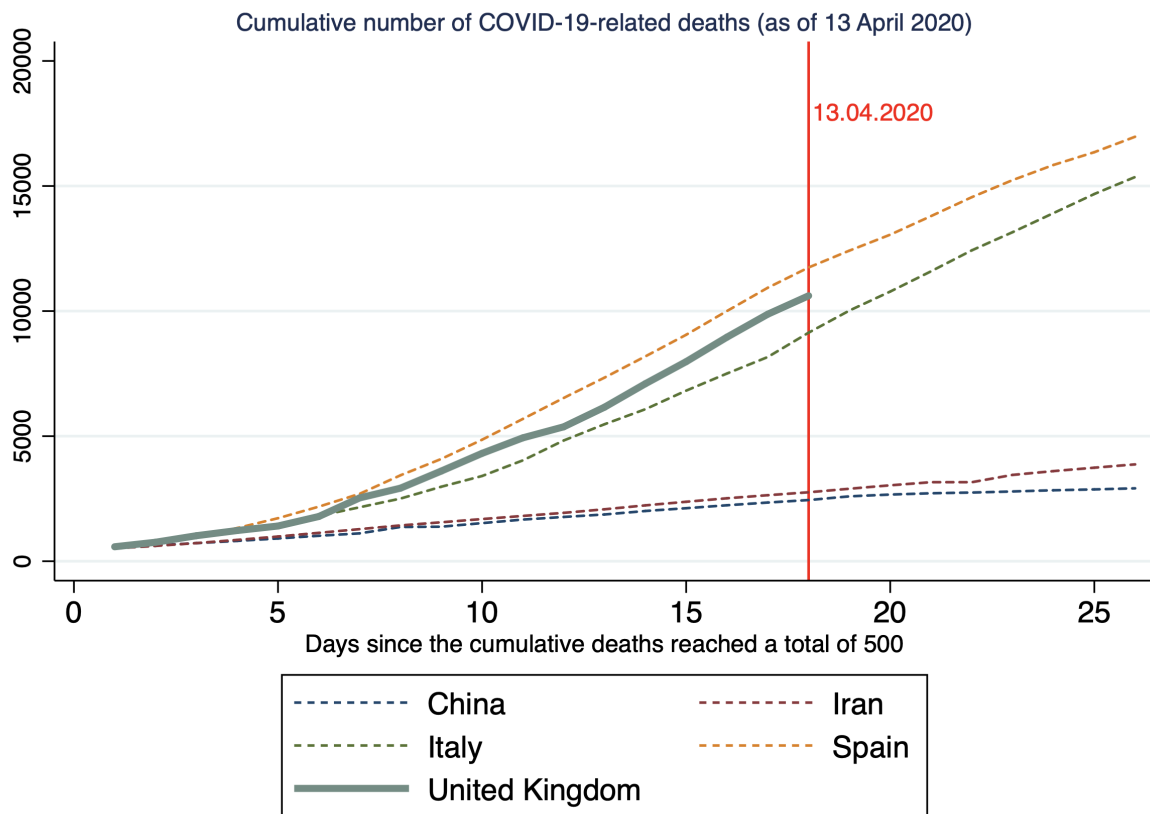


Fig. 2. Evolution of cumulative COVID-19-related deaths since the first day the total number reached 500 in a country. Red line represents the most recent situation in United Kingdom. Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).

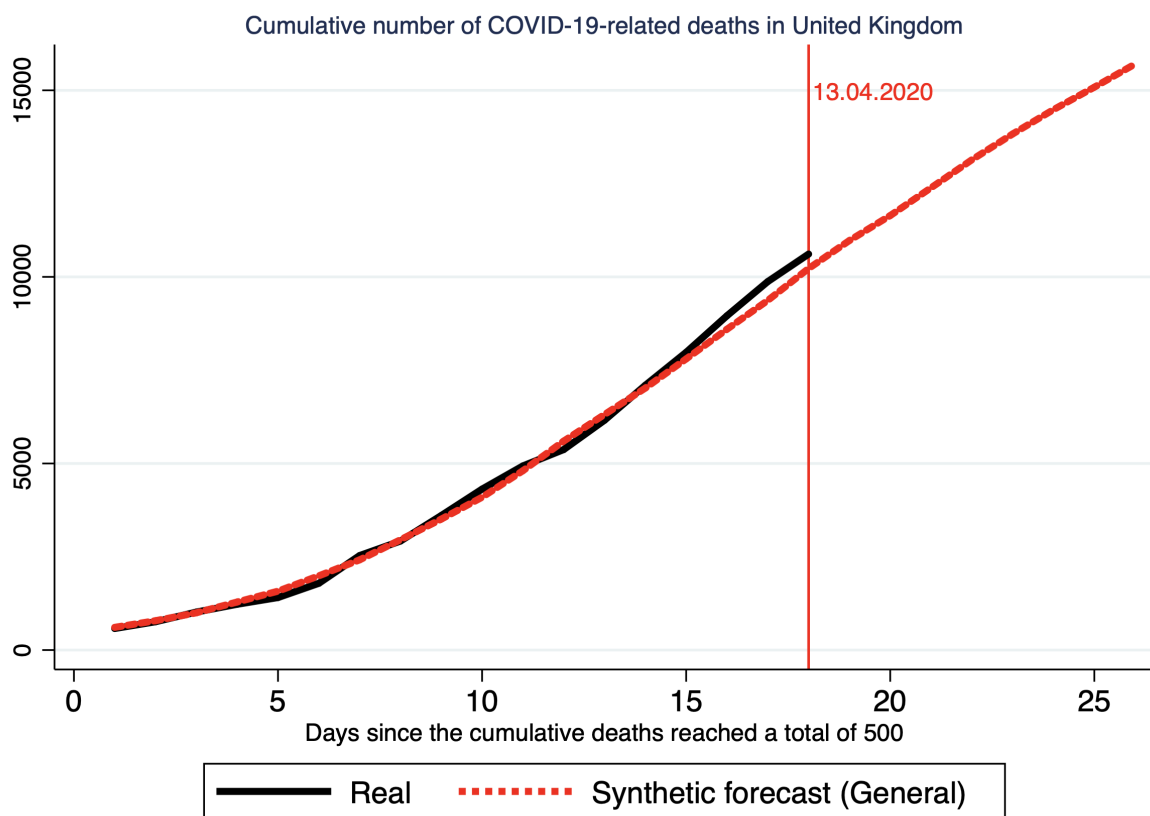


Fig. 3. **Synthetic forecast for cumulative COVID-19-related deaths in United Kingdom for the next 8 days.** Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).

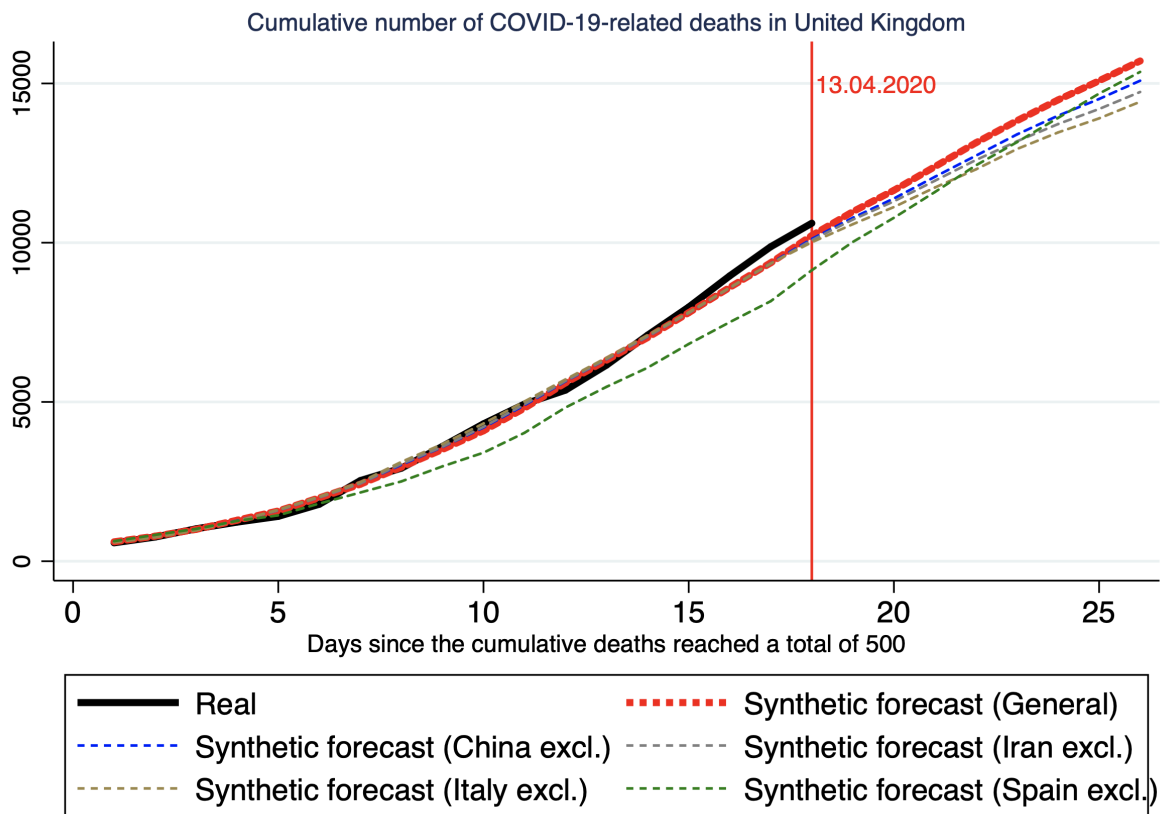


Fig. 4. Synthetic forecast for cumulative COVID-19-related deaths in United Kingdom for the next 8 days: Exclude one leader at a time. Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).

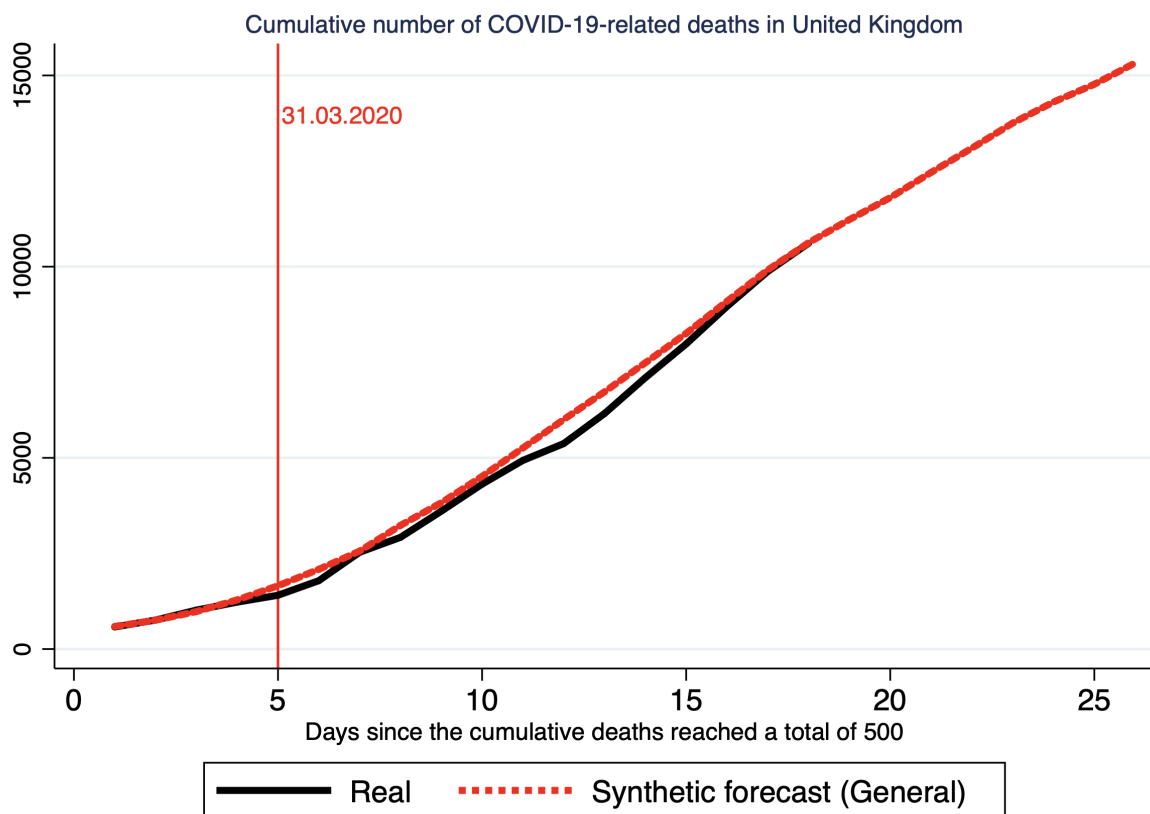


Fig. 5. **Synthetic forecast for cumulative COVID-19-related deaths in United Kingdom for the next 8 days: Estimated only over 5 days in March.** Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).

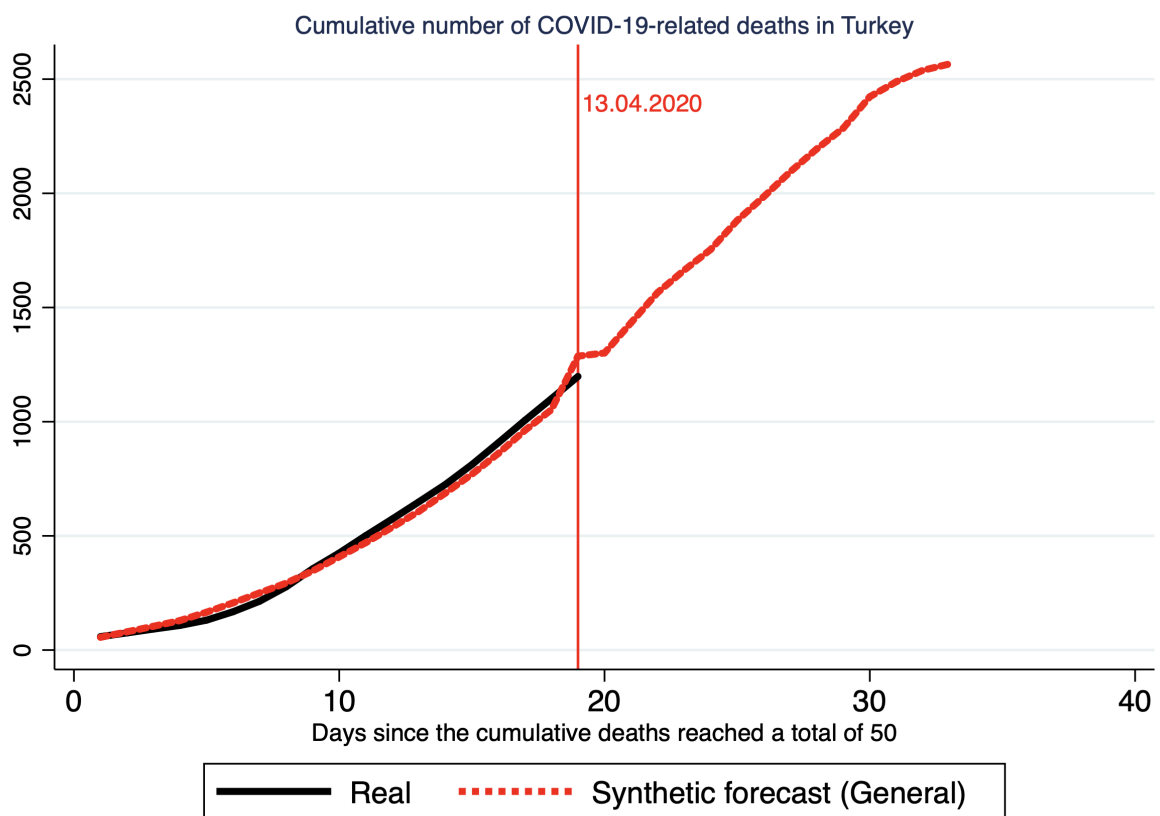


Fig. 6. **Synthetic forecast for cumulative COVID-19-related deaths in Turkey for the next 14 days.** Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).

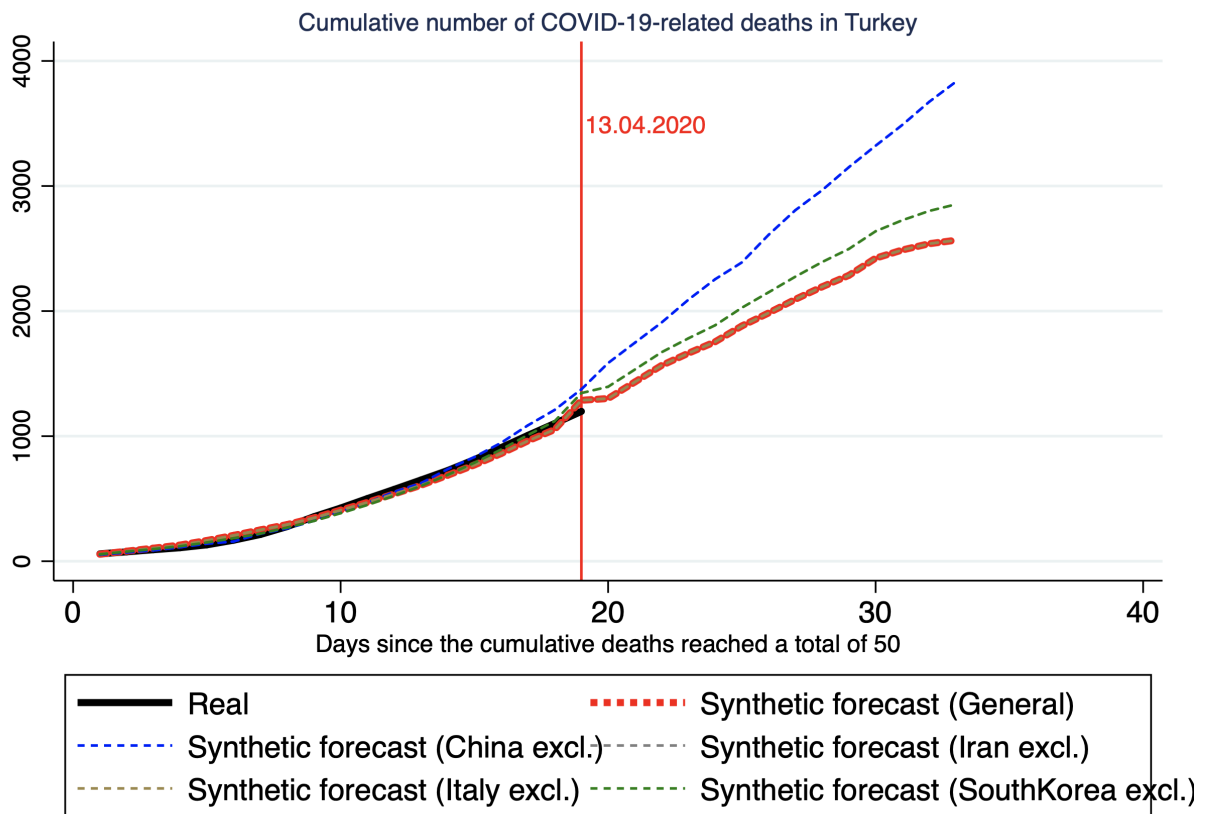


Fig. 7. Synthetic forecast for cumulative COVID-19-related deaths in Turkey for the next 14 days: Exclude one leader at a time. Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).

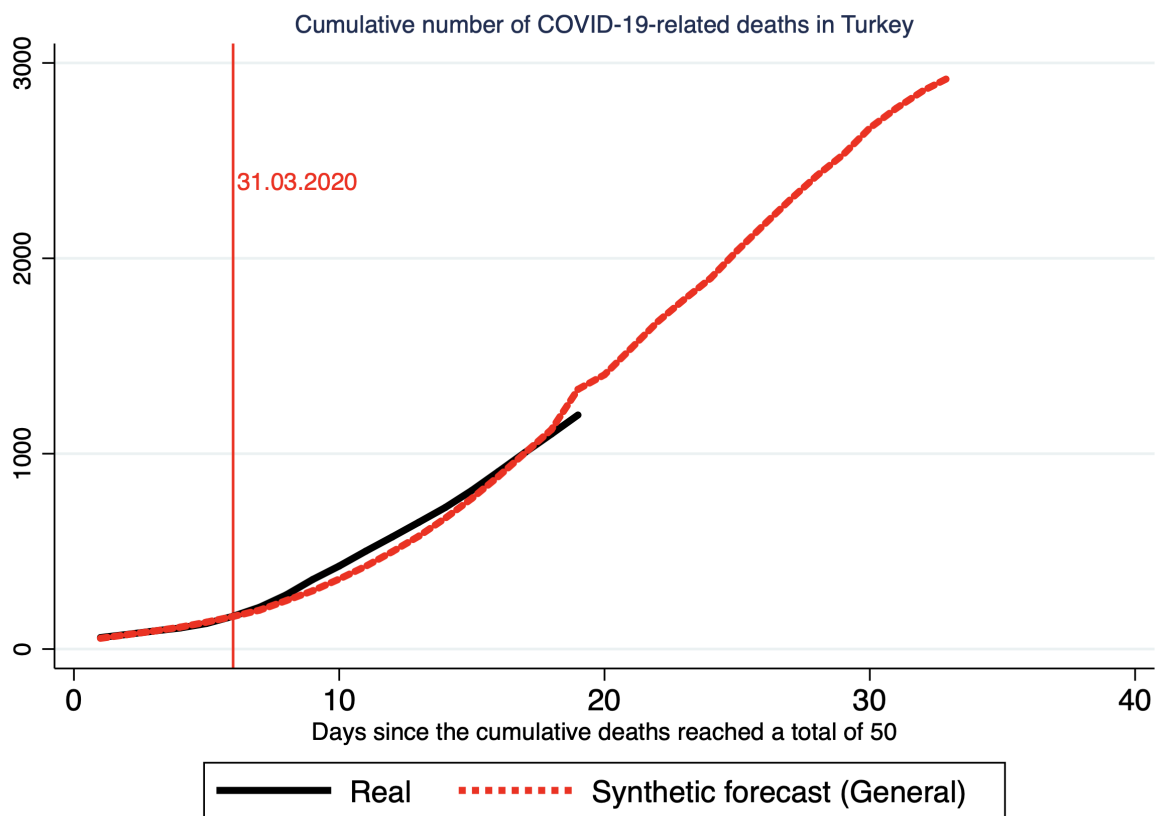


Fig. 8. **Synthetic forecast for cumulative COVID-19-related deaths in Turkey for the next 8 days: Estimated only over 6 days in March.** Source: Author's calculations as of 13 April 2020 with the data provided by the European Centre for Disease Prevention and Control (ECDC).



THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■



The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Tel: +44 (0)20 7405 7686
systemicrisk.ac.uk
src@lse.ac.uk



Systemic Risk Centre