Closed-form Solutions for Optimal Social Distancing in a SIR model of COVID-19 Suppression

Reyer Gerlagh[*]

May 22, 2020

Abstract

I present a stylized suspected-infected-recovered (SIR) model of COVID-19, with symptomatic versus asymptomatic patients, and social distancing intervention. The optimal suppress strategy has low-infection rates, enabling assumptions that support closed-form solutions.

The model predicts high costs of social distancing in comparison to health costs of the disease; it separates public versus private benefits of social distancing, and determines the required level of group immunity for relaxing social distance intervention.

I extend the model with heterogeneous population for preferences over social contacts, health costs, and transmission. Heterogeneity in transmission intensity offers most opportunities for reduced costs under a differentiated social distancing policy.

keywords: COVID-19, SIR model, Suppression, (differentiated) Social distancing
JEL codes: I18 (health policy), E17 (simulation and application), D62 (externalities)

1   Introduction

In the wake of the COVID-19 pandemic, countries have enacted lock down policies unprecedented in modern times; dramatic – and costly – policies impose *social distancing*. The suppression of COVID-19 through restricted interpersonal contacts is studied using epidemiological models such as the canonical susceptible-infected-recovered (SIR) model due to Kermack and McKendrick (1927) and further popularized by Anderson and May (1979).[1] Many research teams now develop quantitative SIR models (For a US example, see Atkeson, 2020), each team adding its own granular details such as age and sector specific network or virus reproduction patterns. Yet, these models turn out dependent on many details, and it becomes increasingly difficult to check robust validity of results,[2] or to discern the general insights among the many detailed policy suggestions. Complementary to the sophisticated simulations, there are questions of more generic nature that need answers.

Are the large economic costs justified, even if these exceed the health costs by large margin? Is it likely that economic damages are exacerbated by employees that seek to protect themselves beyond required levels of social distancing? When jobs have different contact intensities, how much can we gain from differentiated regulation? Can a tighter policy for groups at risk, such as elderly, be used to relax social distancing for others? Conversely, can those with lower risk of transmission, such as children, be allowed a more relaxed regime? To answer these more general questions, I develop a stylized SIR model and derive simple rules that characterize an optimal social distancing strategy. Exploiting the low infection rates characterizing COVID-19 suppression strategies, I develop a version of the model that is mathematically tractable with closed-form solutions that provide intuitive and easy-to-understand results.

The model characterizes the optimal (efficient) policy in terms of primitive model parameters. The optimal rule for social distancing depends on the baseline (before intervention) reproduction rate R_0 and the share of diagnosed infections in the population, or its inverse, the share of asymptomatic infections. Necessary and sufficient conditions for herd immunity follow from the same fundamental parameters. My results also speak to the choice between herd immunity and COVID-19 suppression as policy strategies.

With the basic model set up and running, I can address the balance of costs between

---

[1]The original models also describe an exposure stage, which I leave out for convenience of the analysis. A recent extension of the SIR model with a very extensive set of stages has been used in Giordano et al. (2020) for study of the Italian COVID-19 outbreak.

[2]As case in point, some models point to very large gains of age-specific interventions, while other models find virtually no benefit thereof.

social distancing and infections. Stringent social distancing policies, required to keep the reproduction rate below unity, are costly and therefore bound to invite questions of relative efficiency. May the cure be worse than the ailment; aegrescit medendo?[3] I address those concerns in two ways: by studying the ratio between public benefits and individual benefits of social distancing (that is, by considering the externality value of social distancing), and by comparing the optimal balance between costs of social distancing versus health care costs.

Individuals have private reasons not to contract a disease, especially one that involves non-negligible mortality risks. Chen (2012) and Toxvaerd (2020) study the equilibrium paths of an infectious disease based on individually rational distancing norms. I complement those studies first by focusing on the social optimum, and second by adding an explicit comparison between private and public interests. The model provides sharp predictions with a clear intuition: the public benefits of optimal social distancing exceed the private benefits by an order of magnitude. The disease characteristics unambiguously call for public coordination through policies that are experienced as very tough at the individual level.[4] These same mechanisms are shown to imply very high aggregate costs of social distancing compared to the health care costs, in optimum.

I next extend the model with heterogeneous populations, allowing the analysis of targeted social distancing policies. Feng (2007), for example, studies disease dynamics for selective isolation policies. Fenichel (2013) shows the need for differentiated policies by health group. Berger et al. (2020) focus on the use of test outcomes for differentiated interventions. I complement these studies by presenting general conditions that inform us about the potential gains of differentiated policies in an optimal suppression strategy along three dimensions. First, I consider heterogeneous social preferences, describing differences between economic sectors – think of bars and sport matches where many people gather in close proximity – but also differences between outgoing and withdrawn people. Second, I consider heterogeneous morbidity and mortality risks. The model confirms the intuition that vulnerable populations, such as the elderly, should follow more stringent social distancing interventions. Somewhat surprisingly, the analysis also indicates that while the young generations benefit form a relaxed social distancing, aggregate social costs of social distancing increase under a policy tailor-made for health heterogeneity (i.e. that shields the elderly). Gains mainly come from lower morbidity and mortality of the elderly, an important complementary insight to other recent studies (cf Acemoglu et al.,

---

[3]Keogh-Brown et al. (2010) addresses these questions in a context different from COVID-19.

[4]We can also state the results the other way around. When observing individuals who over-comply, in many cases such is non-rational.

2020). Third, I consider heterogeneous transmission intensity, which I find to be the most promising source for reducing total social distancing costs.

My analysis focuses on a stable suppression strategy, in which the COVID-19 infection rate is kept low. I do not consider eradication, nor do I study herd immunity as a target. My focus is the central case. The former becomes viable when sufficient testing capacity and a vaccine or cure are available. The latter becomes relevant when a long delay is expected before a vaccine or cure is to arrive. I treat the lock down policy as part of the transitionary dynamics, towards a suppression policy. Though I do not go into lock down details, in the appendix I provide some rough lower bound for its costs.

2 Model

2.1 Set up

Standard SIR models cannot be solved in closed form, which is why the literature typically relies on numerical methods (cf Atkeson, 2020). I follow a different approach. Building on the canonical SIR framework, I derive closed-form solutions by making some additional simplifying assumptions that, at the early stages of an epidemic outbreak, are approximately valid. This approach is similar to (Gonzalez-Eiras et al., 2020), though I simplify along different dimensions. The main observation validating my approximation is that along an optimal path, the infection rate, compared to the hazard rate of finding a vaccine, cure, or cheap tests, is low. The implication is that I may consider the level of immunity as evolving sufficiently slowly to be exogenous.

The model is in continuous time, with a month as the unit of time. I normalize the population size to unity. Population is divided in two groups: those who have not been diagnosed with the disease, labeled by superscript A, and those who have, labeled by superscript B. Note that group A includes those who have not been infected yet, plus asymptomatic patients. A share s_1^B of the population has been diagnosed with the disease and has not yet recovered, while s_2^B have been reported as recovered. Of the remainder of the population, s^A, some are also infected and infectious (s_1^A) or recovered (s_2^A), but they are not diagnosed, that is, they are not observationally different from those who have not contracted the disease and are still vulnerable (s_0^A).

Each time, a flow x^B of patients is detected as new cases of infection, while a flow x^A of infections goes undetected. Of those infected, s_1^A, s_1^B, a flow y^A, y^B recovers. Figure 1 presents the flow diagram.

$$s_0^A$$

$$x_A \qquad x_B$$

$$s_1^A \qquad s_1^B$$
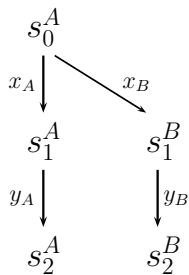
$$y_A \qquad \qquad y_B$$

$$s_2^A \qquad s_2^B$$

Figure 1: Flow diagram for COVID-19

Absent social distancing, the typical person interacts with n_0 other people, in which case infections spread according to:

$$x^A + x^B = \psi s_1^A s_0^A n_0, \tag{1}$$

where ψ is the infection intensity or contagiousness, related but not equal to the reproduction rate in a way explained below. The product of the two population fractions is the probability that an infected person meets a susceptible person.

Of those infected, a share θ is diagnosed and isolated to prevent further infections. The remaining share $1 - \theta$ are asymptomatic; they go unobserved and spread the disease through their social network. I assume a recovery rate of ρ

$$y^i = \rho s_1^i, \tag{2}$$

which, inversely, means an expected duration of the infection of $1/\rho$:[5]

In the original SIR model, one assumes that those recovered have gained immunity, though for a general model, I may add that recovered persons loose immunity after some time, at rate α:

$$\dot{s}_2^i = \rho s_1^i - \alpha s_2^i. \tag{3}$$

At this stage, I consider s_2^i and α to be very small and negligible, $s_2^i, \alpha = 0$.

---

[5]For interpretation: when the average infected person is infectious for a week, one has $\rho = 4$.

2.2   Early phase, prior to policies

In the early phase of the disease, before social distancing policies are implemented, $n = n_0$ and $s_0^A \approx 1$, so infections are given by:[6]

$$x^A = (1 - \theta)\psi s_1^A n \tag{4}$$

$$x^B = \theta \psi s_1^A n \tag{5}$$

$$\dot{s}_1^i = x^i - \rho s_1^i. \tag{6}$$

Note that θ, the share of diagnosed infections, is relied on as a measure for the share of infected patients who are sufficiently isolated to prevent the spread of the disease. For example, all those who are quarantined in response to symptoms are included in θ. On the other hand, those who are infectious but have not developed symptoms and take no special measures to prevent infection are included in $1 - \theta$.

The two groups A and B exist in a fixed proportion, which satisfies:

$$\frac{s_1^A}{s_1^B} = \frac{1 - \theta}{\theta}. \tag{7}$$

By writing out the equation for $\dot{s}_1^A$, one can see that for social distancing summarized by n, the virus spreads exponentially with growth rate:

$$g = (1 - \theta)\psi n - \rho. \tag{8}$$

The individual reproduction rate R is given by the contagiousness per infected person ψn, multiplied by the expected duration $1/\rho$:

$$R = (1 - \theta)\psi n/\rho. \tag{9}$$

These two equations allow me to calibrate the parameters ψ and ρ on data for the initial spread of COVID-19. There is a baseline (before-intervention) reproduction of about $R_0 = 3$ (see Fig. 4 in Pan et al. (2020) and Table 1 in Liu et al. (2020))[7] and a doubling of the number of cases every two days and a half,[8] which yields a per month growth rate of

---

[6]I write n rather than n_0 to introduce a general notation that remains valid for the analysis under social distancing.

[7]But Mizumoto et al. (2020a) estimate $R_0 = 5$ for the early phase of the Wuhan outbreak. For comparison, MERS and Ebola have much lower values for R_0, and much higher mortality rates.

[8]This is the slope of the graphs that present number of infections over time, in log-scale. See for example www.ourworldindata.org.

$g_0 = \ln(2) \cdot 30/2.5 = 8.5$. Normalizing the without-intervention social distancing to $n_0 = 1$ and assuming that in the first phase of COVID-19 the infected population did not take any special measure to prevent its spreading, $\theta_0 = 0$, I can derive the parameters $\rho = 4.3$, $\psi = 12.8$, suggesting that an average patient infects the next in slightly less than a week ($1/\rho$ months).[9]

2.3   Reducing secondary infections by asymptomatic carriers

The parameter θ measures what percentage of secondary infections can be prevented by isolating potential carriers. An obvious policy measure to contain the disease is self-isolation immediately after the onset of COVID-19 symptoms, though this measure is not without its own problems. Mizumoto et al. (2020b) and Nishiura et al. (2020) find that about 20 to 40 per cent (I set $\phi = 0.3$) of COVID-19 carriers does not develop symptoms even though they are infectious. He et al. (2020) calculate that, conditional on developing symptoms at a later stage, about 40 per cent (I set $\xi = 0.4$) of virus shedding occurs before the onset of symptoms (He et al., 2020). Thus, a perfect self-isolation strategy will raise the share of prevented infections to $\theta^* = (1 - \phi)(1 - \xi) = 0.42$.

Strict self-isolation in case of symptoms has costs, economic as well as in other social dimensions. It also implies many false positives, where patients self-isolate who have another (innocent) disease. Stringent self-isolation after self-diagnosis may additionally be unrealistic for open societies, so that I consider the above θ^* an upper bound for the potential of self-diagnosing voluntary policies.

A further tightening of policy can be achieved by letting a share ζ of the population track its contacts, combined with the rule that all contacts of the previous 3 days must be isolated immediately after an individual develops symptoms. Or, as an alternative, when the infection rate is low, new cases can be interviewed to identify contacts.[10] But even if such a scheme is carried out perfectly, it will not prevent every third-order infection. To be precise, a non-negligible part of third-order infections occurs before the onset of symptoms for the primary infected patient. Hence, some of the secondary infected contacts will already have caused third-order infections before alarmed.

To calculate the potential of a track and alarm policy, I walk along various cases. For a symptomatic primary infected individual, immediate self isolation reduces secondary

---

[9]This is consistent with various medical studies. Patients reach their maximal virus load (contagiousness) at the onset of symptoms (To et al., 2020, Fig 2), which is on average 5 days after infection (Chang et al., 2020). Yet virus shedding is asymmetrically distributed, with the average secondary infection appearing after symptoms evolve (He et al., 2020; To et al., 2020).

[10]This would also be useful for mapping the virus spread.

infections by $1 - \xi$. Of the ξ secondary infected people, a share $1 - \phi$ will be symptomatic, and by putting them in isolation, the decrease in third-order infections associated with these equals $\xi(1 - \phi)\mu\xi$, where μ is the share of pre-symptom infections prevented through early isolation. Of the ξ secondary infected people, a share ϕ will be asymptomatic, and by putting them in isolation, the decrease in third-order infections associated with these equals $\xi\phi(1 - \mu\xi)$. The total reduced infections thus amount to $\zeta^2\xi(\phi + \mu\xi - 2\mu\phi\xi) = 0.176\zeta^2$, for $\mu = 0.5$. However, voluntary schemes tend not to achieve large significant participation, so the term ζ^2 quickly reduces the contribution of track-and-isolate policies. It seems that the policy only contributes to containment if made mandatory, as in China and South Korea. In these case, the maximum achievable reduction in infections is achieved when $\zeta = 1$, increasing the potential from isolation to $\theta^{**} = 0.60$.[11]

The general idea emerging from the above discussion is that an optimal social distancing strategy balances the costs of increasing θ, the share of diagnosed infections, with the costs of reducing contacts n, the number of social interactions. In societies with strong concerns for privacy and a preference for voluntary isolation, much lower values of θ can be achieved compared to societies with a more centralist data-driven approach and enforced isolation policies. In the latter type of society, θ may reach values above one half. In the remainder this paper, I will assume $\theta = 0.3$.

2.4   Social distancing

An optimal policy plan balances the costs of early containment and the future benefits of fewer infections. The objective function reads:

$$\min \int_0^\infty e^{-\varepsilon t} \left[C(n(t)) + D(s_1^B(t)) \right] \mathrm{d}t, \tag{10}$$

where $C(n)$ measures the social costs of distancing, $D(s_1^B)$ measures the health care costs including increased mortality of those who do not recover, and ε measures the probability of finding a cure, vaccine, or any other sources for time discounting.[12] I assume $C(n_0) = C'(n_0) = 0, C' \leq 0, C'' > 0$ and $D(0) = 0, D' > 0, D'' \geq 0$.

---

[11]Indeed, the experience in South Korea suggests that a tight track and trace policy can substantially reduce the need for social distancing, but cannot be used as full substitute, $(1 - \theta^{**})R_0 > 1$.

[12]Note that time preference rates are supposedly small compared to the hazard rate for finding a vaccine or cure.

Writing down the Lagrangean for the planner's optimization, I obtain:

$$\mathcal{L} = \min \int_0^\infty e^{-\varepsilon t} \left[C(n) + D(s_1^B) \right] \mathrm{d}t \tag{11}$$

$$+ \lambda \left[\psi(1-\theta)n - \rho \right] s_1^B, \tag{12}$$

where $n(t)$ is the control variable, $s_1^B(t)$ the state variable, and $\lambda(t)$ is the co-state variable. The FOCs are

$$C' = -\psi(1-\theta)\lambda s_1^B, \tag{13}$$

$$\dot{\lambda} = (\varepsilon + \rho - \psi(1-\theta)n)\lambda - D'. \tag{14}$$

The optimal paths are saddle point stable, see the appendix for a brief discussion.

Proposition 1 (Optimal suppression). *The optimal social distancing path is monotonic: for low infection rates it gradually builds up towards the steady state. For infection rates above the steady state, social distancing starts aggressively (lock down, $R < 1$) and relaxes over time.*

The steady state social distancing and treated infectious people are characterized through

$$n^* = \frac{\rho}{(1-\theta)\psi} \tag{15}$$

$$D'(s_1^{B*})s_1^{B*} = -\frac{\varepsilon C'(n^*)}{\psi(1-\theta)} \tag{16}$$

The two equations above contain two distinct key messages about COVID-19. The regulator-induced norm n^* follows directly from substitution of $g^* = 0$ in (8) or $R^* = 1$ in (9). It tells us that containment of the virus requires a reproduction rate equal to or below one, $R \leq 1$, which is equivalent to $g \leq 0$. The social distancing effort is *independent* of the costs of the disease $D(.)$ or its intensity s_1^{B*}. Diseases that are more infectious (large ψ), diseases for which the a patient is infectious for longer (low ρ), and diseases for which a larger share of people does not observe its own infection (low θ) all require more stringent social distancing to achieve containment. For parameter values as discussed above, the required social distancing reduces social contacts n by slightly above 50 per cent from the baseline n_0.

We can rewrite social distancing (15) in terms of initially observed variables and interpret required social distancing in terms of the virus' reproduction rate: isolation of

symptomatic cases plus social distancing must exactly offset the baseline reproduction.

$$n^* = \frac{n_0}{(1-\theta)R_0}. \tag{17}$$

For any other social distancing, the virus develops exponentially. This is a fundamental property that we state as result:

Result 1 (sufficient social distancing). *The level of social distancing that suppresses the virus n^* is independent of the (stationary) stock of infected population s_1, or its costs $D(.)$. For $n > n^*$, the disease spreads exponentially. For $n < n^*$, the disease declines exponentially. For $n = n^*$, new infections are stable.*

While social distancing n^* contains the virus, it does not speak to the (stationary) level of infections x_1^i. That is described in the second equation of the proposition, which compares marginal costs of social distancing C' and the costs of health $D'(.)s_1^B$. Intuitively, a suppression strategy reduces the epidemic to a set of chains. Each infected person is a link in the chain, who on average infects one other person, the next link. The first result, $R = 1$, then means that chains do not divide. The length of the chains is the expected duration of the epidemic in terms of the length of the sequence of infections before a vaccine or cure is found. The flow of infection (proportional to s_1), equals the number of chains; it can be reduced through a longer lock down before the suppression strategy commences. This understanding of the epidemic says that optimal planning is mostly about the trade off between an extra current effort to cut back on the number of chains, as measured by C', versus the gains of reduced future health costs, where D' measures the marginal health benefits, and s_1^B captures the feature that the number of infections increases proportional to the current flow of infections, if suppression is temporarily relaxed. The longer it takes for a vaccine to arrive (lower ε), or the more contagious the virus is, (larger $\psi(1-\theta)$), the larger the benefits of a cut back on the infection flow. For calibrated parameters, we find for the ratio $\psi(1-\theta)/\varepsilon = 108$.

Result 2 (relative costs of social distancing versus health). *On the optimal suppression path, costs of social distancing exceed health care costs by a large margin.*

$$\frac{-C'(n^*)}{D'(s_1^{B*})s_1^{B*}} = \frac{\psi(1-\theta)}{\varepsilon}. \tag{18}$$

2.5   COVID-19 as an externality

Health care capacity constraints constituted a major motivation for stringent lock down measures. However, capacity constraints are largely due to imperfect planning, to social distancing policies coming too late. One should therefore not interpret capacity constraints as a main part of health care costs along an optimal suppression trajectory. Instead, individual costs such as the expected loss of life years associated with increased mortality likely make up the largest part of costs. And since costs of attracting COVID-19 are mostly private, aggregate costs are approximately linear in the number of patients, so I assume $D(s_1^B) = \delta s^B$. Conditional on contracting and being diagnosed with COVID-19, the expected loss of life years is in the order of magnitude of one or a few months. Adding personal health inconveniences and loss of labor days, I conservatively set the costs at three times monthly GDP. Private costs of contracting the disease are given by $\delta/\rho = D'/\rho$.[13] I thus set $\delta = 3\rho = 12.8$.

Despite the dominance of private costs, from a policy perspective the social distancing interventions are a genuine public good, infection being a public bad. These social costs are captured through the dual variable λ. In steady state, the social costs amount to $\lambda^* = D'/\varepsilon$. Private costs equal δ/ρ. A conservative estimate, abstracting from public costs associated with capacity constraints in hospitals, then assumes linear health costs, $D' = \delta$, and I find:

Result 3 (measure of externality). *Public benefits of social distancing exceed private benefits by magnitude ρ/ε.*

Coming back to the image above, where a stable epidemic is compared to a set of chains, I note that the ratio measures their length. It is the expected number of future patients indirectly infected by a current patient. If an infected person on average infects the next person after a week, and if in expectations the epidemic lasts for one year, the ratio between public and private costs is higher than fifty! A laissez fair equilibrium where individual interests drive dynamics as in (Toxvaerd, 2020; Chen, 2012) is therefore far from optimal. On the other hand, it is by no means rational for an individual to go beyond official social distancing guidelines out of concerns for its own safety. Citizens who do so may overestimate the probabilities of infection. Regulated social distancing makes individuals contribute to a public good.

---

[13]To see why I divide by ρ, note that $D(.)$ describes the costs per unit of time for having one infectious person throughout that period.

2.6   Group immunity

In the longer term, group immunity can arise and social distancing relaxed.

Result 4 (relaxing social distancing given group immunity). *The social norm that keeps a stable population of infected people in the model with evolving group immunity is given by*

$$n^* = \frac{\rho}{(1-\theta)\psi}\frac{1}{s_0^A} = \frac{n_0}{(1-\theta)R_0}\frac{1}{s_0^A} \tag{19}$$

The last part of the equation has a second message. Society can return to 'normal life', $n^* = n_0$, when the population reaches an immunity share of about 60 per cent, $s_1^A = 1 - n^*$. Before such levels are reached, social distancing is relaxed very slowly. If the infection rates x^i are kept low so as not to exceed the health care carrying capacity, the strategy also means that the share s_0^A declines only very slowly.

The model allows for an assessment of herd immunity as policy target, as we can derive a lower bound for its costs. A share $1 - n^*$ of the population must have been infected, of which a share θ has experienced substantial health costs. Cumulative health costs then amount to $\delta\theta(1 - n^*)$. One can compare these with the expected costs of a suppression strategy $(C(n^*) + D(s_1^{B*})/\varepsilon$.

Result 5 (suppression vs herd immunity). *Targeting herd immunity is more efficient compared to a suppression if the expected time before a vaccine or cure is discovered ($1/\varepsilon$) is sufficiently long, and/or if health costs δ are very low, and/or social distancing costs are very high.*

3   Differentiated interventions for heterogeneous populations

Within-group infection dynamics play a determining role in the spreading pattern of a disease. Kuchler et al. (2020) find empirical evidence that spatial heterogeneity of social contacts correlate positively with the spread of COVID-19. My model is consistent with such patterns. When a closed group j has a high level of social contacts, n_0^j, it supports a high internal infection growth rate and the laws of exponential growth dictate that its growth rate eventually dominates those of other groups. A description of such group dynamics can be found in Goyal and Vigier (2015); Galeotti and Rogers (2013).[14]

---

[14]These studies consider immunization as a major strategy, which is still not feasible for the COVID-19 case.

In this section, my aim is to generate sharp insights from a focused perspective. I consider a heterogeneous population under full mixing. Assuming one can observe individuals' characteristics, how does optimal policy fine-tune social distancing rules?

I study three dimensions of heterogeneity. First, people may differ with respect to their demand for social contacts, or the cost of distancing. This case can be associated with people working in different sectors, where employees in physical services tend to have more physical contact with clients, compared to employees in e.g. information technology. The second dimension relates to heterogeneous health care costs. For example, COVID-19 is particularly lethal for the elderly. What does that imply for optimal strategies? Third, I consider heterogeneity with respect to transmission, the ψ^i in our model. Children are infected less frequently, have fewer symptoms, are 70 per cent less susceptible to infection, while people older than 65 are 50 per cent more susceptible (Zhang et al., 2020). How does such heterogeneity affect optimal intervention?

3.1   Heterogeneous contact preferences

Consider a heterogeneous population, where each individual i enjoys baseline social contacts n_0^i, with the average normalized to unity, $\mathbb{E}n_0^i = 1$. Furthermore, costs of social distancing also vary between individuals, described through parameters γ^i. For the sake of a concise analysis, I assume that both the depth and breadth of social preferences are perfectly correlated: $n_0^i/\gamma^i = n_0/\gamma$, where the parameters without individual superscripts indicate averages (slightly abusing notation, I use the expectations operator $\mathbb{E}$ for averages, or omit the individual superscripts). Basically the assumption requires that all individuals value their marginal contacts equally when a uniform reduction is imposed: $C'(n^i) = C'(n^j)$ iff $n^i/n_0^i = n^j/n_0^j$. If some individual i values social contacts twice as much as another individual j, then s/he will enjoy twice as many contacts if unconstrained, and s/he is willing to pay twice as much to prevent an intervention that is costly for her/him. Furthermore, I assume the population perfectly mixes various types, meaning that the relative probability of two types i and j to have contact is given by $n^i n^j/\mathbb{E}n^k$. What matters for infection are the contacts multiplied by probability of infection. The dynamics are then described through:

$$\dot{s}_1^i = (1-\theta)\psi n^i \frac{\mathbb{E}[n^k s_1^k]}{n} - \rho s_1^i \tag{20}$$

$$\dot{s}_1^A = (1-\theta)\psi \mathbb{E}[n^k s_1^k] - \rho s_1^A \tag{21}$$

$$\dot{s}_1^B = \theta\psi \mathbb{E}[n^k s_1^k] - \rho s_1^B \tag{22}$$

where I defined $s_1^A \equiv \mathbb{E}[s_1^k]$, and $n \equiv \mathbb{E}[n^k]$. In the first equation, the term with expectations is independent of individual i's characteristics, so that the infection rate is proportional to contacts:

$$\frac{s_1^i}{s_1^j} = \frac{n^i}{n^j}. \tag{23}$$

This implies that those with the highest risk of infection also tend to spread the virus more frequently. Variety in contacts amplifies the spread of the virus!

$$\mathbb{E}[n^k s_1^k] = (1 + \bar{\sigma}_n^2) n^A s_1^A, \tag{24}$$

where $\bar{\sigma}_n = \sigma_n / \mathbb{E}n$; the bar on top of the standard deviation indicates I consider the normalized spread. The above equation revises the virus growth equation (8) and reproduction (9), leading to the following result.

Result 6 (heterogeneous social distancing). *The virus grows faster with heterogeneous social contacts, $\sigma_n > 0$, as compared to a population with the same average social contacts that is uniformly distributed:*

$$g = (1 - \theta)\psi(1 + \bar{\sigma}_n^2)n - \rho, \tag{25}$$

$$R = (1 - \theta)\psi(1 + \bar{\sigma}_n^2)n/\rho. \tag{26}$$

The equations have an interpretation in terms of policies, but also in terms of calibration. With regard to policies, it informs us that if the population has a diverse contact pattern, a more stringent policy is required to maintain the virus spread. The calibration interpretation is more benign. If initial estimates of the disease parameters ψ and ρ are based on an idealized homogeneous population, then the average reduction in social contacts to achieve suppression is less severe if one reduces the spread.

Result 7. *Optimal differentiated social distancing reduces the spread of social contacts if these originate from social contact preferences, also in relative terms, compared to the (before-intervention) baseline. That is,*

$$\frac{\sigma_{n^*}}{n^*} < \frac{\sigma_{n_0}}{n_0} \tag{27}$$

The result suggests that jobs requiring many contacts are suppressed more than proportionally in a virus containment strategy. The result also suggest job-rotating as an

efficient strategy for essential jobs that have high contact intensity. Shelving at the super market, or caring for many patients, is essential yet contact intensive. The aggregate R can be decreased if those jobs are shared among more employees. Taking a part-time job in the supermarket, or in health care, is a contribution to the overall suppression strategy. Surprising as this may sound, it has clear intuition. If a single job in the supermarket is shared among 5 people, the probability of becoming infected will decrease by factor 5 for each employee. The probability that a client is infected by the employee is than also reduced by a factor 5.[15]

3.2   Heterogeneous health costs

Now consider a population with heterogeneous health costs. Morbidity and mortality are strongly correlated with age (Ferguson et al., 2020). I now assume uniform social distancing costs γ and n_0, while allowing for heterogeneous costs of catching the disease through the term δ^i. To study the optimal diversified social distancing strategy, I look at the suppression steady state with average stock of infected $s_1^B = \mathbb{E}s_1^{Bi}$. The flow of costs amounts to

$$C + D = \int_i C^i(n^i) + \delta^i n^i s_1^B \mathrm{d}i, \tag{28}$$

where I substituted the equilibrium property that infections are proportional to the number of social contacts n^i.[16] A suppression strategy maintains the conditions $g \leq 0$ or $R \leq 1$, which means that it should satisfy

$$(1 - \theta)\psi \int_i (n^i)^2 = \rho n \tag{29}$$

Given the objective function (28) and constraint (29), I easily find the FOC determining optimally varied distancing:

$$n^i = \frac{-C_i' + \rho\nu - \delta^i s_1}{(1 - \theta)\psi 2} \tag{30}$$

with ν the dual variable associated with (29). In words, elderly with higher health risks δ^i will require more social distancing; n^{*i} decreases in δ^i.[17] Less vulnerable young people

---

[15]I am indebted to Jeffrey Campbell for suggesting this interpretation of the result.

[16]That is, $ns_1^{Bi} = n^i s_1^B$.

[17]Technically, this is the property of the fixed point.

(lower δ^i) enjoy less social distancing (higher n^{*i}).

An optimally differentiated social distancing intervention is obviously of great benefit. Yet the analysis also points to clear limits to the benefits. Importantly, a differentiated social distancing policy provides no overall relaxation of the social distancing policies! By imposing a varied policy, the average social distance must decrease. That is, aggregate social distancing costs will rise, but these costs fall on the vulnerable population, while the younger generations gain. I summarize these findings in the following result:

Result 8 (social distancing along health characteristics). *Optimal differentiated social distancing based on health characteristics sets tighter distancing policies for elderly, while the younger are allowed a more relaxed social distancing. Average social distancing tightens ($n^* \searrow$) and aggregate social costs of differentiated social distancing increase ($C^* \nearrow$). Social benefits of differentiated social distancing are due to reduced morbidity and mortality of the vulnerable population.*

The results are expressed in terms of social costs, and cannot be carried over one-to-one to economic output. The increased social costs of the elderly are measured through a reduction in GDP; yet benefits from the reduced social distancing policies for the young will accrue to the economy.

As the differentiated policy substantially reduces average morbidity and mortality, and thus the health care costs of a given level of infections, compared to the uniform policies, it allows for an increase in the level of infections s_1^{B*}. In turn, immunity rates will rise faster, especially among the younger population, and herd immunity could be reached at an earlier stage, leading to a relaxation of social distancing. That is, economic benefits are potentially large, but mostly in the longer term, and only if good shielding is possible.[18]

3.3   Heterogeneous transmission

Lastly, I consider heterogeneity in ψ^i, the transmission of the virus, possibly related to age (Zhang et al., 2020; Vogel and Couzin-Frankel, 2020).[19] For convenience of the analysis, I assume symmetry in the probability of being infected and infecting, so I can substitute the spread of $\psi^i n^i$ for the spread of n^i in the analysis for heterogeneous n^i. That is, I

---

[18]This seems to be cause for the large gains calculated by (Acemoglu et al., 2020).

[19]For lack of clear evidence, I leave out the argument in favor of outside activities that are believed to induce fewer infections compared to in-home activities.

establish the equivalence of (25)-(26):

$$g = (1 - \theta)\psi(1 + \overline{\sigma}_{\psi n}^2)n - \rho, \tag{31}$$

$$R = (1 - \theta)\psi(1 + \overline{\sigma}_{\psi n}^2)n/\rho. \tag{32}$$

The equations imply that one can increase social contacts for those with a low transmission coefficient ψ^i, while reducing the standard deviation of the product $\psi^i n^i$. This gives:

Result 9 (relaxed social distancing for those with low transmission). *Optimal differentiated social distancing based on transmission characteristics sets looser distancing policies for children who have lower transmission coefficient ψ^i and tighter policies for elderly with higher ψ^i. Average social distancing is relaxed ($n^* \nearrow$) and aggregate social costs of differentiated social distancing decrease ($C^* \searrow$).*

Different from stratification along health characteristics, a differentiated policy along transmission characteristics allows for a relaxation of average social distancing as it lowers the average transmission intensity! I thus expect potentially large gains. The policy recommendation, in terms of age-related policies, is similar to that born out of differentiated health costs, but the underlying mechanism is radically different and so is the implied average social distancing.

When various heterogeneous characteristics interact, efficient policy making becomes more complex. Zhang et al. (2020) also estimate that children tend to have more contacts (higher n_0^i) when, for example, schools open, which may offset the lower transmission (ψ^i). Under such interaction, the efficient strategy is to keep social distancing for children as well, but to allow them more contacts compared to others, and thus count for a substantial reduction in aggregate social costs of distancing.

4   Quantitative illustration

4.1   Optimal Suppression

For the costs of social distancing, I assume a polynomial cost function of degree $\eta > 1$, $C(n)\gamma(1 - n/n_0)^\eta$, which satisfies $C(n_0) = C'(n_0) = 0$. I do not have much evidence on the costs of social distancing, but it seems reasonable to assume that, after an initial phase of high costs because lack of experience, costs of a maximal lock down range between ten and twenty percent of GDP, which is our unit of measurement; here I assume $\gamma = 0.2$ and strongly convex costs, $\eta = 3$. Calibrating all parameters results in Table 1.

Table 1: Parameters and Targets

Parameter	Description	Value	Evaluation / Target
γ	social distancing costs	0.2	loss of income
δ	health costs	12.8	costs of labor loss and mortality
η	convexity of social distancing costs	3	
ε	hazard rate for finding vaccination	0.08	optimist scenario (12 months)
ρ	recovery rate / inverse lag between infections	4.3	initial spread of disease
θ	share of diagnosed infections	0.3	
ψ	contagiousness of virus	12.8	initial spread of disease
n_0	natural social network size	1	normalized

Optimal policy outcomes are listed in Table 2. Social distancing is determined by the necessary conditions to constrain the spread. Given a baseline (before intervention) reproduction of 3, and given that diagnosis and subsequent isolation of part of the population already reduces secondary infections, $\theta = 0.3$, I find that halving the amount of social contacts suffices, $n^* = 0.48$. Along the optimal path infections only spread slowly; about 0.2 per cent of the population is infected per month. The costs of social distancing are substantial, about 3 per cent of GDP in our calibration, while associated health care costs are only 0.15 per cent, consistent with Result 2.

Table 2: The Optimal Suppression Strategy

Variable	Description	Value	(annual base)
n^*	optimal social distancing	0.48	
$\rho s_1^{B*}/\theta$	flow of new infections	0.0018	
$C(n^*)$	social distancing costs (ppp.month)	0.029	
$D(s_1^{B*})$	health costs (ppp.month)	0.0015	
$(C(n^*) + D(s_1^{B*}))/\varepsilon$	expected cumulative suppression costs	0.37	0.03
$1 - s_1^{A*}$	herd immunity	0.52	
$D(\theta(1 - s_1^{A*}))$	herd immunity health costs	2.0	0.17

The model can also be used to provide some suggestive calculations on the costs of a lock down, required if the policy response is late. Expected cumulative costs depends on the currently infected population s_1^A, or equivalently s_1^B. In the appendix, I show how the model can be used to estimate peak infection rate at the start of the lock down for a sample of countries, based on excess mortality rates, and subsequently, lock-down costs.

5 Conclusion

I used a standard dynamic optimization model that integrates SIR disease spread patterns, social distancing, and heterogeneous populations. The suppression strategy that aims at low infection rates allows me to derive robust general insights that complement more specific policy studies based on simulations in elaborate SIR models.

First, an optimal policy is monotonous, meaning that anticipation of an outbreak induces early social distancing that gradually ramps up over time. An overshooting of infections, and subsequent stringent lock downs, relaxed over time, as observed in most countries is the result of imperfect anticipation.

Second, once the epidemic is stabilized, the stringency of required social distancing does not depend on the number of infections. Specifically, keeping a lower rate of infections does not imply a reduced need for social distancing. The model provides no support for back-to-normal policy, e.g. densely populated social (sports) events after the lock down. While indeed lock downs target a reproduction R strictly below one, and are thus more stringent and costly than a suppression strategy, the latter must maintain the conditions for avoiding the reproduction to rise above one. Over time, when immunity rises, social distancing can be relaxed but to target herd immunity could be a costly objective.

Third, in the optimum, the costs of social distancing substantially exceed the health costs of those infected, including the value of lost labour time and increased mortality. Related, the social benefits of social distancing substantially exceed private benefits. Informing the public and leaving social distancing to be determined by individual trade offs is an unfeasible strategy. The characteristics of COVID-19 explain this finding. Its exponential spread makes it efficient to bear high costs of social distancing, above the individual benefits, to keep future health care costs low.

Fourth, a heterogenous population demands diversified intervention. In general, a uniform distribution of social contacts leads to a reduced spread of the virus, compared to a dispersed distribution, for given average number of social contacts. This finding suggests rotation for contact-intensive jobs as a possible strategy and, in relative terms, more stringent social distancing policies for contact-intensive sectors such as restaurants and professional sport events. Heterogeneous vulnerability indisputably demands dedicated social distancing policies, where those more vulnerable are sheltered. However, such a strategy increases the overall burden of social distancing to society. The reduction in number of contacts by some group, say elderly, cannot be redistributed equally to the others, say the working age groups. Finally, exploiting differences in transmission intensity offers an opportunity for substantial gains. If the young tend not to contribute much to

the transmission of COVID-19, then they can relax social distancing rules.

I remained silent about the implementation of social distancing rules, while clearly macroeconomics costs depend thereon. Given a required reduction in social contacts, the government must find the means to keep essential production afloat and support non-essential production as much as possible, in order to minimize social costs. Combining the above insights provides some room for maneuver. The numerical assessments based on central parameters found herd immunity to be 5 times as costly compared to suppression. If the elderly can be segregated and shielded effectively, health care costs could come down and the ratio might drop in favor of herd immunity for all but the older generations. Yet shielding the elderly should not be taken lightly, as quality time with the family may also increase in value with age.

References

Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. (2020). A multi-risk model with optimally targeted lockdown.

Anderson, R. M. and May, R. M. (1979). Population biology of infectious diseases: Part i. *Nature*, 280(5721):361–367.

Atkeson, A. (2020). What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios.

Berger, D. W., Herkenhoff, K. F., and Mongey, S. (2020). An seir infectious disease model with testing and conditional quarantine.

Chang, D., Mo, G., Yuan, X., Tao, Y., Peng, X., Wang, F., Xie, L., Sharma, L., Dela Cruz, C. S., and Qin, E. (2020). Time kinetics of viral clearance and resolution of symptoms in novel coronavirus infection. *American Journal of Respiratory and Critical Care Medicine*.

Chen, F. (2012). A mathematical analysis of public avoidance behavior during epidemics using game theory. *Journal of theoretical biology*, 302:18–28.

Feng, Z. (2007). Final and peak epidemic sizes for seir models with quarantine and isolation. *Mathematical Biosciences & Engineering*, 4(4):675.

Fenichel, E. P. (2013). Economic considerations for social distancing and behavioral based policies during an epidemic. *Journal of health economics*, 32(2):440–451.

Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z., Cuomo-Dannenburg, G., et al. (2020). Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand.

Galeotti, A. and Rogers, B. W. (2013). Strategic immunization and group structure. *American Economic Journal: Microeconomics*, 5(2):1–32.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., and Colaneri, M. (2020). Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, pages 1–6.

Gonzalez-Eiras, M., Niepelt, D., et al. (2020). On the optimal "lockdown" during an epidemic. *CESifo WP*, 8240.

Goyal, S. and Vigier, A. (2015). Interaction, protection and epidemics. *Journal of Public Economics*, 125:64–69.

He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, pages 1–4.

Keogh-Brown, M. R., Wren-Lewis, S., Edmunds, W. J., Beutels, P., and Smith, R. D. (2010). The possible macroeconomic impact on the uk of an influenza pandemic. *Health economics*, 19(11):1345–1360.

Kermack, W. O. and McKendrick, A. G. (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.

Kuchler, T., Russel, D., and Stroebel, J. (2020). The geographic spread of covid-19 correlates with structure of social networks as measured by facebook. *CESifo WP*, 8241.

Liu, Y., Gayle, A. A., Wilder-Smith, A., and Rocklöv, J. (2020). The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*.

Mizumoto, K., Kagaya, K., and Chowell, G. (2020a). Early epidemiological assessment of the transmission potential and virulence of coronavirus disease 2019 (covid-19) in wuhan city: China, january-february, 2020. *medRxiv*.

Mizumoto, K., Kagaya, K., Zarebski, A., and Chowell, G. (2020b). Estimating the asymptomatic ratio of 2019 novel coronavirus onboard the princess cruises ship, 2020. *medRxiv*.

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A. R., et al. (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19). *medRxiv*.

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., et al. (2020). Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, china. *Jama*.

To, K. K.-W., Tsang, O. T.-Y., Leung, W.-S., Tam, A. R., Wu, T.-C., Lung, D. C., Yip, C. C.-Y., Cai, J.-P., Chan, J. M.-C., Chik, T. S.-H., et al. (2020). Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by sars-cov-2: an observational cohort study. *The Lancet Infectious Diseases*.

Toxvaerd, F. (2020). Equilibrium social distancing. *Cambridge-INET Working Paper Series*, (2020/08).

Vogel, G. and Couzin-Frankel, J. (2020). Should schools reopen? kids' role in pandemic still a mystery. *Science*, doi:10.1126/science.abc6227.

Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., Wu, Q., Merler, S., Viboud, C., Vespignani, A., et al. (2020). Changes in contact patterns shape the dynamics of the covid-19 outbreak in china. *Science*.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The lancet*.

A   Appendix. Saddle point stability of optimum

For a phase diagram in (s_1^B, n)-space, the $\dot{s}_t^B$ locus is the horizontal line at $n = n^*$. The locus for $\dot{n} = 0$ is somewhat more complicated. There are three variables, (n, λ, s_1^B), and I

want to drop the λ. Thus I rewrite the above FOCs as

$$0 = C'' \dot{n} + \psi(1-\theta)\dot{\lambda}s_1^B + \psi(1-\theta)\lambda\dot{s}_1^B$$

$$= C'' \dot{n} + \varepsilon + \rho - \psi(1-\theta)n - \psi(1-\theta)D's_1^B - \frac{C'}{s_1^B}\dot{s}_1^B. \qquad (33)$$

It follows that if n is constant and one increases s_1^B, the term $D's_1^B$ goes up, while $\frac{\dot{s}_1^B}{s_1^B}$ remains constant, so $\dot{n}$ goes up. That is, independently of the slope of the locus for $\dot{n} = 0$ (down or up), to the right of the locus the optimal paths will move up, and to the left they move down.

B Appendix. Costs of a late start

The model can also be used to provide some suggestive calculations on the costs of a lock down, required if the policy response is late. Expected cumulative costs depends on the currently infected population s_1^A, or equivalently s_1^B. I thus estimate the peak infection rate at the start of the lock down for a sample of countries, based on excess mortality rates. I can control for different expected mortality rates for COVID-19 associated with different demographic characteristics of various countries, based on the finding that mortality of COVID-19 is about proportional to the *annual* mortality (Ferguson et al., 2020). I can then write

$$m_t = m_0(1 + 365x_{t-L}) \Rightarrow x_{t-L} = (\widetilde{m}_t - 1)/365 \qquad (34)$$

where m_t is the mortality rate, m_0 is the baseline mortality without COVID-19, $\widetilde{m} - 1$ is the excess mortality rate, x_{t-L} is the *daily* rate of infection about 3 weeks before the mortality peak (Zhou et al., 2020), and 365 converts the annual mortality rate into a daily infection flow. I compare, per week of year, deaths in 2020 with the number of deaths in 2019, and identify the peak excess mortality. In Table 3, I report for some countries the estimated peak infections, per thousand per day.[20] Denmark started its interventions early, and peaked at only 0.32 infections per thousand per day. Spain, on the other hand, was late and only started a (very strict) lock down when infections had reached a ten-fold higher level. The fifth column presents the additional costs associated with a lock down intervention that brings the infection rates down to the suppression

---

[20]A similar exercise for cities finds much higher relative mortality rates, e.g. 9.4 for Bergamo in week 11, 6.4 for New York City in week 15, and 4.4 for Madrid in week 14.

strategy level. Costs are measured in terms of monthly GDP loss, based on the optimal control model, with no frictions and perfect foresight about both the disease's dynamics and costs of interventions.[21] The last column presents the estimated decline in GDP in the first quarter of 2020, relative to the last quarter of 2019. Statistically predicted costs (column 6, per quarter) exceed the modeled costs (column 5, per month), which are based on an assumed balanced set of interventions. When the virus hit, policy makers had to come up with lock down measures under imperfect information. The comparison with model outcomes is unfair as the latter assumes some experience on social distancing has lead to learning about efficient interventions.

Table 3: Peak Infections and Additional Costs

Country	Week	Mortality	Infections	Lock down costs	2020Q1
Austria	15	1.14	0.39	0.043	0.025
Denmark	15	1.12	0.32	0.038	
France	14	1.57	1.56	0.090	0.058
Germany*	14	1.06	0.17	0.022	
Italy	12	2.67	4.58	0.146	0.047
Netherlands*	14	1.76	2.08	0.103	
Portugal	14	1.28	0.77	0.064	
Spain	14	2.09	2.98	0.121	0.052
Sweden	15	1.53	1.46	0.088	
Switzerland	15	1.37	0.99	0.073	
United Kingdom*	16	2.48	4.05	0.139	
US	15	1.18	0.50	0.050	0.012

Notes: Countries with a ∗ have their peak close to the latest reported week data. Mortality in that week relative to the same week the year before. Daily infections per thousand, see main text. Lock down costs additional to suppression strategy, relative to monthly GDP, from model calculations. 2020Q1 data from Eurostat and BEA, negative growth relative to 2019Q4.

---

[21]Calculations are consistent with the envelope theorem: λ equals the marginal costs of higher infection rates around the steady state. Additional costs are convex in the log of infections, but concave in the linear measure of infections.