EIEF WORKING PAPER SERIES

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 20/04

March 2020

The Optimal COVID-19 Quarantine and Testing Policies

by

Facundo Piguillem

(EIEF)

Liyan Shi

(EIEF)

Optimal COVID-19 Quarantine and Testing Policies*

Facundo Piguillem[†] Liyan Shi [‡]

April 15, 2020

PRELIMINARY DRAFT

Abstract

Many countries are taking measures stopping productive activities to slow down the spread of COVID-19. At times these measures have been criticized as being excessive and too costly. In this paper we make an attempt to understand the optimal response to an infectious disease. We find that the observed policies are very close to a simple welfare maximization problem of a planner who tries to stop the diffusion of the disease. These extreme measures seem optimal in spite of the high output cost that it may have in the short run, and for various curvatures of the welfare function. The desire for cost smoothing reduces the intensity of the optimal quarantine while extending it for longer, but it still amounts to reducing economic activity by at least 40%. We then study the possibility of either complementing or substituting the quarantine policy with random testing. We find that testing is a very close substitute of quarantine and can substantially reduce the need for indiscriminate quarantines.

JEL classification: E1, E65, H12, I1

Keywords: COVID-19. Optimal quarantine. Optimal testing. Welfare cost of quarantines.

*We thank Claudio Michelacci, Daniele Terlizzese and Luigi Guiso by providing us with deep thoughtful suggestions in the early stages of this draft. All remaining errors are ours.

[†]EIEF and CEPR (e-mail: facundo.piguillem@gmail.com).

[‡]EIEF (e-mail: liyan.shi@eief.it).

1 Introduction

The arrival of COVID-19 at the beginning of 2020 took most of the world by surprise. It was quickly understood that even though related to SARS, its fatality rate was significantly lower. This drove many governments to deem it as a mild illness, resulting in very few initial measures to stop it.[1] However, soon after the outbreak it became clear that COVID-19 was substantially more contagious than SARS. This worried local and national officials because if the virus could spread freely the hospitals would not be able to treat the large inflow of potential patients: the health systems were facing a capacity constraint.

This per se would not be a daunting feature if it weren't because the fatality rate among untreated elderly (those above 60 years of age) was alarmingly high, with some studies estimating above 5% for patients between 60 and 70, and around 15%-20% for individuals above 70.[2] The combination of rapid diffusion and the need of intensive care units (ICU) to prevent a high mortality rate resulted in many administrations taking aggressive measures to either stop the infection or, at the very least, slow down the diffusion, which is known as "flattening the curve."

The approach to deal with the treatment capacity problem has been heterogenous across countries. China took initial drastic measures stopping all economic activity in the most affected areas, while the Japan and South Korea have implemented policies to slow down the diffusion without greatly affecting economic activities.[3] As long as the number of affected individuals do not reach the treatment capacity constraint, the disease should be manageable. Other countries have taken intermediate approaches, but the common language appears to be to "flatten the curve", without necessarily eliminating the threat, mitigation rather than suppression.

Since any intervention that affects GDP is costly, and exponentially so as the intervention deepens, these different approaches raise many questions about the right policy: should countries follow the Chinese approach taking drastic measures until the virus

[1] One example of this is the fact that the Wuhan doctor who discovered the virus was initially disciplined for "spreading rumors" that could create paranoia. On the other hemisphere, President Trump in public appearances argued that it was no more than a seasonal flu.

[2] All available data shows that the fatality rate among individual under 40, without pre-existing conditions, is no different than a seasonal flu.

[3] The strategy of trying to eliminate the virus is also termed Suppression as in Ferguson et al. (2020), while flattening the curve is termed Mitigation.

is extinct? Or is it better to do enough intervention to keep the affected population under the capacity constraint? If so, how much is enough? Wouldn't a combination of the two be better? What should the role of testing be in this context?

In this paper we aim to provide a preliminary answer to the aforementioned questions and try to rationalize the diverse observed policies. To this end we build on Atkeson (2020) who incorporates a SEIR epidemiology model into well known economic setups.[4] In this environment there is an outbreak of an infectious disease which spreads out continuously over time. Some affected individuals are initially asymptomatic and engage in economic activities (meeting) with healthy, but susceptible, subjects who then contract the illness and pass it to others. Unlike Atkeson (2020) we assume that exposed individuals are also asymptomatic carriers who can transmit the virus to other susceptible agents. Once the subject is symptomatic and recognized as infected, it is contained and cannot transmit the illness. However, in this period she may need medical care. If she is not able to receive medical care, she dies with a higher probability than with proper care. We assume that the country has a capacity constraint on how many people can be treated at a given time. Once the capacity is exceeded, the average fatality rate in the economy starts to sharply rise.

In this context whether to choose a suppression or mitigation strategy depends on the possibility of eliminating the virus. Many epidemiologists argue that it is not possible to completely eliminate the virus, and that it would eventually be endemic. For instance, actions like those taken against the SARS in 2002 and Wuhan in 2020 are futile. Thus, governments should only seek mitigation. To give a chance to the suppression strategy, we further assume that there could be a critical mass of individuals, which we assume is 1 person, such that if at some point the number of contagious carriers is below that critical mass, the virus is completely eliminated.

We stress in this paper the two fundamental issues that we think should drive the optimal intervention. The first one, as we mentioned above, is the *capacity of the health system to deal with a large inflow of patients*. By many accounts COVID-19 does not seem to be an extremely deadly illness when the carriers are properly treated. Hence, the need to incorporate a hospital capacity constraint is a first order issue. Second, but no less important, in the standard SIR model, the main (implicit) friction creating

[4]In two contemporaneous paper Eichenbaum, Rebelo, and Trabandtz (2020) and Alvarez, Argente, and Lippi (2020) rely on the classical SIR model that does not distinguish between symptomatic and asymptomatic individuals.

the need for indiscriminate quarantines, is the inability of the policy maker to distinguish the asymptomatic infected (exposed in the terminology of Atkeson (2020)) from the susceptible but still unaffected. If it could, the decision maker *would quarantine only the affected*, letting the unaffected population to continue with their normal activities. Even though this appears as a natural **information friction**, the technology available to test individuals, identify them and be able to impose personalized quarantines rather than indiscriminate ones, exists and could be a welfare improving substitute of what is nowadays termed lockdowns. Of course, testing the whole population at once would completely eliminate the problem, but it could be prohibitively expensive. But this is a cost-benefit analysis that should be properly addressed in the current state of affairs. We analyze these two issues in sequential order. First, we take as given the information friction, then we analyze policies that relax it.

When the policy maker cannot separate an exposed, but asymptomatic, from a susceptible individual, it directly stops some or all of the economic activity to avoid the spread of the illness. By doing so, it prevents the realization of meetings that reproduce the virus. How much and for how long should production be restricted? In the current jargon, how strict and how long should the quarantine be?

We calibrate the model to the preliminary data arising from the outbreak in Italy and we find that the observed set of policies are in line with our estimations. A complete lockdown for three weeks appears optimal when the planner is not concerned about consumption smoothing and the measure is implemented at the same time as the Italian government did. This policy **aims to achieve suppression**: after three weeks the virus is completely eliminated and there are no further waves.[5] A positive critical mass of 1 person makes sure that the virus is eliminated once and for all. Instead, if the critical mass is 0, i.e, the virus cannot be eliminated, the optimal policy with linear utility resembles that of the planner with concave preferences seeking mitigation.

When the planner has concave preferences that favor cost smoothing, the intensity of the optimal policy is substantially reduced and closely follows the pattern of the number of infected individuals. Placing the results in the Italian context, the intervention starts mildly at the beginning of March. Then sharply increases as the number of infected cases accumulates, to reach around 50% of GDP at the beginning of April.

[5]The first documented case in Italy was on February 24^{th}. Several epidemiologist have inferred that the virus was circulating in the country by the third week of January. The Italian government implemented the quarantine in March 8^{th}, which is about 48 days after the presumed arrival of COVID-19 to Italy.

Then it starts to reduce the intensity again, by the beginning of May is around 15% of GDP and remains relatively stable, but reducing, until July. All in all, the intervention lasts for 90 days. It is worth noting that the measures introduced by the Italian government in March 8^{th} affected around 16% of the economic activity, the tightening of the intervention in March 22^{th} increased the intensity to around 50%, and in April 10^{th} the Italian Prime Minister announced a relaxation bringing it down to around 20% of GDP. Unlike the linear utility case, the optimal policy with **concave preference aims to achieve mitigation**, or "flattening the curve". Thus, it builds a large mass of immune individuals, with the implied cost of a substantial number of fatalities.

When comparing both policies with the observed quarantines, we find that the current quarantine is too "soft" to achieve suppression and too "harsh" to be an optimal mitigation strategy. We also find that if the "value of life" is sufficiently large, above 8 years of annual income, the concave planner should also pursue suppression, implementing a policy that shuts down $2/3$ of economic activity for two months.

We estimate that without any intervention there could be as many as $660,000$ fatalities.[6] With a suppression policy, either long (concave) or short (linear), there are between $2.000 - 2.700$ fatalities. With a mitigation policy the number of fatalities is substantially larger, with at least 250.000, much larger than the 27.700 that we obtain if the current intervention were to continue for at least three months.

We want to emphasize that the decision between mitigation and suppression is by no means trivial. The welfare functions that we compute are not concave. Thus, seemingly small changes in the parameters can have very important consequences for the optimal policy. The linear utility planner has a tendency to choose suppression, except when the critical mass is zero. The concave planner has a tendency for mitigation, except when the value of life is large enough. If the value of life where above a reasonable 8 years of annual income, the concave planner would also choose suppression. In other words, depending on how policy makers value fatalities, they could choose extreme policies, even when they are concerned about cost smoothing.

Yet, these policies are drastic with large costs in terms of output, which falls by more than 50% at the peak of the intervention. This brings about the possibility of complementing the quarantines with massive testing to simultaneously decrease the speed

[6] This result is line with the calculations by the panel of experts in Walker et al. (2020), who estimate around $645,000$ fatalities for Italy without any intervention

COVID-19's reproduction and be able to put to work a larger share of the population. To do so we take seriously that the main problem is an information friction. We consider the possibility that the government can initiate intense screening to identify the exposed individuals. Once identified as positive, the subject is required to endure a (personal) quarantine. This is done by randomly selecting individuals for which there is no information yet: those who have never tested positive before.[7] Identifying a positive case has two beneficial effects. 1) It is possible to quarantine the individual, even in a stricter way than the rest of the population to slow down the reproduction of the virus. And, 2) once the individual is able to eliminate the virus from its biological system, she is immune and can be allowed to work without any restriction, helping to moderate the extent of the recession. This last contribution, is often overlooked and it could be of considerable relevance, see for instance Dewatripont et al. (2020), especially when many subjects could remain asymptomatic during the whole duration of the infection. Without the testing, we would quarantine many individuals that are immune and could be working.

We lack reliable data on the cost of a test. Thus, we assume that the marginal cost of the first unit is 1 day of daily output per worker and grows quadratically. The speed at which the marginal cost grows is chosen in such a way that it would be economically infeasible to test the entire population at once. We find that **testing is intensively used as a substitute of indiscriminate quarantines** and generates substantial welfare gains. With the cost function that we assume, the output gains are so large that lockdowns could be completely avoided. In our favorite scenario, testing is used intensively, an average of 2% of the unidentified population is tested every day, with a final cost of 1% of GDP.

1.1 Literature review

The literature on epidemiology control dates back to the model proposed by Kermack and McKendrick (1927), also known as the SIR epidemiology model. However, to the best of our knowledge there has not been many applications to economics. With the recent outbreak of COVID-19, economic researchers have started to incorporate SIR

[7]If a subject has been tested before but the results were always negative, it is still susceptible to the illness, and therefore is in the same situation as another who has never been tested. While those who have been affected and tested positive and recover, are from then on immune, so that there is no need to test them again.

models in economic environments to assess the potential economic implications of COVID-19.

Atkeson (2020) computes the projected paths of the disease and evaluates its economic impact. We build on his work deepening the information structure. In this work, it is implicitly assumed that only the symptomatically infected individuals are contagious. We instead assume that, as it happens in many countries, the symptomatically infected are isolated and therefore do not contribute to the speed of contagion. It is what we call exposed but asymptomatic individuals, potentially unidentified without testing, who actually fuel the spreading of the disease. This extension allows us 1) to better fit the dynamics of the disease and 2) to have a well-defined information friction calling for the need for testing. In addition, we optimize over the set of policies rather than focusing on some, yet interesting, paths.

Eichenbaum, Rebelo, and Trabandtz (2020) construct what they call a SIR-macro model with endogenous consumption and labor supply. The competitive equilibrium in their model is suboptimal due to fact that agents do not fully internalize the externality of their economic interactions. They consider the optimal consumption tax policy that can correct the externality. We differ in many dimensions. First, as Atkeson (2020) they consider only the actively infected as potential carriers. Second, they use a meeting technology that does not allow for congestion. Thus, the dynamics of how infection spreads doesn't feature our dilutive effect that kicks in later as immuity expands. Third, we consider policies that directly control economic activities, rather than altering marginal decisions.

As us, Alvarez, Argente, and Lippi (2020) also study the optimal lockdown policy. To this end they use a meeting technology similar to Eichenbaum, Rebelo, and Trabandtz (2020). They assume linear preferences, which weights output and the cost of disease. They also explicitly consider the possibility that in some countries the lockdown could be less effective or harder to implement. We differ from them in some dimensions. Our structure and meeting technology allow us to focus on the fundamental information friction in distinguishing types, which necessitates quarantines to stops the contagion or testing to overcome it. We consider also concave preferences, which generate a need to smooth the costs of the intervention, and show that is very relevant reducing the intensity and increasing the length. Similarly, to incorporate the hospital capacity problem, they assume an exogenous fatality rate linearly increasing in the number of infected. Instead, we model hospital capacity explicitly and calibrate it to the Italian situation. Finally, they solve the optimal control problem, without restricting the policy space, while we look for the optimal lockdown intensity in a restricted policy space.

Dewatripont et al. (2020) propose that testing, either prioritized or random, is essential to restart the economy. They argue that mass testing is technological feasible and a mere logistic issue of scaling up. We assess a random testing policy and find that some degree of testing joint with quarantines are welfare improving.

2 A SEIR model of disease contagion

Time is continuous and runs indefinitely, $t \in [0, \infty)$. At time t the economy is inhabited by a population N_t with an initial mass of one: $N_0 = 1$. Since the spread of the illness is so fast that it can be measured by the day, we use the convention that one unit of time is one day. At any given time, each individual can be one of five types: susceptible, exposed, infected, and recovered. We denote by S the number of individuals still unaffected but susceptible to the virus. There are two types of carriers of the virus: exposed asymptomatic E and infected symptomatic I. They both infected and thus infectious. When an individual first becomes infected, it always starts in the group E, it may develop symptoms and become I, or it may never show any symptoms, in which case remains in E until it recovers.

When a subject recovers it becomes immune. Depending of the symptomatic history, R will denote the number of immune recovered agents that were previously symptomatically infected I, and R^u the also immune recovered subjects, but who where previously only symptomatic E. The first groups have observable signals that make them identifiable by the government, while R^u are immune individuals that without additional information could remain unidentified. For this reason the distinction between R and R^u is important. Clearly, it must be the case that

$$N_t = S_t + E_t + I_t + R_t + R_t^u.$$

At $t = 0$, the economy is hit by a disease due to a deadly virus. If the exposed population is above a critical mass, $E > \underline{E} \geq 0$, then it starts to spread. Otherwise, it's self-contained and all patients gradually recover. The virus spreads through meetings between exposed and unaffected individuals. To avoid this, the government can impose quarantines, which can be directed or indiscriminate. Directed quarantines single out a proportion of the carriers and force them to stay at home in isolation. Thus, a proportion q_t^I of infected individuals and a proportion q_t^E of exposed asymptomatic are forced into quarantines. In addition, the government can force a proportion q_t of the whole population to refrain from engaging in economic activities. This q_t is what we call indiscriminate quarantine. We assume that symptomatic individuals, I_t infected agents, can be identified by their symptoms and are forced into a full quarantine, so that $q_t^I = 1$.[8] In this section we start by assuming that the government has no information about the identity of either E_t or R_t^u individuals, hence, it has no choice but setting $q_t^E = q_t^{R^u} = 0$.[9] Therefore, the virus spreads **only** through meetings between exposed asymptomatic and unaffected individuals. Finally, those who are initially exposed after an incubation period become symptomatically infected. Those who already have it and are recovered become immune permanently.

The number of meetings in the economy depends on the level of economic activity, which in turn depends on the number of workers, L_t, who are not indiscriminately quarantined: $1 - q_t$. We denote by $\lambda m(L_t, E_t)$ the meeting function between workers and carriers of the virus. Potentially, the total number of carriers allowed to work is $(1-q_t^I)I_t + (1-q_t^E)E_t$. Because we are assuming that $q_t^I = 1$ and $q_t^E = 0$, the total number of carriers is just E_t, hence the second term in the meeting function rather than $I_t + E_t$.[10] Not all of these meetings generate an infection, since only $\frac{S_t}{L_t}$ of the workers are susceptible and the government only allows a proportion q_t of the individuals to work, at every instant there are $\lambda \frac{S_t}{L_t} m(L_t, E_t)(1 - q_t)$ meetings generating new affected (exposed) individuals. Once exposed, an individual becomes symptomatically infected with intensity γ per unit of time, and can recover with intensity σ without

[8] This assumption is in line with the preventive measures taken by all the governments as soon as they detect an infected subject.

[9] Alvarez, Argente, and Lippi (2020) assume that $q_t^I = 0$, as they merge all the carriers E_t and I_t into a single group I_t that is allowed to work as long as they are not subject to an indiscriminate quarantine.

[10] In Atkeson (2020) only the infected individuals can transmit the virus, so that the infectious meetings in his economy are $m(L_t, I_t)$. He does not distinguish between symptomatic and asymptomatic carriers though. We borrow his notation and give it a different interpretation.

ever been symptomatic.[11] Thus, the law of motion of the exposed type satisfies:

$$dE_t = \begin{cases} \left[\lambda \frac{S_t}{L_t} m(L_t, E_t)(1-q_t) - (\sigma+\gamma)E_t\right] dt, & \text{if } E_t \geq \underline{E} \\ -(\sigma+\gamma)E_t dt, & \text{if } E_t < \underline{E}. \end{cases} \tag{1}$$

We assume that the symptom appearance γ and the recovery σ are independent of time and the state of the economy. They just reflect the individual's strength to fight the virus inside their biological system. The same is true for the intensity of contagion λ, which is a scale parameter capturing the level of interactions among agents in their daily economic activity. The speed at which the illness spreads is clearly state dependent, increasing in the number of exposed E_t and the share of the population which are still susceptible. The function $m(L_t, E_t)$ could incorporate potential "congestion" effects. For instance, one may think that when most of the population are already affected, most meetings would be between individuals who are either immune or already infected and thus would not generate new infections. We later propose a functional form for $m(\cdot)$ but we experiment with different alternatives.

We want to emphasize that the presence of the minimum critical mass $\underline{E}$ could be very important for the prescribed policy interventions. When $\underline{E} = 0$ the virus never dies, it could be forced to affect a negligible number of people, but it would be always around to re-surface and spread again. Instead, when $\underline{E} > 0$ it could be possible to take drastic measures to force the affected population below the critical mass, so that the virus disappears and the infection is definitively defeated. Instead, when $\underline{E} = 0$, since the virus would eventually spread anyway, a policy maker could choose to simply regulate the speed at which the number of exposed and infected subjects arrive. This would be important when we bound the capacity of the health system to treat the illness.[12]

Exposed subjects become symptomatically infected at rate γ. Once they are infected they would require medical assistance and potential hospitalization. When treated

[11] The asymptomatic status prior to becoming symptomatically infected clarifies the effects in production, and it is instrumental in Section 5 when we analyze the information friction. Otherwise, we could merge them into a single type as in Alvarez, Argente, and Lippi (2020).

[12] Another way to think about $\underline{E}$ is as a way to prevent the modeling strategy from forcing policy prescriptions. For instance, because growth is proportional, one we can divide a positive number indefinitely by other positive number and it would always be strictly positive. In the context of our model we could end up with less than a person infected, which is not physically possible, but it would imply that the infection would reappear in the future. $\underline{E} > 0$ makes sure that, whenever fewer than a minimum amount of person are infected, the disease would disappear.

individuals recover at rate η and die at the rate Δ_t per unit of time. The law of motion of (symptomatic) infected individuals satisfies:

$$dI_t = [\gamma E_t - (\eta + \Delta_t) I_t] dt.$$

As with γ and σ, here again η is independent of the economy's state. The process by which the body is able to eliminate the virus from the system is not affected by the health system, it only depends on the strength of the subject's immune system, conditional on surviving. But notice that Δ_t does depend on the state of economy. One may think that the way in which the illness affect a particular individual depends only on her/his biological characteristics and therefore should be independent of how other individuals are affected. However, here we assume that the death rate depends on the capacity of the health system to treat patients.

Hospitals can optimally treat only H patients at a time. Once that capacity is exceeded the treatment received by each patient is diluted resulting in a suboptimal treatment. Those who are optimally treated die with intensity θ, while does who are treated in an overcrowded system die with intensity $\delta > \theta$. As a result, the average daily death intensity in the economy satisfies:

$$\Delta_t = \theta \underbrace{\min\left\{1, \frac{H}{I_t}\right\}}_{\text{fraction treated}} + \delta \underbrace{\max\left\{1 - \frac{H}{I_t}, 0\right\}}_{\text{fraction untreated}}. \tag{2}$$

Given the previous assumptions, the number of recovered patients and total population evolve according to:

$$dR_t^u = \sigma E_t dt, \tag{3}$$

$$dR_t = \eta I_t dt, \tag{4}$$

$$dN_t = -\Delta_t I_t dt. \tag{5}$$

From the previous structure it is straightforward, see Appendix A, to compute the average death rate from the illness and the duration of sickness. This of course would depend on whether the patients are treated or not. When all sick individuals are treated, a patient recovers in $\frac{\eta}{(\eta+\theta)^2}$ days, and on average a fraction $\frac{\theta}{(\eta+\theta)}$ of the patients die. When left untreated, the recovery happens in $\frac{\eta}{(\eta+\delta)^2}$ days, and the average death rate is $\frac{\delta}{(\eta+\delta)}$. These are moments that it is possible to match with the already available data. Note that due to selection, patients would appear to recover faster in countries with higher fatality rates.

For simplicity we assume that the production technology is linear. Each meeting produces one unit of output per individual involved in the meeting. Infected hospitalised individuals are unable to produce, since $q_t^I = 1$. In normal times the total production would be $Y_t = L_t = N_t$ per day. However, during the spreading of the virus, only the unaffected, fully recovered and those still undetected but yet exposed can produce, so that $L_t = S_t + E_t + R_t + R_t^u$. Hence, if the government allows for undistorted economic activity, i.e. $q_t = 0$, the total production would be $Y_t = S_t + E_t + R_t + R_t^u$. To prevent the spread of the virus the government bans certain activities. It does so by forcing quarantines among the population. Since the government is unable to distinguish S_t and R_t^u from E_t, it cannot condition the quarantine on each individual status, it simply ban a fraction $q_t \in [0, 1]$ of all economic activity. As result, the total production after a policy intervention is $Y_t = (1 - q_t)(S_t + E_t + R_t^u) + R_t$.[13]

We assume a closed economy. The only produced good is non-storable, and there is no possibility of borrowing or saving in financial assets. This implies that consumption is equal to production in every period: $C_t = Y_t$. All individuals, and therefore also society as whole, discount the future at rate $\rho > 0$. The government chooses a path $\{q_t\}_{t=0}^{\infty}$ to maximize society's welfare:

$$\max_{\{q_t : t \geq 0\}} \int_0^{\infty} e^{-\rho t} u\left((1 - q_t)(S_t + E_t + R_t^u) + R_t\right) dt; \qquad \text{(P1)}$$

subject to equations (1), (4), and (5).

Notice that this setup allows for a variety of possibilities. A solution could be a forced quarantine for every individual for a limited period. For instance, we can think about Wuhan's **suppression** policy intervention as setting $q_t = 1$ for all $t \leq \bar{\tau}$ and $q_t = 0$ for all $t > \bar{\tau}$, for some $\bar{\tau} > 0$. In this case if in some point $E_t < \underline{E}$ the virus dies and never recovers. This problem would reduce to choosing the optimal length $\bar{\tau}$ of a complete lock down. Alternatively, one can think about policies that are more

[13] Notice that we are allowing the recovered subjects to return to work. This is clearly optimal and the recover status is fully observable. In spite of this, most, if not all, countries include inefficiently the recovered in the mandatory quarantines.

moderated, **mitigation** policies, with $q_t < 1$, but that last for a longer interval. Here the strategy would be to try to maintain I_t below H at all t until most of the population becomes immune.

The choice of the welfare function is by no means trivial. In Problem (P1) we have purposely excluded the welfare lost due to fatalities. We have done so because it is, at the very least, highly controversial how to compare losses due to foregone consumption with welfare losses due to fatalities. What is the value of a life? If one believes that a human life is more important than everything else, then the correct welfare function should only minimize the number of fatalities. In this case, as long as $\theta > 0$ the solution to (P1) is almost trivial, setting $q_t = 1$ for as long as is needed to locate E_t below $\underline{E}$. If $\theta = 0$ and $\delta > 0$, then only policy paths that maintain $I_t \leq H$ would be part of the solution. Here the output cost becomes the relevant factor pinning down the optimal path.

However, (P1) is also problematic because current and past choices reveal that societies are willing to trade off human life for economic activity. For instance, the U.S. Center for Disease Control and Prevention (CDC) estimates that between 12.000 and 61.000 people die annually due to influenza. Yet, governments are not willing to stop the economic activity to prevent it. Similarly, in 2018 around 36.000 people died in car accidents in the U.S. But there has never been a discussion about banning circulation in motor vehicles. One can interpret these choices as balancing individual and collective responsibility. As long as the fatalities are not too large, society prefers to delegate the choice of the "acceptable risk" to the individual, while if the fatality rate is too high there maybe some frictions that prevent individuals from properly asses the risk. Then, it becomes a collective responsibility and the government must intervene. For this reason we also consider and alternative problem where the policy maker trades off economic activity and lives:

$$\max_{\{q_t : t \geq 0\}} \int_0^\infty e^{-\rho t} \left[u\left((1-q_t)(S_t + E_t + R_t^u) + R_t\right) - v\left(\Delta_t I_t\right)\right] dt; \qquad \text{(P2)}$$

$$\text{subject to equations} \quad (1), (4), \text{and } (5).$$

Here the function $v(x)$ would be key in determining the number of acceptable deaths. Admittedly, it is hard to parameterize it, but we use data for alternative activities that generate fatalities to discipline its implications. In the sense that, if for instance, an activity is allowed when the fatalities are caused by influenza, it should also be allowed when caused by COVID-19.

2.1 Functional forms

We consider alternative functional forms. We start assuming that welfare is linear in consumption, so that the planner is only concern about productive efficiency. In this case:

$$u(c) = c$$

Of course, variations of consumption across time could also be important for the planner. Thus, we also present welfare results in which the elasticity of intertemporal substitution is not zero. In this case we use two alternatives:

$$u(c) = \log(c) \quad \text{and} \quad u(c) = c - \frac{b}{2}c^2$$

All the functional forms are mathematically tractable and meaningful in one dimension or another, allowing for a wide range of interpretations.

The choice of $v(x)$ is less straightforward. One maybe think that a quadratic loss function $v(x) = \frac{d}{2}x^2$ would be appropriated, because the cost grows exponentially with the number of fatalities. However, it also has the potentially unappealing feature that the planner would be wiling to accept many fatalities if they are sufficiently spread out over time, while it wouldn't accept it if all fatalities happen in a concentrated interval of time. For instance, 1 dead today and 1 tomorrow is much better that 2 dead either today or tomorrow. An alternative is to use a linear function $v(x) = dx$, and choose d in such a way that there is an "optimal" upper bound for the number of fatalities. Alternatively, we could consider the cumulative number of fatalities and define

$$\int_0^{\infty} e^{-\rho t} v\left(\Delta_t I_t\right) dt = \frac{d}{2}\left[\int_0^{\infty} \Delta_t I_t dt\right]^2.$$

Regarding the meeting function we do most of our calculations using a standard proportion function:

$$m(L_t, E_t) = E_t.$$

Finally, we present robustness using a Cobb-Douglas meeting function which allows for more flexibility at targeting congestion on diffusion technologies.

3 Quantitative implications

3.1 Parametrization

There are several key parameters in the model. First, the parameters γ and σ determine how long an individual can be contagious without potentially showing any clear symptoms. The parameter γ is related to the incubation period, which is 6.5 days of generation time according to Ferguson et al. (2020), so that $1/\gamma = 6$. For η, which is the recovery rate of symptomatic subjects we also follow Ferguson et al. (2020) and we set it such that on average it takes 9 days for a subject to recover. The precise value for it would depend on the death rate which we discuses below. Nevertheless, for any average fatality rate d, the daily fatality rate θ and the average time to recovery η are jointly determined by the relationship show in the previous section. The relationship is:

$$\frac{\eta}{(\eta+\theta)^2} = Days; \quad \frac{\theta}{(\eta+\theta)} = d.$$

The solution to this system is:

$$\eta = \frac{(1-d)^2}{Days}; \quad \theta = \frac{d}{1-d}\eta,$$

where $Days$ is the average number of days until recovery and d is the average death rate. Thus, given any estimate for θ we can recover the implicity η that generate 9 days of recovery time using the last equation.

There are 4 parameters for which there is considerable uncertainty: θ, σ, λ and the date of the first infection. The first is related the expected number of deaths for a given number of infected agents. The second, determines for how long an asymptomatic agent can be contagious and therefore is important for determining the dynamics. The third, has been the subject of considerable debate since it determines the reproduction factor of the virus R_0. One may think that σ is a fundamental property of the virus, in the sense that it should be the same across countries. However, both θ and λ can be very specific to each country. For instance, the age structure of the population can affect the average death rates, while the nature of the social interactions would determine λ.

To address this issue we estimate these four parameters to match the observed dynamics of COVID-19 in Italy. Ideally one should fit the dynamics of infected cases and compute the implied death rate with the number of fatalities. The problem with this approach is that the information for "cases" only reflects the number of individuals who have been tested and generated a positive result. There are many reasons to believe that this measure would not reflect the real state of the country in terms of infected individuals. First, those who are asymptomatic are never tested and therefore are not recorded. Second, it is widely known that initially there was a scarcity of test kits, which forced the authorities to tested only subjects that were likely to be infected or vulnerable. Thus, many mildly symptomatic individuals were left untested. Finally, the tests weren't very precise at the beginning which delivered many false negative.

To avoid this problem we target instead the path of fatalities. Given the parameters of the model, there is a one to one mapping from the number of infected to the number of fatalities. To be precise we target four moments to pin down the four parameters. The number of fatalities at the beginning February 24^{th}, at the introduction of the first intervention on March 8^{th}, the tightening of the restrictions on March 22^{th} and the status two weeks later on April 10^{th}. Loosely speaking the first two moments are mainly determined by the initial time of the outbreak and λ, which controls the number of meetings generating infections. The outcomes after the consecutive interventions shed light on the number of asymptomatic agents, determined by σ and the fatality rate θ.[14]

To implemented this strategy we need three additional pieces of information. First, for a given number of initially recorded fatalities there are many combinations of initial mass and outbreak date that are consistent with the observation. To avoid this ambiguity we assume that $e_0 = 1/60.000.000$. That is, there was initially only one exposed individual (patient 0). Second, the effects after the interventions depend on

[14]Here is important that most of the symptomatic agents are in individual quarantines. If there were no asymptomatic $\sigma = \infty$ the infection would died out rapidly, while with many asymptomatic $\sigma = 0$, the infection would keep spreading quickly. Similarly, the changes in the number of fatalities would be larger the larger is θ.

their intensities. On this we rely on information provided by Guiso and Terlizzese (2020) who estimate that the initial intervention on March 8^{th} affected 16% of the sectors and the second one on March 22^{th} reached 40% of sectors. We adjust these values upwards because these estimations do not consider the effect of school closing. As stated by Basile and coauthors, the fact that workers must remain home taking care of their children had an important impact on GDP. Thus, we assume that the initial intervention was $q = 1/4$ and the second $q = 1/2$.

Finally, we assume that hospital capacity was binding at the moment of the initial intervention, and that when the capacity is binding $\delta = 2 \times \theta$. In the next subsection we describe in detail why we make this assumption. This implies that initially the number of fatalities were larger because some patients were untreated. In our model we do not distinguish between patients who require critical care v.s. those who don't. For this reason we need to scale the observed capacity to the number of infected. At the time of the outbreak in Italy there were $5,343$ beds for intensive care and a population of $60,000,000$ people.[15] Since only $0.3 \times 4.4\% = 1.32\%$ of the infected need critical care, the country can treat no more than $5,343/0.0132$ infected individuals at a time. Thus, the country is prepared to treat only $100 \times (5,343/0.0132)/60,000,000 = 0.67\%$ of the population. For this reason, in the baseline scenario we assume that h is constant and equal to 0.00674. Of course, governments are taking measures to increase this capacity. We will analyze this issue later. For instance the increased capacity in Italy after the outbreak would implied that the new h is $0.0106 \approx 1.1\%$.[16]

The resulting parameter values are shown in Table 1. We obtain $\lambda = 0.6515$, $\theta = 0.212$, $\eta = 0.1070$, $\sigma = 0.22$ and the day of the outbreak is January 21^{st}. A few comments are worth mentioning. First, the calibrated initial reproduction factor is $R_0 = 1.7$, which is relatively low compared with other estimates that situate it between 2 and 2.5. Thus, the fact that σ is more than twice of η plays an important role. First, it implies that there is roughly one asymptomatic individual for each symptomatic. Second, but because they recover very fast, the quarantines become very effective. It is precisely the large drop in cases after the successive interventions what identifies the value of σ. Last but not least, the estimated daily death rate implies an average fatality rate

[15] Source: President Conte's national speech on March 24th, 2020. He also mentioned that due to the outbreak the numbers of beds increased to 8,370.

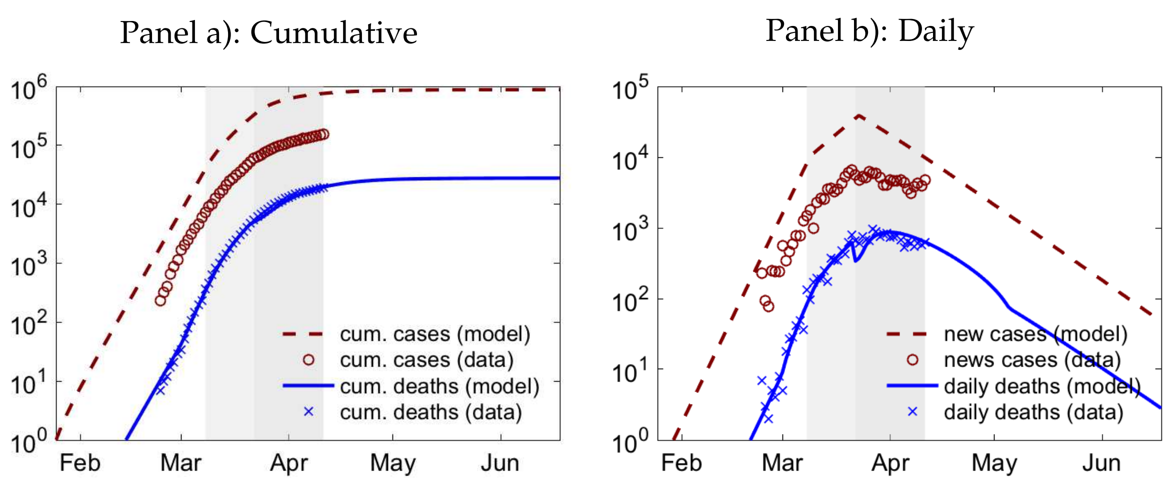
[16] Note that this is a very conservative estimation, the age distribution in Italy also increases the number patients in need of critical care. However, we maintain the estimations from China to account for emergency increases in the number of beds.

Table 1: Parameter values

Parameter	Symbol	Value	Moment
Contagion rate	λ	65.15%	Fit fatality's path
Exposed to infected rate	γ	16.67%	6 days incubation period
Recovery rate	η	10.70%	9 days to recovery for I
Recovery rate	σ	22.00%	4.5 days to recovery for E
Daily death rate if treated	θ	0.21%	1.9% fatality rate if treated
Daily death rate if untreated	δ	0.42%	3.8% fatality rate if untreated
Hospital capacity	h	0.00674	5,343 ICUs for 60 million population
Initial exposed	E_0	1/60mn	One individual in population
Critical mass	$\underline{E}$	E_0	Minimum possible number
Daily discount rate	ρ	0.05/365	interest rate

of 1.9% of the infected individuals. This rate may seem high compare with many studies which situate it around 1%. However, in Subsection 3.1.1 we do model-free estimations that deliver 1.92%. Most of the difference with other estimates is due to the age structure of the Italian populations.

Figure 1: **Calibrated paths**



The implied path of fatalities by the model and the realized one in Italy can be seen in Figure 1. In Panel a) we the show cumulative numbers, while in Panel b) we present the daily numbers. Recall that the calibration only target cumulative numbers. Thus, all daily changes are not targeted. In both panels the observed path of fatalities is depicted with the blue "+" mark. Each mark corresponding to one observation. The red dashed line corresponds to the model prediction that assumes $q_t = 0$ until the government intervenes on March 8^{th}, where the intensity increased to $q = 0.25$ until March 22^{nd}, when the intensity further increases to $q_t = 0.5$ and remains that way until the end of the simulation. **In both panels the scale is logarithmic.** As expected, since it was calibrated to do so, the model fits very well the cumulative data. It predicts that if the government continuous with the current intensity of the intervention the number of fatalities would reach 27.700 individuals by July. It also predicts that without intervention the number of fatalities would be around 650.000. This last scenario will be discussed in the next section.

In Panel b) of Figure 1 we show the implied **daily** paths by the model, this time including the model infected individuals. The actual "measured" infected cases are plotted with the red circles. There are two important takeaways from this figure. The daily predictions are fairly good. The model predicts that there is at least one more month in which the daily fatalities would be above 100 individuals. More importantly, the model implied cases are substantially above the actual measured cases (recall that the scale is logarithmic). To put it in a context, at the pick of the "measure" increase in infected individuals, the data states that 6.550 became infected, while the model states that the same day there were 29.000 new infections.

3.1.1 Estimating fatality rates

There is much controversy about the "true" value of the fatality rate, especially when all the available data is too raw to provide a concrete answer. Most studies tend to state that on average 1% of the infected die. However, this value could significantly change with the demographic structure of the population. In particular, the fatality rate appears to sharply increase with age and the lack of proper treatment. Since we are focusing on the Italian case, both factors are first order issues for our estimations. The study by Ferguson et al. (2020) estimates that the average fatality rate in Wuhan is around 0.99%. The same paper states that around 4.4% of the infected subjects require hospitalization. They also estimate that 30% of the hospitalised cases require critical care; and even when the patient receives proper critical care she dies with 0.5 probability. If we assumed that without critical care the subject dies with certainty, that implies that the fatality rate for the untreated is twice the analogous for the treated. This indicates that $\delta \approx 2 \times \theta$, providing a first support to our calibration.

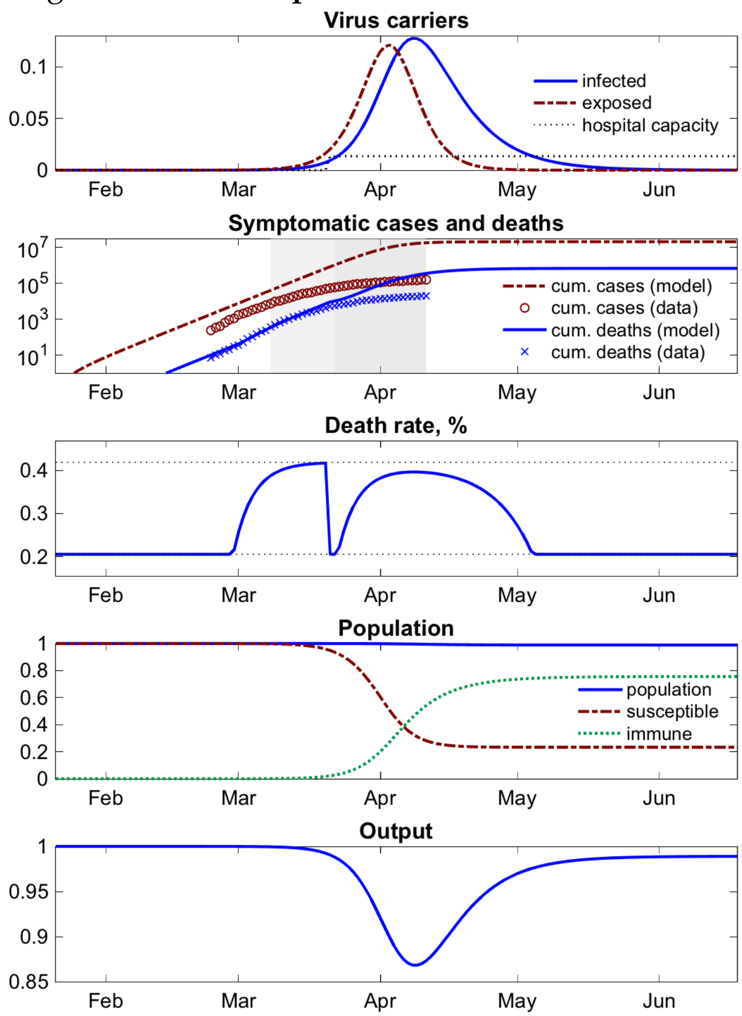
Another calculation to determine the difference between θ and δ is to compare the fatality rates in a country that didn't reached the health capacity with another that did. A candidate for the first is South Korea, while Italy is a clear candidate for a country in which the health system was overwhelmed. In Appendix B we present the information for the fatality rates by age for both countries. To obtain the average fatality rate, we multiply each age-specific rate by the relative weight of that age group in the population. We obtain that the average rate for South Korea is 1.22%, while for Italy is significantly larger at 4.09%. But, how much of this total difference is due to the different age distributions and how much due to the difference in the health systems? We made an intermediate calculation where we recompute the average death rate for South Korea, but using the population weights of Italy, which delivers 1.92%. We interpret the difference $1.92\% - 1.22\% = 0.7\%$ as the pure age composition effect. This number is by itself substantial and informative about the significant risk that COVID-19 represents for an "old" country as Italy. It is striking that only the age adjustment generates a death rate almost identical to the calibration using the virus dynamic's information.

Still, the observed fatality rate in Italy is, so far, more than 4% and, if our adjustment is correct, the additional two percentage points are not explained by the age composition. If we attribute the difference to the lack of proper medical attention, we obtain again that $\delta \approx 2 \times \theta$. Since these two independent sources deliver consistent estimations, we calibrate our model with $\delta = 2 \times \theta$.

3.2 Simulated paths without intervention

With these parameter values we can estimate what would be the evolution of the illness, and its economic impact, if it were to spread in an unrestricted fashion, i.e., with $q_t = 0$ for all t. In Figure 2 we plot the proportion of infected and exposed subjects at each day. To clarify the dynamics we also the cumulative proportion of Total cases, susceptible and immune individuals at each day. The economy starts with an initial mass of $\frac{100}{60.000.00}\%$ of exposed individuals and 0 infected. Initially the exposed move around and engaged in economic activities without necessarily knowing that they are carriers. Soon after, some "confirmed" infected start to arise, but still those numbers are very small, and definitively smaller than the number of exposed individuals. In this period, the growth rate of the infection is large, around 100% per day, but the quantities do not seem alarming due to the still small number of affected individuals. After 45 days, the number of infected are about the same as the number of exposed. At this point, if a policy maker takes a picture of the situation, it can only see the type i individuals, but the number of carriers is $2 \times i$.

Figure 2: **Potential path: no intervention scenario**



Nevertheless, the number of infected cases are still small, although growing over time. The initial slope is steep, with the growth in the number of total cases in an apparent explosive path. The situation deteriorates after around 45 days when the hospital capacity is reached. The fatality rate that was low at the beginning starts to rise due to the infected that are either untreated or badly treated. After 80 days the number of infected is at its maximum, with around 13% of the population symptomatically infected and 10% have been affected, but do not show symptoms yet.

At this point, the growth rate of the infected starts to decrease. The main reason for this is that the number of susceptible people reached a point, around 75% of the population with our calibration, such that the reproduction rate of the virus $\lambda(S/N)$ is smaller than its death rate, given by $\eta + \Delta_t$. After that, the virus starts to die by itself. Eventually, it disappears with 75% of the population who have developed and immunity. The cost in lives is large, without intervention 1.1% of the population dies, with the analogous effect on total production and consumption.

There are two important takeaways from these simulations. First, the virus needs unaffected individuals to reproduce. As the infection spreads, the number of susceptible unaffected individuals decreases. More and more meetings start to happen between exposed and already immune individuals. It is true that still some new subjects become infected, but every period less individuals are becoming infected than the people who are either recovering or dying. For our calibration, 75% of the population immune is enough to eradicate the viruses. Of course, that depends on the meeting function. With other meeting functions that could be different. This is important because it determines the feasibility of a *mitigation* (flattening the curve) policy. To take that approach, one must have a very good knowledge about the features of the meeting function. Second, notice that there is a rebound in production after the illness reaches the peak. This happens because the recovered patients are allowed to return to their jobs. The contribution is not trivial and it must be properly considered when designing contention policies. For instance, the current quarantines do not distinguish between subjects that are already immune from those with uncertain status. If the recovered status is known to the planner, they must be allowed to work.

4 Optimal intervention

4.1 Simple quarantines

In most countries affected by COVID-19 the approach has been to impose quarantines for a determined period of time. Moreover, the intensity of the quarantine has been changing over time and it is expected to be decreased over time to a point in which all restrictions would cease to exist. The response of the intervention has been mainly driven by the information about the number of infected cases. As a result, we can think about these type of policies as restricting the set of policies q_t to be a three-parameter step function such that, for some $\tilde{q} \in [0,1]$, $b \in \mathbb{R}$ and $\tau \geq 0$, the government intervention satisfies:

$$q_t = \begin{cases} \tilde{q} + b \times I_t, & \text{if } t \leq \tau \\ 0, & \text{if } t > \tau. \end{cases} \tag{6}$$

For instance, a complete shutdown of all economic activities for two weeks would be represented by $\tilde{q} = 1$, $b = 0$ and $\tau = 15$, and any fixed intensity intervention would be characterized by the set of policies with $b = 0$. The parameter b would capture the time varying intensity component.[17] Here we are assuming that q_t depends only on I_t, why not to make it depend on other state variables? Because of the structure of the model, all variables are deterministically linked, thus all state variables contained the same information. However, different state variables are potentially shaped in different ways, which could help to better approximate the optimal unrestricted policy. To understand this potential concern we have estimated the optimal policy assuming the dependency of q_t on E_t and $\Delta_t I_t$, we found that the policy depending on I_t as in (6) generates the highest welfare. In Appendix C we characterize the optimal path for q_t in an unrestricted functional space.[18] Nevertheless, in this section we compute the optimal "quarantine" duration and time varying intensity to provide some intuition about the forces at play and to compare it with policies that have actually been implemented.

One important determinant of the optimal policy is the existence of the critical mass $\underline{E}$. As long as $\underline{E} = 0$ the suppression policy has little chances, no matter how small is the exposed population, as long as it remains positive there would be new waves of contagion. The only long term solution is to build a mass of immune individuals to prevent the reproduction of the virus. However, when $\underline{E} > 0$ the intervention could aim to completely eliminate the virus without building a large stock of immune subjects and still preventing further waves. To assess these two very different strategies we assume that the critical mass is $\underline{E} = \frac{1}{60\text{million}}$. This means that if the government manages to reduced the number of exposed to less than one individual in the popu-

[17] We have run experiments with q_t functions that depend directly on t and allow time changing shapes. We found that the optimal policy is initially increasing and then decreasing. This drives us to believe that the shape embodied in I_t is not overly restrictive.

[18] See Alvarez, Argente, and Lippi (2020) for a solution in the unrestricted policy space.

lation the virus disappears.

The main results are shown in Table 2 and Figure 3. We assume that the quarantine is implemented in day 48, which corresponds to March 8^{th} in the calendar of our calibration: the day in which the Italian government decided the first intervention. We present two sets of results. The first set, in columns (2) and (3) of Table 2, is the optimal quarantine if the intended outcome is to **suppress** the virus, which is always feasible when $\underline{E} > 0$. The second set, columns (4) and (5), shows the optimal quarantine if the intervention only seeks to **mitigate** the spread. This corresponds to the "flattening the curve" strategy. In both cases, we show the results with linear and logarithmic utility. Finally, column (1) shows some analogous statistics for the scenario without intervention. An important consideration when reading these results is that the value of life $v(\cdot)$ is innocuous shaping the dynamics of the optimal q_t. As we discuss later the value of life generates discrete changes in the long term strategies. When $\underline{E} > 0$ and $v(\cdot)$ is below a certain threshold $\underline{v}$, the optimal intervention calls for mitigation, while above $\underline{v}$ the optimal intervention is suppression. But conditional on either suppression or mitigation the optimal q_t is almost invariant to $v(\cdot)$.

Moreover, the estimated number of fatalities in the no intervention scenario, which ranges from $600,000$ to $780,000$, is in line with the calculations by the panel of experts in Walker et al. (2020), who estimate around $645,000$ fatalities for Italy without any intervention.

There are many interesting results worth mentioning. First, suppose that $\underline{E} > 0$ and the policy is intended to suppress the virus. Then, the optimal intervention takes the form of a fixed intensity for a determined time span. With linear utility, column (2), it is optimal to implement a complete lockdown for three weeks (exactly 23 days), which is very close to the initial recommendations by experts in epidemiology and the announced initial time of quarantine by most governments. This policy is extremely effective in reducing the number of symptomatic total cases from 41% to 0.11% of the population, reducing also the total fatalities from 1.1% to 0.0035% of the population. However, comparing with column (3), when there is a desire for cost smoothing, the optimal policy reduces the intensity by about a third and it duplicates the length of the quarantine. This "mild" quarantine last for more than seven weeks rather than three weeks. The smoothed policy has a cost in terms of lives though, the fatality rate is 0.0045% rather than 0.0035%, which implies 30% more fatalities. To put it in con-

Table 2: Optimal fixed intensity quarantine

	No Intervention (1)	Suppression ($\underline{E} > 0$) Lin. Util. (2)	Log Util. (3)	Mitigation ($\underline{E} = 0$) Lin. Util. (4)	Log Util. (5)
Quarantine:					
Initial day	-	48	48	48	48
Duration	-	23	51	89	75
Maximum q	-	1	0.71	0.41	0.4
Average q	-	1	0.7	0.15	0.16
Symptomatic rate (per pers.)	33%	0.11%	0.15%	20%	20%
Symptomatic ppl. (number)	20mn	66,000	90,000	12mn	12mn
Immunity rate (per pers.)	76%	0.25%	0.35%	46%	46%
Immune ppl. (number)	45mn	150,000	210,000	28mn	28mn
Death rate (per pers.)	1.1%	0.0035%	0.0045%	0.53%	0.55%
Total fatalities	660,000	2,000	2700	320,000	330,000
Welfare gain (consumption equiv.)	-	0.82%	-0.02%	0.42%	0.38%

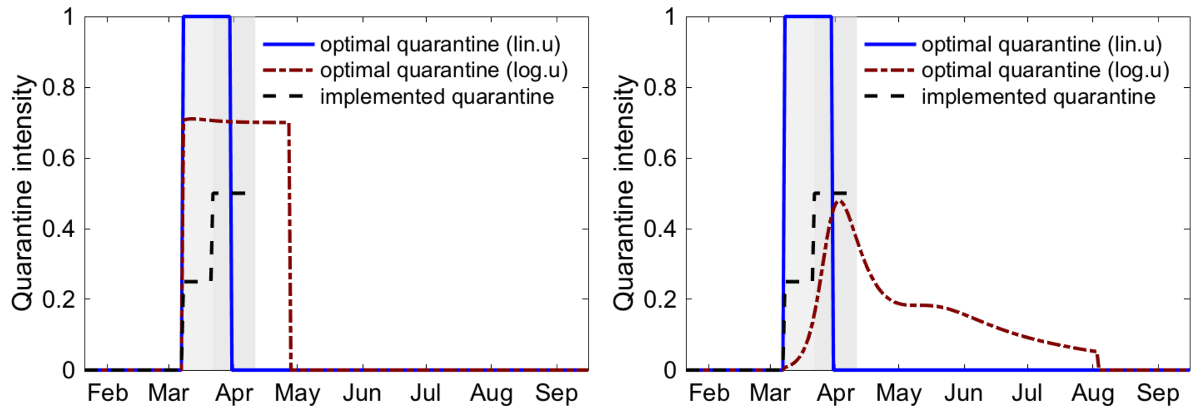
text, with the Italian population, the number of deaths without intervention would be around $660,000$ people, with the optimal linear utility intervention is only 2.000 individuals and with the optimal smooth intervention is 2.700 individuals.

In panel (a) of Figure 3 we plot both optimal policies and we compare them with the calibrated intensity of the observed quarantine. A simple inspection reveals that this quarantine is not enough to generate suppression. However, we are abstracting from other measures, like social distancing, that could bring the implemented q (which measures only economic impact) closer to the optimal suppression policy.

To avoid the high cost in output, or when $\underline{E} = 0$, the optimal intervention could aim to mitigate the spread of the virus. This possibility is shown in columns (4) and (5) and panel (b) of Figure 3. In this case it is not very important whether there is a concern for cost of smoothing. With both utility functions the optimal policy starts slowly and grows fast reaching an intensity of $q_t = 0.4$ by the end of March and then starts to reduce the intensity lasting until June. Overall the intervention lasts between 75 and 90 days with an average intensity of 0.15. This policy has a large cost in terms of both lives and output. The number of fatalities are only half of those without intervention whatsoever, 330.000 vs. 660.000, but there is a large build up of

Figure 3: **Optimal quarantines**

Panel a): Suppression Panel b): Mitigation

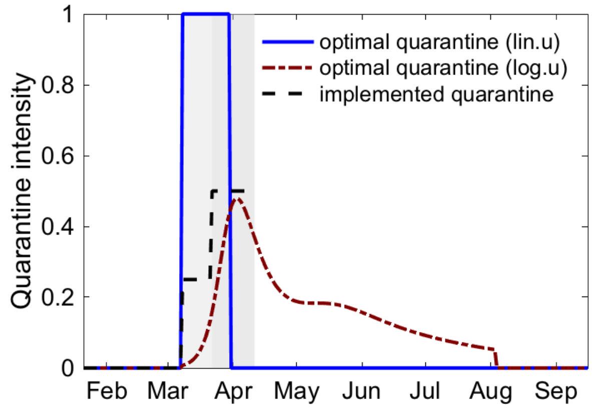


immune individuals, around 28 millions, that prevent the arrival of future waves. The desire for cost smoothing only makes the quarantine slightly shorter keeping more or less the same intensity. The optimal policy with mitigation depends on the implicit dynamics of the model rather than on the structure of the welfare function.

When comparing both suppression and mitigation with the observed policy, it is clear that the observed policy is too strong to be a mitigation policy and too soft to be a suppression policy. There are two caveats to this interpretation. First, we are only considering policies that prevent economic interactions without affecting how the remaining interactions take place. The policy of "social distancing" can be understood as a reduction in λ, so that each interaction generates less infections. This would considerably reduce the number of fatalities. Nevertheless, the suppression results are less affected by the way in which interactions take place: if there are no meetings, it is not important how they happen. But, when the intended policy is mitigation social distance is a very important complement. In this sense, we see these results as stressing the relevance of social distancing as a fundamental complement to indiscriminate economic quarantines.

The second caveat is that we have not considered the extra value of life yet. As we show in the last row of Table 2, the suppression policy is optimal when the utility is linear, but when the utility is logarithmic the mitigation policy is optimal. In fact, when the utility is logarithmic, the suppression policy is worse than the no interven-

Figure 4: **Larger value of life**

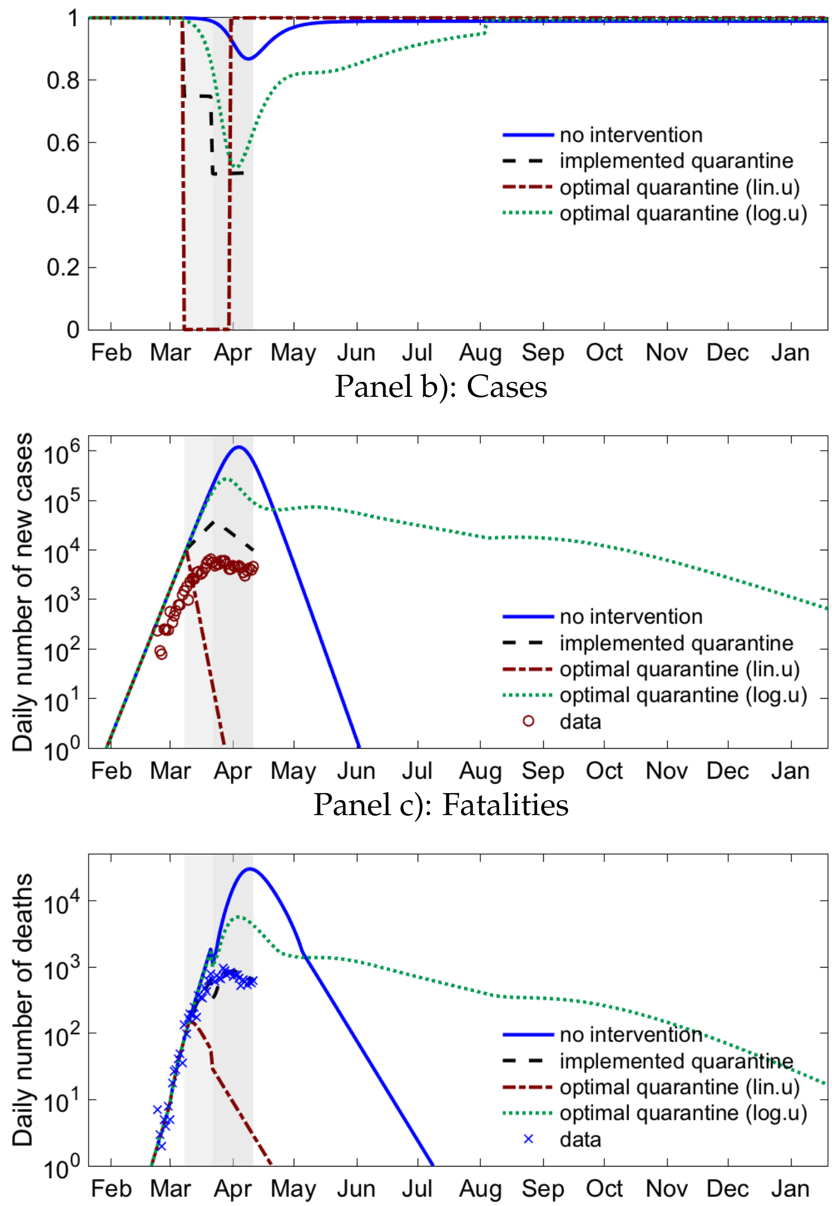


tion scenario. In unreported results we show that suppression is always optimal with linear utility, except when $\underline{E} = 0$, in which case a linear-planner would seek also mitigation. Instead, when the utility is logarithmic the mitigation strategy is always optimal when $\underline{E} = 0$, and when $\underline{E} > 0$ and the extra value of life is below a threshold $\underline{v}$. With our calibration that threshold is $\underline{v} = 2.800$ days of output. That is, as long as the statistical value of a life is smaller than 8 years of annual income the log-planner would seek mitigation. Of course, because lives are more costly, the intensity of the quarantine is larger.

In Figure 4 we plot the optimal quarantines with linear and log utility when the value of life is 2.700 days of output. Including the value of life does not change the optimal linear-policy. If the strategy is suppression and the number of deaths are reduced to a minimum, how much those lives are valued does not significantly change the optimal intervention. Although, it does reassure that suppression is optimal. When the planner has a desire for smoothing, the optimal policy is still mitigation, but now the intensity is higher, peaking at around 0.5. The intervention also last longer, remaining until August, although the intensity is reduced to 0.15 in May, and from June on is below 0.1. This stronger policy reduces the number of fatalities from 330.000 to 260.000, which is substantial, but still a large number. Increasing the value of life further would not increase the intensity of the mitigation policy, it would just generate a switch from mitigation to suppression even for the log-planner. Since the observed policy seems to resemble better the optimal mitigation of a log-planner with a "high" valuation of life, in Section 5 we will take this calibration as a benchmark.

Figure 5: **Intervention and effects**

Panel a): Output



In Figure 5 we plot the effect of the optimal interventions described in Table 2 over output, Panel a), number of cases, Panel b) and Fatalities, Panel c). All the patterns are fairly intuitive. The linear utility intervention generates a large drop in output, but also quickly reduces the number of cases and fatalities. In contrast, the log-intervention has a smaller impact on production, but accumulates more infected cases and generates more fatalities.

5 Combining testing and quarantines

The main assumption generating the necessity of an indiscriminate quarantine policy, is the inability of the policy maker to distinguish the exposed subjects from those that are susceptible, but have not been affected yet. If the government knew at each time who are the virus' carriers, it could simply quarantine the exposed subjects and allow everyone else to work to avoid the output cost. The technology to do so is certainly available, but it could be prohibitively costly to undertake such an approach over a vast proportion of the population. However, since the immediate output cost of the quarantine appears to be also very large, it is worth evaluating how much the planner would be willing to spend on testing to reduce the cost of the quarantine.

To deal with this problem we divide the population of exposed individuals in two groups: the unidentified exposed and the exposed population that has been designated as a *positive* carrier of the virus. We maintain the notation E_t for those subjects that carry the virus, but do not know it. These individuals are indistinguishable from those in the group S_t and therefore the government must still set $q_t^E = q_t$. The same rule must also apply to individuals who where previously E_t and recovered without ever exhibiting symptoms. To separate them, the government can test randomly a subset of individuals in the set $S_t + E_t + R_t^u$. If the test result is positive, it means that the subject carries the virus, it is identified with the new group E_t^p, and it is forced into mandatory quarantine, as the group I_t, until she fully recovers. i.e., the group E_t^p is assigned a quarantine measure q_t^I. This group of individuals maybe asymptomatic and they may remain so until they are fully recovered or develop symptoms. Notice that we assuming that the testing technology cannot detect antibodies, for all practical purposes when an individual test negative it could be either S_t or R_t^u, which remains unknown to the tester. Now the total population is:

$$N_t = \underbrace{S_t + E_t + R_t^u}_{\text{unidentified}} + \underbrace{E_t^p + I_t + R_t}_{\text{identified}}.$$

To understand the relevance of testing it is useful to first present the new law of motion for E_t. Suppose the government randomly screens α_t of the individuals in the group $S_t + E_t + R_t^u$, it can identify $\alpha_t E_t$ individuals as positive carriers. Then, the new law of motion for E_t is:[19]

$$dE_t = \begin{cases} \left(\lambda \frac{S_t}{L_t} (1-q_t)^2 - (\gamma + \sigma + \alpha_t)\right) E_t dt, & \text{if } E_t \geq \underline{E} \\ -(\gamma + \sigma + \alpha_t) E_t dt, & \text{if } E_t < \underline{E}. \end{cases} \tag{7}$$

Equation (7) shows the first positive contribution of testing to welfare. Recall that only the group E_t can spread the disease, so the smaller the group, the smaller the contagion rate. Comparing (7) with (1) is evident that testing adds a downwards drift α_t to the population of asymptomatic individuals. Before the group was reducing only when they were either becoming actively infected at rate γ or recovering at rate σ, but now some individuals are also exiting the group because some are identified, at rate α_t, as positive carriers and, hence, cannot infect anyone else.

As the unidentified exposed, the positively identified subjects can eventually become symptomatic and join the group of infected at rate γ, or recover, at the same rate σ as the E. Unlike the E subjects, when an E^p recovers she joins the group of recovered R_t rather than R_t^u, and therefore she is allowed to work. The law of motion for E_t^p and the new law of motion for R_t satisfy:

$$dE_t^p = \alpha_t E_t dt - (\gamma + \sigma) E_t^p dt \tag{8}$$

$$dR_t = (\eta I_t + \sigma E_t^p) dt \tag{9}$$

$$dR_t^u = \sigma E_t dt \tag{10}$$

Comparing equation (9) to (4) we can see the second important contribution of testing. Since the recovered are immune and allowed to work, as they recover they rejoin the labor force at rate σ, which is useful in reducing the output costs of the quarantine. In short, the group of positively tested individuals generate a bulk that reduces the speed of contagion and increases the available resources to get by the quarantine times. This is especially important when the exposed may never be symptomatic. Without the testing, they would never be sick, and therefore they would always be treated as susceptible population subject to quarantines. In this new environment the law of motion of infected is slightly modified to:

$$dI_t = [\gamma (E_t + E_t^p) - (\eta + \Delta_t) I_t] dt \tag{11}$$

[19] Here we assume for $m(\cdot)$ the functional form described in Section 2.1.

The only difference with the previous section is the inflow of positive exposed subjects which happens at rate γ. The population's law of motion remains exactly the same $dN_t = -\Delta_t I_t dt$, since the infection only affects the population by the death rate; and to die a subject must show symptoms first, which only happens if they previously were part of the I_t group. Finally, the production feasibility set remains the same as before, with the mass $E_t + S_t + R_t^u$ subject to quarantines but the R_t allowed to work. We only subtract the cost of the tests.

Suppose the government test x_t individuals at each instant, then the flow cost is governed by the convex cost function $\Phi(x)$, with $\Phi(0) = 0$, $\Phi'(x) > 0$ and $\Phi''(x) > 0$. Given the previous description that the government screens the population with intensity α_t, the number of tests at each instant are $x_t = \alpha_t (S_t + E_t + R_t^u)$. As a result, the feasibility constraint becomes:

$$Y_t = (1 - q_t)(S_t + E_t + R_t^u) - \Phi(\alpha_t (S_t + E_t + R_t^u)) + R_t$$

We maintain the previous parametrization, see Table 1. We maintain the assumption that q_t satisfies equation (6) and we assume a similar structure for the testing function:

$$\alpha_t = \begin{cases} \tilde{\alpha} + b_\alpha \times I_t, & \text{if } t \leq \tau \\ 0, & \text{if } t > \tau. \end{cases} \tag{12}$$

We do not have information about the cost of each test. We start by assuming that the minimum marginal cost is 1 day of output, and then the cost grows in a quadratic way after that. We parameterize the quadratic component in such a way that testing the whole population might not be economically feasible.

In Table 3 we present the optimal policies. We compare the optimal combination of quarantine and testing policies in columns (4) and (5) with the optimal policy that uses only indiscriminate quarantines, in columns (2) and (3), and the outcome without intervention in column (1). The first thing to notice is that testing is used intensively, 62% of the unidentified are tested when the utility is linear and, at the peak, 6% when the utility is logarithmic. These numbers, even with the logarithmic utility, are far larger than the observed testing strategies. Also, in both cases there are welfare improvements. With the logarithmic utility the consumption equivalent gain is 10% above the value of the policy without testing. The important takeaway from this

Table 3: Optimal quarantine and testing policies

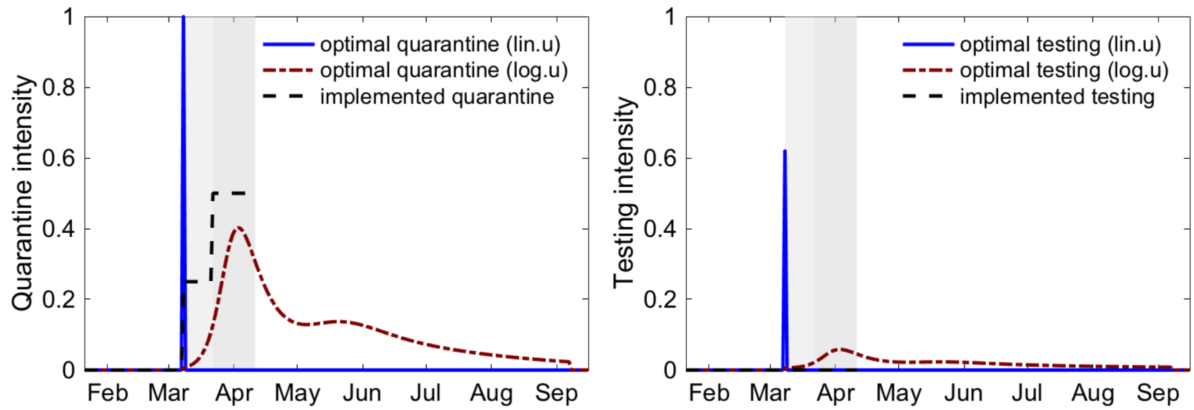
	No	Quarantine Only		Quarantine & Testing	
	Intervention (1)	Lin. Util. (2)	Log Util. (3)	Lin. Util. (4)	Log Util. (5)
Intervention:					
Initial day	-	48	48	48	48
Duration	-	23	149	1	184
Quarantine:					
Maximum q	-	1	0.48	1	0.4
Average q	-	1	0.17	1	0.11
Testing:					
Maximum α	-	-	-	0.62	0.06
Average α	-	-	-	0.62	0.02
Total cost (% of GDP)	-	-	-	0.7%	1%
Symptomatic rate (per pers.)	33%	0.11%	19.6%	0.11%	19.1%
Symptomatic ppl. (number)	20mn	66,000	11.8mn	66,000	11.4mn
Asymptomatic rate (per pers.)	0	0	0	0.04%	1.6%
Asymptomatic ppl. (number)	0	0	0	22,000	970,000
Immunity rate (per pers.)	76%	0.25%	45%	0.25%	44%
Immune ppl. (number)	45mn	150,000	27mn	150,000	26mn
Death rate (per pers.)	1.1%	0.0035%	0.43%	0.0035%	0.42%
Total fatalities	660,000	2,000	260,000	2,000	250,000
Welfare gain (consumption equiv.)	-	0.8%	0.3%	1.1%	0.32%

result is that *testing is a substitute rather than a complement of quarantines*. Looking at the fatality numbers and the total infected cases is evident that these numbers are very similar to those in which testing is not allowed. *The main difference lies on the path for output, which is what generates the welfare gains*.

Since testing is costly, there are important differences depending on the curvature of the utility function. When the utility is linear, so that the concern is more about productive efficiency, testing completely replaces the duration of the quarantine. Rather than doing indiscriminate and inefficient quarantines, it is optimal to stop production for one day, test unidentified subjects and resume production as soon as possible. The time path for this testing policy can be seen in the blue line of Figure 6, panel b). Testing spikes for a day, then it is reduced to zero from then on. In panel a) we plot the simultaneous quarantine intervention in the blue line. Now, the quarantine is less intense than without testing. The first day all production is completely shutdown, but after the second day there is no more intervention. Testing is costly, with our calibration it amounts to 0.7% of GDP.

Figure 6: **Optimal quarantine and testing**

Panel a): Quarantine　　　　Panel b): Testing



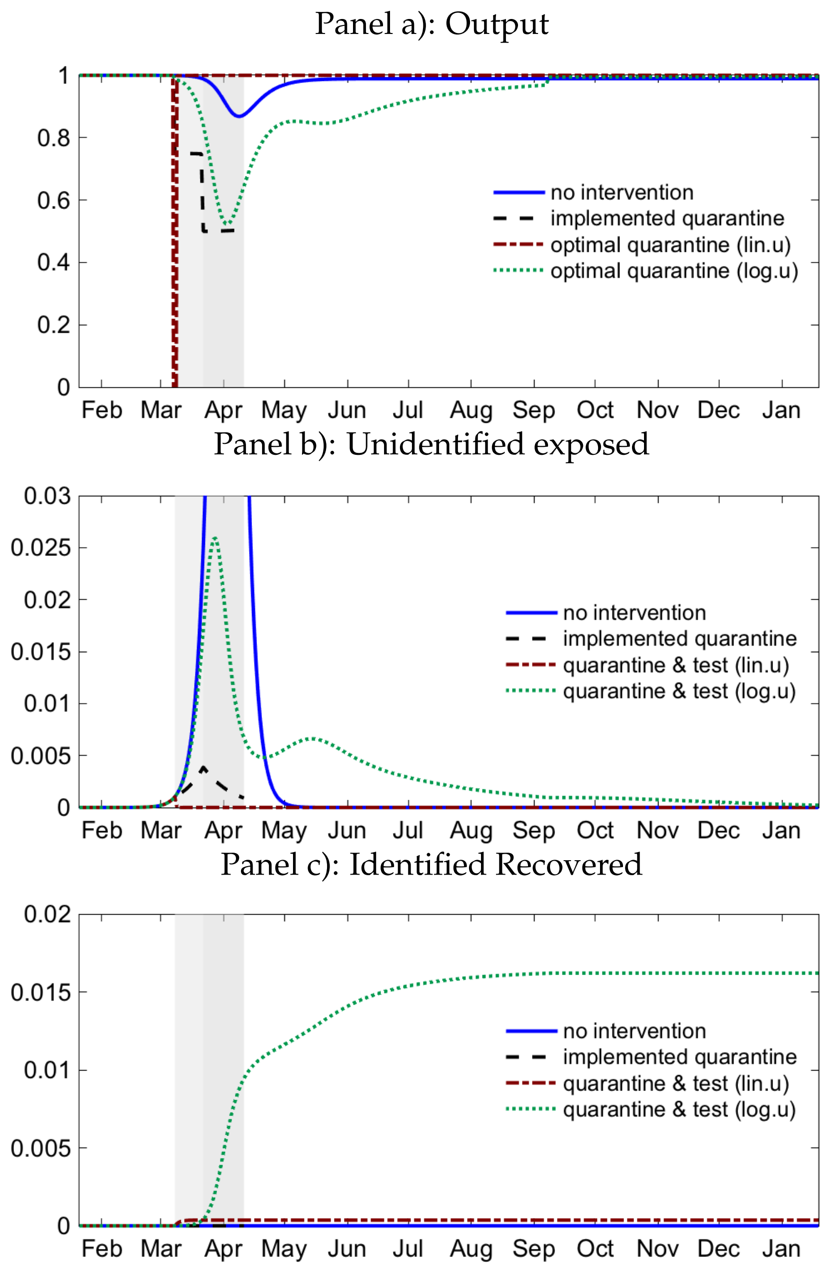
When smoothing is a concern, the optimal policy generates the same duration as without testing, but with a slowly growing testing policy. The quarantine has a similar shape as when testing is not possible, but the intensity is reduced (compare with Figure 4). The lower intensity is replaced with a continuous testing policy that entails to about 6% of the unidentified population at its peak, but it is most of the time around 2%.[20] One way to think about this policy is to consider a situation with long lasting restriction, around 6 months, with some constraints in economic activity and continuous testing of unidentified subjects.

Again, the calibrated policy is expensive, at the end it amounts to around 1% of annual GDP. Since we are assuming random testing, a planner could do it better using additional information, such as the likelihood that an individual has been exposed or the relevance of the subject in the production network. These considerations would only tilt our result more in favor of testing rather than indiscriminate quarantines. In any case, notice that the optimal testing policy follows the path of potentially exposed individuals. The larger the fraction of exposed, the more likely that a test is successful at identifying a positive case. The peak of testing is not at the beginning

[20] We want to emphasize that the percentage is with respect to the unidentified $s_t + e_t$, not with respect to the entire population. So that the number of tests is continuously decreasing over time.

of the outbreak, but rather later on, in our case study mid April, enough time for the governments to plan a testing strategy.

Figure 7: **Output with testing**



The implied output and number of unidentified asymptomatic individuals by the optimal testing strategy can be seen in Panels a) and b) of Figure 7, respectively. Panel c) presents an additional measure that only exist with testing: the asymptomatic individuals what were previously identified as positive and then recovered. Because now it is know that they are immune, they are allowed to work. This measure becomes particularly relevant with the long quarantines. After three months, it amounts to 1.5% of the labor force.

6 Conclusions

In this paper we have extended the standard epidemiologic SIR model allowing for asymptomatic subjects to be tested and consider the trade-off with output losses. We show that if the government has not means to identify the carriers of the virus, the observed mandatory quarantines around the world seem to be close to what it can be considered optimal.

However, if the government can increase the intensity of testing over subjects, that is a far superior strategy. We acknowledge that ultimately this statement depends on the cost of actually performing those tests. The results of this paper indicate that carefully analyzing and assessing this possibility should be a priority.

References

References

Alvarez, Fernando, David Argente, and Francesco Lippi. 2020. "A Simple Planning Problem for COVID-19 Lockdown." Discussion paper, EIEF.

Atkeson, Andrew G. 2020. "What will be the economic impact of COVID-19 in the US?" Working Paper, UCLA.

Dewatripont, Mathias, Michel Goldman, Eric Muraille, and Jean-Philippe Platteau. 2020. "Rapid identification of workers immune to COVID-19 and virus-free: A priority to restart the economy." Discussion paper, Universit Libre de Bruxelles.

Eichenbaum, Martin S., Sergio Rebelo, and Mathias Trabandtz. 2020. "The Macroeconomics of Epidemics." Author's website, Northwestern University.

Ferguson, Neil M., Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunuba, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, and Han Fu a. 2020. "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand." On behalf of the imperial college covid-19 response team, Imperial College of London.

Guiso, Luigi, and Daniele Terlizzese. 2020. "Sfruttare la chiusura per riaprire prima possibile." *Il Foglio*, no. March 24th.

Kermack, William. O., and Anderson. G. McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115 (772): 700–721.

Walker, Patrick GT, Charles Whittaker, et al. 2020. "The Global Impact of COVID-19 and Strategies for Mitigation and Suppression." On behalf of the imperial college covid-19 response team, Imperial College of London.

Appendix

A Model calculations

Mapping to data moments: If untreated, the density function for dying after s units of time is

$$f^u(s) = \delta e^{-(\eta+\delta)s}$$

The death rate is

$$\int_0^\infty f^u(s)\, ds = \int_0^\infty \delta e^{-(\eta+\delta)t} ds = \frac{\delta}{\eta+\delta}.$$

The density function for recovering after s units of time is

$$g^u(s) = \eta e^{-(\eta+\delta)s}.$$

The average recovery duration is

$$\int_0^\infty g^u(s)\, s ds = \int_0^\infty \eta e^{-(\eta+\delta)s} s ds = \frac{\eta}{(\eta+\delta)^2}.$$

Similarly, if treated, the death rate is $\frac{\theta}{\eta+\theta}$. The average recovery duration is $\frac{\eta}{(\eta+\theta)^2}$.

B Death rate data

Table 4: Fatality rates South Korea and Italy

Classification		South Korea						Italy					
		Cases		Fatal cases			Weight	Cases		Deaths		Lethality	Weight
		Number	(%)	Number	(%)	Rate (%)	age group	Number	(%)	Number	(%)	(%)	age group
All		**9,137**	**100**	**126**	**100**	**1.38**	**1**	**35,731**	**100**	**3,047**	**100**	**8.5**	**1**
Age	**Above 80**	406	4.4	55	43.65	13.55	0.0342	5,352	15	1,243	50.2	23.2	0.0717
	70–79	611	6.7	39	30.95	6.38	0.0672	7,121	19.9	1,090	35.8	15.3	0.0988
	60–69	1154	12.6	20	15.87	1.73	0.1198	6,337	17.7	312	10.2	4.9	0.1216
	50–59	1724	18.9	10	7.94	0.58	0.1648	6,834	19.1	83	2.7	1.2	0.1549
	40–49	1246	13.6	1	0.79	0.08	0.1626	4,396	12.3	25	0.8	0.6	0.1531
	30–39	943	10.3	1	0.79	0.11	0.1405	2,525	7.1	9	0.3	0.4	0.1172
	20–29	2473	27.1	0	0	0	0.1327	1,374	3.8	0	0	0	0.1027
	10–19	475	5.2	0	0	0	0.0954	270	0.8	0	0	0	0.0956
	0–9	105	1.2	0	0	0	0.0828	205	0.6	0	0	0	0.0843

C Optimal control problem

Choose the path of quarantine policies:

$$\max_{\{q_t: t\geq 0\}} \int_0^\infty e^{-\rho t} u(c_t)\, dt$$

$$\text{s.t, } c_t = n_t - i_t - s_t q_t$$

and (1), (4), and (5)

Utility options: linear, log, quadratic $c - \frac{b}{2}c^2$.

Hamiltonian:

$$\max_{\{q_t : t \geq 0\}} \int_0^\infty e^{-\rho t} U(n_t - i_t - s_t q_t)\, dt$$

subject to

$$\begin{aligned} de_t &= \left[\lambda \frac{s_t}{n_t}(1 - q_t) - \gamma_t\right] e_t dt \\ di_t &= [\gamma e_t - (\eta + \Delta_t) i_t]\, dt \\ dr_t &= \eta i_t dt, \\ dn_t &= -\Delta_t i_t dt. \end{aligned}$$

$$\mathcal{H}(q, e, i, r, n) = U(n - i - sq) - \phi_1 \left[\lambda \frac{s}{n}(1-q) - \gamma\right] e - \phi_2 [\gamma e - (\eta + \Delta) i] + \phi_3 \eta i - \phi_4 \Delta i.$$

Keep notation $c = n - i - uq$.

$$[q] \qquad U'(c) \lesseqgtr \phi_1 \lambda \frac{e}{n}$$

$$[e] \qquad U'(c)\, q + \phi_1 \lambda \frac{e}{n}(1-q) - \phi_1 \left[\lambda \frac{s}{n}(1-q) - \gamma\right] - \phi_2 \gamma = \rho \phi_1 - \dot{\phi}_1$$

$$[i] \quad -U'(c)(1-q) + \phi_1 \lambda \frac{e}{n}(1-q) + \phi_2 \left(\eta + \Delta + \frac{\partial \Delta}{\partial i} i\right) + \phi_3 \eta - \phi_4 \left(\Delta + \frac{\partial \Delta}{\partial i} i\right) = \rho \phi_2 - \dot{\phi}_2$$

$$[r] \qquad U'(c)\, q + \phi_1 \lambda \frac{e}{n}(1-q) = \rho \phi_3 - \dot{\phi}_3$$

$$[n] \qquad U'(c)(1-q) + \phi_1 \lambda \frac{e}{n} \frac{n-s}{n}(1-q) = \rho \phi_4 - \dot{\phi}_4.$$

This will imply that with concave utility for example log utility, since in the beginning $n = s$ and $i = 0$, $c = n(1-q)$, the optimal q is in the interior.

Steady state: $e = 0$, $q = 0$, $\phi_4 = \frac{1}{\rho} U'(c)$, $\phi_3 = 0$,

$$-U'(c) + \phi_2 (\eta + \Delta) - \frac{1}{\rho} U'(c) \left(\Delta + \frac{\partial \Delta}{\partial i} i\right) = \rho \phi_2$$

$$\phi_2 = \frac{1}{\rho} \frac{\Delta + \rho}{\eta + \Delta - \rho} U'(c)$$

$$U'(c) - \phi_1 \left[\lambda \frac{s}{n} - \gamma\right] - \phi_2 \gamma = \rho \phi_1$$

Initial conditions: $e(0) = e_0$, $i(0) = 0$, $r(0) = 0$, $n(0) = 1$.

Terminal condition: $e(T)=0$, $i(0)=0$, $q(T)=0$, $\phi_3(T)=0$, $\phi_4(T)=\frac{1}{\rho}U'(n(T))$

Variables: e, i, r, n, q, ϕ_1, ϕ_2, ϕ_3, ϕ_4.

Simplifying, only when the eq [q] is with equality:

$$[q] \qquad U'(c)=\phi_1\lambda\frac{e}{n}$$

$$[e] \qquad U'(c)-\phi_1\left[\lambda\frac{s}{n}(1-q)-\gamma\right]-\phi_2\gamma=\rho\phi_1-\dot{\phi}_1$$

$$[i] \qquad \phi_2\left(\eta+\Delta+\frac{\partial\Delta}{\partial i}i\right)+\phi_3\eta-\phi_4\left(\Delta+\frac{\partial\Delta}{\partial i}i\right)=\rho\phi_2-\dot{\phi}_2$$

$$[r] \qquad U'(c)=\rho\phi_3-\dot{\phi}_3$$

$$[n] \qquad U'(c)(1-q)\left(1+\frac{n-s}{n}\right)=\rho\phi_4-\dot{\phi}_4.$$