Economic Theory (2024) 77:9–48
https://doi.org/10.1007/s00199-023-01533-w

RESEARCH ARTICLE

Static and dynamic inefficiencies in an optimizing model of epidemics

Pietro Garibaldi[1] · **Espen R. Moen**[2] · **Christopher A. Pissarides**[3]

Received: 30 October 2022 / Accepted: 23 August 2023 / Published online: 6 November 2023

Abstract
Several externalities arise when agents shield optimally to avoid infection during an epidemic. We classify externalities into static and dynamic and compare the decentralized and optimal solutions when agents derive utility from social interaction. For low infection costs agents shield too little; for high costs they shield too much because of a "rat race to shield": they delay social action until other agents contract the disease and society reaches herd immunity. Other externalities drive more wedges between the private and social outcomes. The expectation of a fully effective vaccine that ends the disease faster changes results, reversing excessive shielding.

Keywords SIR models · Matching model · COVID-19 · Social distancing · Rat race · Herd immunity

JEL Classification A12 · I10 · J18 · D61-D62

1 Introduction

In this paper we model the transitions in an optimizing forward-looking model of an epidemic, in the three-state SIR framework originally proposed by Kermack and

Research support from Collegio Carlo Alberto is gratefully acknowledged. We thank Per August Moen for excellent research assistance, and Espen Henriksen for advise on the computational methods. We thank online seminar participants at CEPR, Polytechnic of Torino, Collegio Carlo Alberto, Oslo Macro Group, LSE, University of Chicago Search Conference, Marseille, Norwegian Business School and the Frisch Centre. Moen gratefully acknowledges financial support from the Norwegian Research Council (Grant Number 82061).

✉ Christopher A. Pissarides
c.pissarides@lse.ac.uk

[1] Collegio Carlo Alberto, University of Torino, IZA and CEPR, Turin, Italy

[2] Norwegian Business School, CEPR, Oslo, Norway

[3] London School of Economics, University of Cyprus, IZA and CEPR, London, UK

McKendrick (1927). We focus on the consequences of individually optimal behaviour for the dynamics of the epidemic, and in particular the externalities that arise when private agents act in their self interest to shield themselves from the epidemic. Although our model can be applied to the study of externalities in COVID-19, it is not specific to this. It is motivated more generally by the SIR model of epidemics, in which infections follow "social" contact and carry some cost to the individual.[1]

Following our derivation of the individually optimal transition rates, we show how optimizing behaviour in the absence of policy influences outcomes. We then compare the decentralized optimizing model and a social planning solution, with the social planner having access to the same information set as private agents. We obtain striking contrasts, which we derive formally and illustrate with simulations. We derive four different types of externalities, which are classified below and are the main theme of this paper.

In our formulation agents can be in one of four states, but by applying a modelling trick we collapse the model to one that is close to the original three-state SIR model. We work in discrete time. In the first state a mass S_t of agents are healthy but "susceptible" to the disease; in the second state a mass I_t of agents are "infected" but without symptoms, and they can pass the disease on to susceptible individuals after social contact; in the third state the I_t infected individuals develop symptoms, which are costly in terms of lifetime utility but not time; they cross from the asymptomatic infected state in period t to recovery in period $t + 1$, by bearing a cost which is a fraction of their lifetime utility and may be interpreted as the probability of dying from the disease.

Our interest is in deriving the impact that agents have on the transitions across the SIR states, and particularly whether these impacts are socially optimal. Individuals respond in their own self interest, with rational expectations about the future. Our focus is on the transition from the susceptible to the asymptomatic infected state, which is influenced by contacts between the susceptible and the infected. The decisions are taken without information about their state (whether they are susceptible or asymptomatic infected) and with full knowledge of future transitions in the event of an infection. For simplicity of exposition, we call individuals who are in either the susceptible state or the asymptomatic infected state, *vulnerable*, so the mass of vulnerable individuals is $V_t = S_t + I_t$.

We borrow ideas from search and matching theory (Pissarides 2000) which we embed into a SIR framework in line with the solutions in the literature that followed Kermack and McKendrick's (1927) pioneering work. The main difference between our model and those in the early epidemiological literature is that agents in our model are able to reduce their probability of infection by avoiding contacts with vulnerable individuals. We model behaviour in such a way that this difference is picked up by a single variable x_t, which we call social activity. It is action that yields utility to the agent but in order to complete it, the agent needs to come into contact with other agents. For concreteness we can think of it as consumption, which involves shopping for goods and consuming them in the company of others. We normalize social action in the non-optimizing model by $x_t \equiv 1$ in all periods of life. In our optimizing model

[1] In Garibaldi et al. (2023) we survey pre-COVID and post COVID research with these types of models.

we derive an optimal $x_t < 1$ which varies over time, depending on incentives. When $x_t < 1$ we say that there is *social distancing* or *shielding*

The key contribution of the paper is our comparison between the decentralized solution and the solution chosen by the social planner. The two differ substantially. This is because private agents ignore two types of impacts of their actions on aggregate outcomes, which give rise to deviations from the optimal path. One of these impacts operates "within" a given point in time and the other "across" time. At a given point in time, individuals ignore the fact that when they take social action, other people may randomly come into contact with them and be infected. This may give rise to a static externality. We show that with increasing returns in the contact function, defined here as an elasticity of contacts with respect to social action greater than one, static externalities will *cet par* make the social planner recommend more social distancing. But with constant returns (unit elasticity), the static externalities are internalized. Across time, when private agents shield, they reduce the future infected population and increase the future susceptible population. These relative population changes alter the flows across states: the lower infected pool lowers future infections and lowers congestion in hospitals, whereas the higher susceptible pool slows down the adjustment to the end of the epidemic. Private agents ignore these secondary effects of their actions whereas the social planner takes them into account. We call these *dynamic externalities.*

By working in discrete time and deriving the optimal policies from Bellman equations, we are able to distinguish between three different kinds of dynamic externalities. The *contagion* externality is caused by the impact of private actions on the stock of infected individuals and the subsequent spread of the disease. The *medical congestion* externality is due to the fact that *medical treatments* are less good when there are more patients, given that medical resources and personnel are not immediately adaptable to changed circumstances. Finally, the *immunity* externality is caused by the impact of private actions on the stock of susceptible agents, which in turn influences the dynamics of the disease. These externalities interact with each other to produce deviations between the private and social outcomes that can go either way and can change in the course of the epidemic.

Consider first decisions made in the absence of a vaccine, when herd immunity is the only path to eradication of the disease. With a sufficiently high but not unreasonable cost of attracting the disease, our simulations show that the immunity externality may dominate the contagion externality, and agents in the decentralized solution shield more than in the optimal solution in the height of the pandemic. Because of forward-looking behaviour, agents know that eventually the disease will end, with a fraction of people who never experience the disease. Hence agents shield more than other considerations would imply, to increase the probability that they belong to the infection-free group when the disease is eradicated. This has the features of a rat race, and we refer to it as a *rat race to shield*. If, say, 60% of the people need to get the disease before it is eliminated, an efficient mechanism would be to allocate the 60% randomly across the population (in the absence of heterogeneity). But no individual wants to be one of those randomly selected. So instead of racing to be first as in the traditional rat race, here there is a racing to be the last, by shielding. How much inefficient shielding there is because of this rat race depends on the severity of medical costs. Our simulations illustrate the sensitivity of the model to plausible costs.

If a full vaccine is expected to arrive with sufficiently high probability, the herd immunity state becomes irrelevant because agents will become immune through vaccination rather than recovery. Still there is a "short-term" interval between the current period and the arrival of a vaccine (which we model as a Poisson process), during which the externalities are present. During this period, the static, medical and contagion externalities are still powerful. But he immunity externality by contrast is weakened substantially because its power lays in the fact that by choosing more infections today, there will be fewer infections in the more distant future. With a vaccine expected, this action is likely to cause more overall infections because the vaccine will arrive (in probability) and end the disease before the society can reap the reward of fewer infections later on. With a much weaker immunity externality, the optimal policy now requires more shielding than the decentralized solution, until the vaccine arrives.

The rest of the paper is organized as follows. Section 2 briefly discusses some related literature. Section 3 describes the model in more detail and derives the individual maximizing choices in a world in which no vaccine is expected to arrive, so that the epidemic ends through herd immunity. Section 4 derives the welfare maximizing choices of a central planner. Section 5 analyses the model when a vaccine is expected to arrive with some probability. Section 6 simulates the model highlighting the key role of the static and dynamic externalities in driving a wedge between the decentralized and social optima. The last section concludes.

2 Related literature

There has been a very large number of working papers by economists on epidemics since the outbreak of the COVID-19 pandemic in early 2020. Providing a survey of this literature is beyond the scope of this paper, but we put our paper into context and relate it to those papers that address epidemics with models that have common features with ours.[2] Although there is some overlap with other independently written papers, we believe that our classification of the full set of externalities that arise in optimizing models of epidemics, and the role of the immunity externality in particular, are new to this paper

We should first note that although most economists became interested in epidemiological models because of COVID-19, a small number of papers before COVID-19 addressed issues in epidemiology by making use of the key distinguishing feature of economic models, the change in agent behaviour in response to the disincentive of catching the disease. Most papers before COVID-19 modelled influenza epidemics and focused on a negative contact externality similar to our static externality: too little social distancing by self-interested agents that do not internalize the costs of transmission to others. Examples of papers in that tradition include Chen (2012), who like us obtains the externality by invoking increasing returns to scale in contacts, and Rowthorn and Toxvaerd (2012), who use a contact function with fixed proportions and a linear cost of disease prevention borne by the individual. Reluga (2010) introduced

[2] A partial survey of this literature is Garibaldi et al. (2023).

the idea of an aggregate disease transmission function that depends on endogenously determined social distancing. Quercioli and Smith (2006) go further and touch on one aspect of the dynamic externalities. In this research social distancing decisions depend on the state of the disease, very much along the lines of post-COVID research. Within a SIS variant of the model, Goenka and Lin (2012) integrate an epidemiological model with a macro growth model and show that the economy can feature both cycles and chaos.

Another disease that attracted the interest of economists is HIV. Kremer (1996) Geoffard and Philipson (1996) and Greenwood et al. (2019) are three key contributions. The framework in which these authors modelled behaviour was one of heterogeneous individuals, in terms of their aversion to risk, and the implications of multiple partners. An interesting feature of HIV, when contrasted with COVID-19 or influenza, is that HIV is contracted by a susceptible agent only after a voluntary decision to engage in sexual behaviour with a person who might be infected. In COVID-19 or influenza infections might arise through proximity without an intentional person-to-person contact. The implications of this difference for the externalities involved have not been studied, but it is clearly the case that the externalities listed in this paper would not be the same as in HIV epidemics. At a macro level, Chakraborty et al. (2010) propose a general equilibrium model of HIV infection transmission, prevention investment and rational behaviour.

Post-COVID-19 the economics literature exploded, working mainly with the SIR model proposed by Kermack and McKendrick (1927) or later variants (see Garibaldi et al. (2023) for details). Several papers focused on deriving optimal policies from a social welfare function that encapsulated the trade-off between the costs of the disease (usually deaths or restrictions on behaviour that require social interaction) and GDP losses from social distancing.[3] Unlike our paper, the objective of these papers is to derive directly optimal policies without commenting on whether the decentralized solution replicates them.

Another group of papers derives non-cooperative maximization policies for economies that face similar trade-offs between the costs of the disease to the individual and the costs of shielding. Some also derive the socially optimal solutions and make comparisons, so they are more directly comparable with our paper. In the case of homogeneous agents, Eichenbaum et al. (2021) solve for both the centralized and decentralized equilibrium, but unlike our paper, they do not derive analytical expressions for any externalities. They adopt a quadratic meeting technology in the spirit of Diamond and Maskin (1979), and infections come from social interaction during consumption activities.

A closer and simultaneous paper to our own is by Farboodi et al. (2021). Their setting is in continuous time, and the meeting technology is again quadratic. Their treatment of the static externality, which, as in our model, arises in their model because of the

[3] See, e.g., Alvarez et al. (2021) and Hritonenko and Yatsenko (2022) for homogeneous agents and Acemoglu et al. (2021), Favero et al. (2020) and Makris (2021) for heterogeneous agents. Within the same class of papers, Federico and Ferrari (2021) treat the transmission rate as diffusive stochastic state variables, Bosi et al. (2010) consider the role of altruistic behaviour in a SIS variant of the model, and La Torre et al. (2021) consider a planner that chooses not only the degree of social distancing but also that of therapeutic treatment.

increasing returns to scale in meetings, is similar to ours. But they do not disentangle the dynamic externalities, which does not enable them to identify the forces that might lead a social planner to command more social interaction early on in the course of the pandemic. Another paper in this strand, with homogeneous agents, is by Rachel (2020), who solves for the optimal lockdown and highlights the role of social policy in the presence of infection externalities. It has both a static and a dynamic dimension and derives ambiguous overall effects on the planner's choice. He derives his results from a less general model than in this paper, with only a two-state (high or low) social action, which does not enable him to give a full listing of the dynamic externalities. The same applies to Lebeau (2020) paper. Nævdal (2020) shows that there are increasing returns to scale to social distancing and to other means to control the epidemic, which are a cause of externalities.

Also, in this class of models, some papers model heterogeneous agents; for example, Alfaro et al. (2020) and Brotherhood et al. (2020). The latter consider optimal policies when there are two groups, young and old, and the old suffer medical costs from infections. Their focus is the role of testing and the interactions between young and old. They also consider the decentralized and centralized solution and provide quantitative measurement of the externalities.

Finally, mention should be made of several papers that focus on the labour market in isolation. Jones et al. (2020) explicitly link working from home with optimal containment policies. Kapicka and Rupert (2020) model the labour market dimensions of the pandemic, and solve for the centralized program that takes into account an infection externality similar to our contagion one, which arises because agents do not take into account that once infected, the probability that others will get infected increases. The labour market response to the COVID-19 pandemic within a search environment is also studied by Gregory et al. (2020), but without reference to the planner solution. The potential uncertainty about one' s infection status, which plays an important role in our paper, and the role of testing, are also studied by Berger et al. (2020) and von Thadden (2020).

3 Decentralized equilibrium in a model of epidemics

In this section we develop a model of transitions with forward-looking individual decision making and no vaccine against the disease. We first set up the basic epidemiological framework. Then we model individually optimal activity levels given an exogenous relationship between own activity level and the risk of attracting the disease. Next we define a contact technology that maps the activity levels of individuals into a transmission rate for the disease. Finally we define equilibrium.

3.1 Basic epidemiological framework

We work in discrete time and define the period to be the length of time that an infected person is asymptomatic. In terms of COVID-19, the length of the period is therefore about two weeks, although it could be longer. During this period infected individuals

are unaware that they are contagious, and we refer to them as infected. At the end of the period of infection symptoms arrive and the individual bears an overnight cost. The next period her status changes, and she has recovered. In the recovery state the individual is immune to the disease.

The sequence of events is as follows. In the initial state, the population is normalized to 1. Of these, all but a very small number ε are susceptible, with mass $S_0 = 1-\varepsilon$.[4] The ε individuals are infected with no symptoms, so they belong to state I. In the following period some susceptible individuals transition from state S to state I, because of contacts with the ε infected individuals. All individuals in state I in period t who survive make a transition to recovery a period later and join state R, after bearing a disease cost between periods t and $t+1$. The transition out of I and the recovery cost between periods depend only on medical conditions related to the disease that the individual cannot influence.

Key in our analysis is that the spread of the disease depends on the average activity level of the individuals in society. We will discuss this at some length below. For now we state that the "basic reproductive number" R_{0t} is time dependent. The rate at which people in the economy become infected in a given period t is given by $R_{0t} S_t I_t$, where R_{0t} is endogenously determined in equilibrium.

We follow the influenza and COVID-19 epidemiological literature and assume that the population is constant. Deaths in these epidemics is such a small fraction of infections that including them in the dynamic equations would greatly complicate the theoretical model for trivial extensions to the results.

With transition probability from state S to state I given by $R_{0t} I_t$, the number of people in state S falls each period by the same fraction. This is also the number of people who join the I state, whereas a period later infected individuals are removed from both states S and I. In discrete time, the dynamics of the system can be written as[5]

$$S_{t+1} = S_t - R_{0t} I_t S_t \tag{1}$$

$$I_{t+1} = R_{0t} I_t S_t. \tag{2}$$

$$R_{t+1} = R_t + I_t \tag{3}$$

The effective reproductive number, or R-number, is given by $R_{0t} S_t$. This number plays a key role in the dynamics of the disease. From (2), infections begin to fall when the R-number drops below 1. A key element in our analysis is to determine R_{0t}.

3.2 Individual behaviour

In our economy, individuals choose their activity levels. We will discuss examples of activity levels below. At an abstract level, activities are actions that give the

[4] We use S, and later I and R, to denote both the state and the mass of individuals that belong to it. In general, when reference is to the state there is no subscript but when reference is to the mass it is dated with a time subscript.

[5] The equations are slightly simpler than in the standard SIR model, reflecting that all infected indivduals in period t have recovered in period $t+1$.

agents utility but also increase the risk of attracting the disease. An activity level of x_t in a given period gives the agent a utility in that period equal to $\phi(x_t)$, where $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ is continuously differentiable and strictly concave in x_t. We assume that $\lim_{x_t \to 0^+} \phi(x) = -\infty$, and that $\phi(x_t)$ has a maximum at a value of x_t normalized to 1. We further assume that $\phi(1) > 0$.[6]

As noted above, infected agents are without symptoms, hence they cannot distinguish between being in state S or in state I. However, before an infected agent recovers, she incurs a cost, and hence observes that she moves to the recovered state. So individuals perceive to be in one of two states: not recovered or recovered and immune. Not recovered individuals are either healthy but susceptible to the disease or infected without symptoms. We introduce a new term for these individuals, *vulnerable*. All vulnerable individuals choose the x_t that maximizes their lifetime expected returns subject to the probabilities of belonging to state S or I.[7]

Vulnerable agents perceive a risk of infection from contacts with infected individuals. This probability depends on the activity level of the agent in question, the activity level of the other agents in the economy, and the number of infected people in the economy. In general, an agent who is in state S in period t enters state I in period $t+1$ with probability,

$$p_{t+1} = p(x_t, \cdot), \tag{4}$$

where x_t is the choice of social activity made by the agent, and $\cdot$ denotes a set of variables that are exogenous from the point of view of the individual agent, to be specified below. We assume that $\frac{\partial p(x_t,.)}{\partial x_t} \geq 0$. Note that p_{t+1} is the probability that the susceptible agent attracts the virus in period t and is infected in period $t+1$, and so it is predetermined at the start of period $t+1$.

The NPV of utility of a recovered individual is denoted by W_t^R. Assuming infinite lives and immunity, the recovered individual chooses x_t so as to maximize $\phi(x_t)$ every period.[8] By assumption the solution is $x_r = 1$. With a discount factor β we write

$$W^R = \frac{\phi(1)}{1-\beta}. \tag{5}$$

The NPV of utility of an infected individual is denoted by W_t^I, while the NPV of utility of a susceptible individual is denoted by W_t^S. Consider now a vulnerable individual in period t. She chooses x_t without knowing whether she is in the infected state or in the susceptible state. The expected NPV of utility of vulnerable individuals is denoted by W_t^V. The vulnerable person knows that she was susceptible in the previous period $t-1$, otherwise she would have been sick and recovered between $t-1$ and t, and she knows the probability p_t that she contracted the virus in that period, which is now predetermined. She chooses the present-period x_t by maximizing the expected value of the PDVs given her p_t :

[6] In Garibaldi et al. (2020) we discuss the micro-foundation of $\phi(x)$.

[7] Testing the population not in the recovery state would reveal the information needed to distinguish between the susceptible and the infected agents, but it is assumed absent.

[8] For an economic model with waning immunity in SIRS model see Goenka and Nguyen (2022).

$$W_t^V = \max_{x_t} \left\{ p_t W_t^I + (1 - p_t) W_t^S \right\}. \tag{6}$$

Both infected and susceptible individuals obtain utility from activity $\phi(x_t)$. An individual who is infected in period t will suffer overnight a utility loss associated with treatment. An infected individual is assumed to enter period $t + 1$ with expected utility $(1 - \delta(I_t))\, W^R$. The expression $\delta(I_t) W^R$ is the expected cost of contracting the disease. We assume that $\delta'(I_t) > 0$, to indicate medical congestion. The medical congestion is due to fixed hospital space and medical personnel, at least in the time scale of the disease. It therefore follows that the NPV of utility of an infected individual, W_t^I, is,

$$W_t^I = \phi(x_t) + \beta(1 - \delta(I_t)) W^R. \tag{7}$$

Recall that a susceptible individual is infected with probability p_{t+1}, hence

$$\begin{aligned} W_t^S &= \phi(x_t) + \beta \left[p_{t+1} W_{t+1}^I + (1 - p_{t+1}) W_{t+1}^S \right] \\ &= \phi(x_t) + \beta W_{t+1}^V. \end{aligned} \tag{8}$$

Substituting (8) and (7) into the maximization problem of a vulnerable individual, (6), this problem becomes,

$$W_t^V = \max_{x_t} \left\{ \phi(x_t) + \beta p_t (1 - \delta(I_t)) W^R + \beta(1 - p_t) W_{t+1}^V \right\}. \tag{9}$$

Differentiation of (9) for period $t + 1$, using (4), gives

$$\frac{\partial W_{t+1}^V}{\partial x_t} = -\beta p'(x_t) \left[W_{t+2}^V - (1 - \delta(I_{t+1})) W^R \right].$$

The first order condition for x_t that is obtained from problem (9) is therefore given by,

$$\frac{\phi'(x_t)}{1 - p_t} = \beta^2 p'(x_t) \left[W_{t+2}^V - (1 - \delta(I_{t+1})) W^R \right]. \tag{10}$$

It is clear from the first order conditions that in the case of an infectious disease vulnerable agents restrict their activities to avoid infection. We refer to this property as *social distancing* or *shielding*.

3.3 Contact technology

We now specify the contact technology that yields the infection probability $p(x_t, .)$. This parallels the matching function of labour economics (Petrongolo and Pissarides, 2001), but with some important differences. In the matching function of the labour literature, more workers looking for jobs reduces the success probability of a single worker because of congestion externalities in the application process. Here more

individuals coming out in the marketplace increases the chances of infection because a single exposed individual can infect many people; the infectious disease is "non-exhaustible," in the sense that many people could acquire it from a single person at the same time.

To provide an intuitive derivation of our contact function suppose $\bar{x}_t$ stands for the number of trips outside the house that each person does on average.[9] Assume that with $\bar{x}_t$ trips, each person experiences on average m_t contacts, defined by a well-behaved function $m_t = m(\bar{x}_t)$, with $m'(\bar{x}_t) \geq 0$. The function $m(.)$ is similar to the matching function of labour economics in the sense that it depends on the structure of the marketplace, including density of population, transportation facilities, types of establishments etc.[10] Some of these contacts are between susceptible and infected people, which lead to the infection of the susceptible agent with some positive probability that depends on the infectiousness of the disease.

Our formulation of the contact function implies that the presence of recovered individuals does not influence the infection rate of susceptible individuals. This is a reasonable assumption in situations in which meetings do not crowd out each-other, and is also the most common assumption in the literature. However, the assumption may be less reasonable when meetings are one-to -one. Alternative formulations of the contact technology are discussed further in Garibaldi et al. (2023). Numerically, the exact formulation of the contact function in this respect matters little.

Consider now the choice of activity level made by a single individual. Here we follow the method used in search theory to choose the optimal search intensity (Pissarides 2000, chapter 5). With $m(\bar{x}_t)$ representing the total number of contacts for $\bar{x}_t$ outings, each outing on average generates $m(\bar{x}_t)/\bar{x}_t$ contacts. So if the individual chooses an activity level x_t, her contacts are on average $x_t m(\bar{x}_t)/\bar{x}_t$. On average the fraction of contacts that are infected is equal to the fraction of persons in set I in the population. With the normalization of the population size to unity, we obtain that the probability that a contact is with an infected person is simply I_t.[11] Finally, suppose that the probability that a contact between a susceptible and an infected person leads to the infection of the susceptible person is an exogenous medical constant k. The transition from the susceptible to the infected state for the person who chooses x_t becomes,

$$p_{t+1} = k \frac{x_t m(\bar{x}_t)}{\bar{x}_t} I_t. \tag{11}$$

Hence we assume that the probability that a person is infected is proportional to the number of people she meets. One way of motivating this is as follows: Since for each contact there is a probability $(1 - I_t)$ that the person does not meet an infected person, there is a probability $(1 - I_t)^{xm(\bar{x})/\bar{x}}$ that the person does not meet any infected persons

[9] Appendix 1 derives a special example of a contact function from the urn-ball game, which satisfies the main properties of the general form discussed here.

[10] The dependence of $m(.)$ on a single variable parallels the contact function used by Diamond (1982) in his famous "coconut" paper. He assumes that there are b agents with a coconut each coming into contact pairwise, for a contact technology $m(b)$ with $m'(b) > 0$.

[11] If death was included in the formal analysis, the fraction of infected people would be slightly higher than I_t. That would be the only implication of allowing for deaths.

in her x outings. If I is a small fraction of the population, this is approximately equal to $\exp\{Ixm(\bar{x})/\bar{x}\}$, so the probability of meeting an infected person is $1-\exp\{.\}$ and for small transition probability this is approximately equal to the expression in the text.[12]

3.4 Equilibrium

We assume that the agents have rational expectations regarding the dynamics of the aggregate state variables S_t, I_t and R_t given by (1)–(3). In a symmetric Nash equilibrium all agents choose the same policy, so $x_t = \bar{x}_t$. For notational simplicity we drop the bar from $\bar{x}_t$. Hence, from (11), $p'(x_t) = k\frac{m(x_t)}{x_t}$, which inserted into the first-order condition for x_t, (10), gives that

$$\frac{\phi'(x_t)}{1-p_t} = \beta^2 k \frac{m(x_t)}{x_t} I_t \left[W^V_{t+2} - (1-\delta(I_{t+1}))W^R \right]. \tag{12}$$

Furthermore, from (11), we have that

$$p_{t+1} = km(x_t)I_t, \tag{13}$$

where x_t is the solution to (12). It follows that $R_{0t} = km(x_t)$ in (1)–(3). This completes the specification of behaviour in our model. We are now in a position to define our decentralized equilibrium.

Definition 1 A decentralized epidemic equilibrium is a set of sequences of state variables $\{S_t, I_t.R_t\}_{t=0}^{\infty}$, a set of value functions $\{W_t^V, W^R\}_{t=0}^{\infty}$, a sequence of social contacts $\{x_t\}_{t=0}^{\infty}$, and a sequence of infection probabilities $\{p_t\}_{t=0}^{\infty}$, such that, for given initial conditions $S_0 = 1-\epsilon$, $I_0 = \epsilon$,

1. S_t, I_t and R_t solve Eqs. (1)–(3) with $R_{0t} = km(x_t)$.
2. The value functions W^R and W_t^V solve Eqs. (5) and (9), respectively.
3. x_t solves the first order condition (12)
4. p_t solves Eq. (13).

Note that the derivative of p_t with respect to x_t along the equilibrium path, as defined by (13), does not have to be equal to the derivative of p_t with respect to x_t at the individual level, given by (11); this will depend on the scale properties of the matching function and will be discussed below.

Some obvious properties of the model, given our functional assumptions, can easily be derived. There is more social distancing (lower x_t), for higher k and higher I_t, and for higher recovery costs.[13]

[12] This approximation is reasonable for the infection probabilities we encounter in the simulations of the optimizing SIR model. With the lowest cost of infections, the highest infection probability is around .13 calculated using (11). The exact probability using the Poisson distribution is .122.

[13] We have not shown existence or uniqueness of equilibrium. However, our simulations show that equilibrium exists, and do not indicate the existence of multiple equilibria.

4 Externalities and deviations from social efficiency

4.1 Formal statement

As in other models of pairwise interaction, we would expect the decision strategies derived in the preceding section to be subject to externalities and inefficient outcomes. We derive the socially optimal strategies by assuming the existence of a social planner, who chooses social activities for all agents. The information that the social planner possesses about agent identities and the future path of the economy coincides with that of private agents. As with private agents, the planner chooses the same x_t for all vulnerable agents and a separate x_{rt} for the recovered. However, since the recovered are immune, and they do not influence the transition rates of the susceptible or infected individuals, the social planner will choose $\phi'(x_{rt}) = 0\,\forall t$. This matches private choices so it can be ignored and we can focus our analysis on the vulnerable agents only.[14]

There are two reasons for why the social solution deviates from the decentralized solution. First, one person's activity level may influence other people's meeting rate. Second, the planner is aware that her actions today influences the future dynamics of S_t and I_t, through Eqs. (1)–(2). These facts are ignored by private agents in the decentralized equilibrium. We called externalities that might arise from the first channel *static externalities*. The externalities due to the second channels are the *dynamic externalities,* which we subdivided into the medical congestion, contagion and immunity externalities.

Given W^R, the planner cannot influence the utility of an infected individual. The NPV utility of all vulnerable individuals is the same value function W_t^V as defined above. Hence we can derive the social optimum choice of x_t by having the social planner step into the shoes of the vulnerable agent and select x_t to maximize W_t^V. The planner's controls are the activity levels $\{x_t\}_{t=0}^{\infty}$. In any period, p_t, S_t, and I_t are pre-determined state variables.[15] The objective of the planner is the constrained maximization of the value function

$$\begin{aligned} W_t^V(S_t, I_t, p_t) = \phi(x_t) + \beta p_t W^R(1-\delta(I_t)) \\ +\beta(1-p_t)W_{t+1}^V(S_{t+1}, I_{t+1}, p_{t+1}), \end{aligned} \tag{14}$$

subject to the laws of motion (13) and (1)–(2) for p_t, S_t and I_t, respectively. The first-order condition for x_t then reads

[14] As discussed in Sect. 3, some types of meeting technologies may imply that a high activity level of immune individuals may reduce the transmission rate in society. If so, the planner may want to increase the activity level of the recovered individuals above 1. In this paper we do not consider this special case.

[15] Since S_t, I_t and p_t are predetermined variables, we include them all as state variables. Note that with identical activity levels for all agents, p_t is equal to the fraction of infected to non-recovered individuals ($p_t = I_t/(S_t + I_t)$). Still it is very convenient to include p_t as a separate state variable. In the decentralized solution, agents consider p_t (on individual level) as a state variable that depends on their activity level in the previous period. Including p_t as a state variable in the planner's problem thus facilitates the comparison of the planner's solution and the decentralized solution.

$$-\phi'(x_t) = \beta\,(1-p_t)\left\{\frac{\partial W^V_{t+1}}{\partial p_{t+1}}\frac{\partial p_{t+1}}{\partial x_t} + \frac{\partial W^V_{t+1}}{\partial I_{t+1}}\frac{\partial I_{t+1}}{\partial x_t} + \frac{\partial W^V_{t+1}}{\partial S_{t+1}}\frac{\partial S_{t+1}}{\partial x_t}\right\}. \quad (15)$$

The left-hand side of (15) measures the utility cost of the social distancing in the current period, the deviation of $\phi'(x_t)$ from the unconditional optimum 0. The right-hand side gives the gains in expected lifetime returns from the social distancing, which accrue to the susceptible individuals (the fraction $1-p_t$ of all vulnerable), from the next period onwards, discounted to the present at β.

We show:

Proposition 1 *he social optimum level of social activity is the solution to the following equations:*

$$\begin{aligned} -\frac{\phi'(x_t)}{1-p_t} &= -\beta^2 k m'(x_t) I_t\left[W^V_{t+2} - (1-\delta(I_{t+1}))W^R\right] \\ &\quad + \beta\left(\frac{\partial W^V_{t+1}}{\partial I_{t+1}} - \frac{\partial W^V_{t+1}}{\partial S_{t+1}}\right) k m'(x_t) S_t I_t. \end{aligned} \quad (16)$$

$$\frac{\partial W^V(S_t, I_t, p_t)}{\partial I_t} = -\beta p_t \delta'(I_t) W^R - \phi'(x_t)\frac{m(x_t)}{m'(x_t) I_t} \quad (17)$$

$$\begin{aligned} \frac{\partial W^V(S_t, I_t, p_t)}{\partial S_t} &= \frac{1-p_{t+1}}{p'_{t+1} S_t}\left[\phi'(x_t) + \beta(1-p_t)p'_{t+1}\frac{\partial W^V_{t+1}}{\partial p_{t+1}}\right] \\ &\quad + \beta(1-p_t)\frac{\partial W^V_{t+1}}{\partial I_{t+1}}, \end{aligned} \quad (18)$$

where the dynamics of p_t are given by (13).

Equation (16) follows from the first order condition (15), the fact that $\frac{\partial W^V_{t+1}}{\partial p_{t+1}} = \beta(W^V_{t+2} - (1-\delta(I_{t+1}))W^R)$ from (14), and that $\partial p_{t+1}/\partial x_t = k m'(x_t) I_t$ from (13). Proofs of Eqs. (17) and (18) are given in Appendix 2.

If we compare the market solution given by (12) and the planner's solution given by (16), we note the following: the first term on the right-hand side of the planner's first order condition is similar to the right-hand side of the decentralized solution, with the exception that $k\frac{m(x_t)}{x_t}$ is replaced by $k m'(x_t)$. If the two differ, this is the source of the static externality. The last term in the planner's solution is absent in the decentralized solution and reflects the dynamic externalities.

Equation (17) expresses the social cost of having one more infected individual in period $t+1$. This cost has two parts. First, more infected people may increase the cost of being infected for all infected agents. This is the medical externality and is captured by the first term. The second term is the contagion externality, the impact of higher I_{t+1} on future infections. One additional unit of I_t increases the number of

infected people in the next period with the same number as $\frac{m(x_t)}{m'(x_t)I_t}$ additional units of x_t.[16] The unit cost of reducing x_t is $\phi'(x_t)$. The marginal cost of increasing I_t due to the potential for more infections in period $t+1$ is the product of the two, as in the second term on the right-hand side of (17).

Finally, (18) is the basis of the immunity externality associated with the induced changes in S_t, which is the most subtle of the externalities. For a given (optimal) sequence of activity levels $x_t, x_{t+1}, \ldots$, a unit decrease in S_t does not influence the pay-offs in period t, nor the probability p_{t+1}, since $p_{t+1} = x_t k m(x_t) I_t$. However, S_t does influence the number of infected people in period $t+1$, $I_{t+1} = k m(x_t) S_t I_t$, and hence also the probability of contracting the virus in that period, $p_{t+2} = I_{t+1} k m(x_{t+1})$, and in later periods as well. A fall in S_t thus reduces infections in later periods, and takes the economy closer to herd immunity.

To give some intuition for the expressions in (18), we note that the planner can costlessly react to an increase in dS_t by increasing x_t by $\frac{1-p(x_{t+1})}{p'(x_{t+1})S_t} dS_t$ units, thereby keeping S_{t+1} unchanged. But this increases I_{t+1} by one unit and p_{t+1} by $p'(x_t)\frac{1-p_{t+1}}{p'_{t+1}S_t} dS_t$ units. The per-unit gain to the planner of increasing x_t is $\phi'(x_t)$ units. Together this gives (18). This argument is also made rigorous in the Appendix. Note that $\frac{\partial W^V_{t+1}}{\partial I_{t+1}}$ is given by (17), and $\frac{\partial W^V_{t+1}}{\partial p_{t+1}}$ follows readily by taking the derivative of (14).

Our main interest is the net effect of the immunity externality and the contagion externality. To this end, consider the case in which $\delta(I) = \delta_0$ (no medical externalities). Define $Z_{t+1} \equiv \frac{\partial W^V_{t+1}}{\partial I_{t+1}} - \frac{\partial W^V_{t+1}}{\partial S_{t+1}}$ as the net dynamic externalities; the difference between the contagion and immunity externalities. If $Z_t \leq 0$ for some t, the contagion externality dominates over the immunity externality at this point in time, and the social planner will want to impose more social distancing than in the decentralized solution.

In Appendix 2 we show that along the planner's optimal path, the net dynamic externality Z_t evolves according to the following difference equation

$$\begin{aligned} Z_{t+1} \quad = \quad & -\phi'(x_{t+1})\frac{m(x_{t+1})}{m'(x_{t+1})I_{t+1}} + \beta(1-p_{t+1})\phi'(x_{t+2})\frac{m(x_{t+2})}{m'(x_{t+2})I_{t+2}} \\ & + \beta(1-p_{t+1})(1-p_{t+2})Z_{t+2}. \end{aligned} \tag{19}$$

For given vectors of variables S_t, I_t, and x_t, (19) is a relatively complex first order difference equation. As is standard in rational expectations models, the associated homogenous equation is not stable, as $\beta(1-p_{t+2})(1-p_{t+1}) < 1$. The value of Z_t at any point in time depends on all future values of Z_t, albeit with decreasing weights, due both to discounting and to the fact that fewer and fewer people are still susceptible.

We can calculate the value of Z_t *at the limit*, denoted Z_∞, with $I = I_\infty = 0$ and $S = S_\infty$, and hence obtain a terminal value for Z_t. Clearly, at the limit, the immunity externality is 0, as one more susceptible individual will not influence the utility NPV of a vulnerable individual when there are no infected people around. However, the contagion externality will still be strictly positive. Even though the epidemic is over,

[16] This follows from the fact that $dI_{t+1} = k m(x_t) S_t dI_t + k m'(x_t) S_t I_t dx_t$.

one infected individual may still infect others before the illness again dies out. Suppose the number of infected individuals rises from 0 to δI. The δI individuals will in expectation pass on the illness to $R_{0t} S_\infty \delta I < 1$ individuals, who in turn will each pass it on to the same expected number of individuals. Since society has obtained herd immunity at this stage, the process converges. In the Appendix we show that Z_∞ can be written as

$$Z_\infty = -\frac{\beta^2 R_0}{1 - \beta R_0 \bar{S}} \delta_0 W^R \quad < 0, \tag{20}$$

so contagion trivially dominates.

4.2 Discussion of externalities

As already pointed out, the first order condition for the planner, given by (16), deviates from that of the agents in the decentralized solution, given by (12), in several ways.

First, in the first term on the right-hand side, the factor $k\frac{m(x_t)}{x_t}$ in the decentralized solution is replaced by $km'(x_t)$ in the planner's solution. The two solutions give identical outcomes when the meeting technology exhibits constant returns to scale:

$$\frac{x_t m'(x_t)}{m(x_t)} = 1. \tag{21}$$

This requirement parallels the familiar elasticity condition from matching theory, often referred to as the Hosios (1990) condition, which applies to situations of pairwise matching (Pissarides 2000, chapter 8). If the condition is not satisfied, the social planner will want to impose more or less social distancing because of this (static) externality.

To show how the returns to matching and social distancing bring about this externality, we differentiate the infection probability of a single agent, (11), with respect to other agents' actions, $\bar{x}_t$. With constant returns in $m(\bar{x}_t)$, the partial is zero, so less social distancing by others in the market does not influence this person's infection probability. But with increasing returns, $\partial p_{t+1}/\partial \bar{x}_t > 0$; a person is more likely to be infected when other people in her community reduce their social distancing. This is the essence of the externality: it arises in situations in which a change in a typical individual's social distancing has an impact on other people's infections probability.[17]

Consider next the dynamic externalities. Intuitively, the contagion externality will make the social planner want more social distancing. When the private agent reduces her social activity, she reduces the number of infected people next period, so makes some other people better off. She ignores this effect of her actions, so the social planner will want her to reduce her social activities further.

But the immunity externality will have the opposite effect. When the private agent reduces her social action she raises the number of susceptible people and delays the end of the epidemic. The social planner, in contrast to the agents in the decentralized

[17] See the pioneering work of Diamond and Maskin (1979), in which they explore meetings with linear and quadratic "search technologies."

solution, takes into account the impact of the lower x_t on the path of susceptible agents and the prolongation of the epidemic. If the immunity externality is stronger than the other externalities present, so that the planner (for given state variables S_t and I_t) shields less than the agents in the decentralized economy, we say that we have a race to shield.[18]

The medical congestion externality, which we have ignored so far, introduces yet another dimension to the divergence between private and social outcomes. Medical congestion arises when infections are high, so the planner's objective is to reduce peaks in infections. This is something that private agents do not take into account in their decision making. Given that a certain number of infections needs to take place before herd immunity, for the social planner a flat infections curve is better for medical reasons than one that goes up and down. This will typically imply more social distancing by the regulator early on, when infections are rising fast, but less later, when they flatten out.

5 Vaccination

Our discussion of the dynamic externalities in the preceding section has made it clear that they depend on the time that it takes the disease to end. We get the externalities because the social planner chooses a different path to the end of the disease from the one implied by private shielding decisions. A vaccine is therefore expected to have a large impact on the externalities, because it alters the end game of the disease.

Vaccines typically have limited effectiveness; for example, a vaccine might be less than 100% effective at avoiding the disease, or it might reduce the health costs to the individual once contracted.[19] These situations are easily handled in our model. A vaccine that gives limited immunity is one that reduces the infectiousness of the disease, the k in our model, to a number that is still positive, but less than the number before its introduction. One that reduces the medical cost can be modelled as a fall in the disease cost $\delta(I)$.

Rather than repeat our derivations for smaller k or $\delta(.)$ parameters, which would be trivial, we focus on a more interesting case. What if there was no vaccine as yet, but there was *expectation* that one would arrive. We are looking to find the impact of this expectation on our externalities, before the vaccine arrives. We assume that the process that brings the vaccine is Poisson and is correctly anticipated by both private agents and the social planner.

Two other assumptions help to make our analysis sharper. One is that once a vaccine arrives it eliminates the spread of the disease instantly. This avoids the impact of delays in the administration of the vaccine, or limitations in its effectiveness, which, although important in real life, would complicate results and lose clarity about the externalities

[18] Some considerations in this direction were stated in Britain at the very beginning of the COVID-19 pandemic, including the "eat out to help out" campaign in the summer of 2020. See https://www.politico.eu/article/dominic-cummings-uk-coronavirus-herd-immunity-hearing-committee/ for the government views about the prospect for herd immunity.

[19] For an optimal vaccination policy when the vaccine does not give immunity forever see Federico et al. (2022)

that arise when there is expectation of a vaccine arrival. Note though that the vaccine is not a cure, so it will not help those who are in the infected state.

The second assumption is that the expected time before the introduction of the vaccine is short enough so the disease is expected to end by the vaccine and not by herd immunity.

To derive the impact of these assumptions on the externalities, suppose there is a probability $\lambda > 0$ that a fully effective vaccine will arrive between this period and the next. The expected duration of the disease is therefore $1/\lambda$. If a vaccine arrives in period t, it does so before the agents make their activity choice x_t. We develop the theory and subsequently simulate the solution for a λ value that implies that in expected value terms, the disease will end through vaccinations and not through herd immunity.

Private agents might shield a little more when the expectation is that the disease will end long before herd immunity, because the continuation value of staying susceptible increases. The planner might also be subject to these considerations and reduce social action a little more. In addition, a concern for the planner before was to reduce the length of time that society needed to get to herd immunity. This length of time is now likely to be shorter and independent of private actions, since it is the vaccine that will stop the epidemic and not immunity through infections.

So for a high value of λ, we expect to find large differences between the planner's solution with and without a vaccine. To derive formally the impact of the vaccine arrival rate, denote by $\bar{W}_t^{VV}$ the net present value of utility of a vulnerable agent in period t if a vaccine arrives between $t-1$ and t, and W_t^{VN} the utility if it does not arrive. Define

$$W_t^{VV} \equiv \lambda \bar{W}_t^{VV} + (1-\lambda) W_t^{VN}. \tag{22}$$

Hence, W_t^{VV} is the expected NPV of utility of entering period t as vulnerable, in the absence of a vaccine in period t and without knowing that a vaccine will arrive in t.

5.1 The decentralized solution

Suppose the vaccine arrives in the beginning of period t. Because it immediately stops the disease, the continuation pay-off for an individual is βW^R if uninfected and $\beta(1-\delta(I_t))W_R$ if infected. Both are independent of her activity level x_t that period. Hence the individual sets $x_t = 1$ and obtains an instantaneous utility $\phi(1) = (1-\beta)W^R$. It follows that

$$\begin{aligned} \overline{W}_t^{VV} &= \phi(1) + \beta(1-p_t)W^R + \beta p_t(1-\delta(I_t))W^R \\ &= [1 - \beta p_t \delta(I_t)]\, W^R, \end{aligned} \tag{23}$$

that is, the NPV utility of the recovered less the utility loss that accrues if the agent is already infected.

Now suppose the vaccine has not yet arrived in period t. With predetermined probability p_t, the individual got infected, and her continuation value is $\beta(1-\delta(I_t))W^R$, independently of whether or not the vaccine arrives. With the complementary proba-

bility, the agent is not infected, and obtains a continuation pay-off of βW_{t+1}^{VV}. It follows that

$$W_t^{VN} = \max_{x_t} \left\{ \phi(x_t) + \beta p_t \left(1 - \delta(I_t)\right) W^R + \beta(1 - p_t) W_{t+1}^{VV} \right\} \tag{24}$$

It follows from Eqs. (22), (23), and (24) that

$$\begin{aligned} \frac{\partial W_t^{VV}}{\partial p_t} &= -\lambda\beta\delta(I_t) W^R - (1-\lambda)\beta \left[W_{t+1}^{VV} - (1 - \delta(I_t)) W^R \right] \\ &= -\beta \left[W_{t+1}^{VV} - (1 - \delta(I_t)) W^R \right] - \lambda\beta(W^R - W_{t+1}^{VV}). \end{aligned} \tag{25}$$

So the first order condition for the agents' maximization problem as defined by (24) reads

$$\frac{\phi'(x_t)}{1 - p_t} = \beta^2 p'(x_t) \left[W_{t+2}^{VV} - (1 - \delta(I_{t+1})) W^R + \lambda(W^R - W_{t+2}^{VV}) \right]. \tag{26}$$

The first two terms inside the square brackets reflect the cost of being infected if a vaccine does not materialize before the next period, and correspond to the analogous terms in (12). The last term reflects the expected gain to the susceptible from the vaccine's arrival, which the infected will not take part in.

The decentralized equilibrium can be defined analogously to the definition with no vaccine, given by Definition 1, with W_t^V replaced by W^{VN}.[20] Furthermore, the first order condition for x_t, (12), is replaced by (26).

5.2 The planner

Before a vaccine arrives, the planner chooses activity level so as to maximize the net present value of utility of the representative vulnerable person W^{VN} defined by (24). As above, the planner maximizes W^{VN} given the constraints (13), (1), and (2) (the law of motion for p_t, S_t and I_t). Parallel to (15), the first order condition for x_t reads

$$-\frac{\phi'(x_t)}{1 - p_t} = \beta \left\{ \frac{\partial W_{t+1}^{VV}}{\partial p_{t+1}} \frac{\partial p_{t+1}}{\partial x_t} + \frac{\partial W_{t+1}^{VV}}{\partial I_{t+1}} \frac{\partial I_{t+1}}{\partial x_t} + \frac{\partial W_{t+1}^{VV}}{\partial S_{t+1}} \frac{\partial S_{t+1}}{\partial x_t} \right\}. \tag{27}$$

Parallel to (16), we expand the first order condition to,

[20] It is slightly more cumbersome to characterize W_t^{VN} than W_t^V. W_t^{VN} is given by (24), which again depends on W_{t+1}^{VV}, defined by (22), and which includes $\overline{W}_{t+1}^{VV}$, defined by (23), while W_t^V is defined simply by(9).

$$\frac{\phi'(x_t)}{1-p_t} = \beta^2 km'(x_t)\left[W^{VV}_{t+2} - (1-\delta(I_{t+1}))W^R + \lambda(W^R - W^{VV}_{t+2})\right] + \beta\left(\frac{\partial W^V_{t+1}}{\partial I_{t+1}} - \frac{\partial W^V_{t+1}}{\partial S_{t+1}}\right) km'(x_t)S_t I_t. \quad (28)$$

The two dynamic externalities captured in the last term can be derived with the same methodology as when vaccines are absent, but taking into account that there is a probability $1-\lambda$ that the externalities unfold in the next period and $(1-\lambda)^2$ that they unfold two periods ahead. It follows that

$$\frac{\partial W^{VV}(S_t, I_t, p_t)}{\partial I_t} = -(1-\lambda)\phi'(x_t)\frac{m(x_t)}{m'(x_t)I_t} - \beta p_t \delta'(I_t)W^R \quad (29)$$

$$\frac{\partial W^S(S_t, I_t, p_t)}{\partial S_t} = \frac{(1-\lambda)(1-p_{t+1})}{p'_{t+1}S_t}\left[\phi'(x_t) + \beta(1-p_t)p'_{t+1}\frac{\partial W^{VV}_{t+1}}{\partial p_{t+1}}\right] + (1-\lambda)\beta(1-p_t)\frac{\partial W^{VV}_{t+1}}{\partial I_{t+1}}.$$

Naturally, the medical externality (the last term in the first equation) is unaffected by the prospect of vaccines, as it concerns those who are already infected when it arrives. The contagion externality is scaled down by a factor of $1-\lambda$, the probability that the vaccine has not yet arrived. That is also the case for the first term in the immunity externality. However, the second term of the immunity externality is scaled down by a factor $(1-\lambda)^2$ (since $\frac{\partial W^{VV}_{t+1}}{\partial I_{t+1}}$ is scaled down by $1-\lambda$). It follows that the planner is expected to reduce her optimal activity level by more than private agents are expected to do, and so weaken the race to shield.

To sum up, we would expect that private agents shield a little more when the expectation is that the epidemic will end before herd immunity, because the overall gain from staying susceptible increases as the likelihood of getting infected in the future falls. The planner is also subject to these considerations. The immunity externality will weaken, because it arises when private agents' shielding increases the number of susceptible individuals in the future, which is now more of a good thing as a vaccine then may have arrived. Hence we expect that the prospect of a vaccine reduces the activity level set by the planner more than the activity level set in the decentralized solution.

6 Simulations

6.1 Parametrization and calibration

In this section we parameterize the model, calibrate some important parameters, and simulate the model with different assumptions regarding the severity of the illness,

arrival rates of vaccines, returns to the contact functions, and the strength of the medical externalities.

We use the following parametrization of the net semi-indirect payoff function: $\phi(x_t) = A + lnx_t - x_t$. With this function the optimal x_t for all recovered agents satisfies the normalization $x_r = 1$. The PDV after recovery is $W^R = (A-1)/(1-\beta)$. We set $\beta = 0.998$. If a period is two weeks, this corresponds to an annual discount rate of 0.05.

The parameter A is certainly important, as it influences the value of life relative to the value of social activity. Hall et al. (2020) argue that the statistical value of life is 6 times higher than the NPV of consumption weighted by the marginal value of consumption. Although at a stretch, we interpret x_t as consumption, as do Farboodi et al. (2021). Furthermore, we interpret $\ln x_t$ as the utility of consumption, and x_t as the utility cost of making consumption available (for instance the utility cost of reduced leisure). We evaluate the marginal value of consumption at $x_t = 1$, the consumption level obtained in the absence of an epidemic and when recovered. The per period consumption value weighted by the marginal value of consumption is then 1. The per period utility is $A - 1$. Hence, for the per period utility (or NPV utility) to be 6 times higher than the per period level of consumption (or NPV value of the level of consumption), we must have that $A = 7$. We therefore set $A = 7$.

We model the cost of catching the disease (as a proportion of lifetime utility) as $\delta(I_t) = \delta_0 e^{\delta_1 I_t}$, where δ_0 and δ_1 are non-negative parameters. We simulate the model with a high and a low value of δ_0, the high value equal to 0.0083 (5/6 %) and the low value equal to 0.00177 (1/6 %). If the cost of the disease is death, the high and low costs can be interpreted as fatality rate of 0.0083 and 0.00177, respectively. The values span the fatality rates typically used in the recent epidemiological literature, as well as the COVID-19 death rates in Europe.[21] In the welfare analysis below we will use both the high and low values of δ_0.

Recall that the $R_{0t} = km(x_t)$, where R_{0t} is the basic reproductive number in period t. We write $R_{0t} = R_0 x_t^{\alpha}$, where the constant R_0 is the basic reproductive number when $x = 1$ and $S = 1$ (the basic reproductive number at the outbreak of the epidemic). As in most studies we set $R_0 = 2.4$.[22]

We begin our main simulations by focusing on the net impact of the two dynamic externalities, which quantitatively is the main new contribution of this paper. In order to do this as cleanly as possible, we set, in what we refer to as our baseline simulations, $\delta_1 = 0$ and $\alpha = 1$, thus shutting down the static externality and the medical congestion externality.

6.2 Preliminary simulation: the role of optimizing behaviour

As a preliminary to our main simulations, we first compare the baseline simulations of the decentralized solution of our model to the SIR model solutions without social distancing. We assume that no vaccine is expected to arrive, and set δ_0 at the high

[21] See https://covid19.who.int/.

[22] In order to economize on notation, we now let k denote the product of the infection probability conditional on a contact and any constants that might belong to $m(x_t)$.

value. The plots are shown in Fig. 1. In each chart the red/broken line refers to the decentralized outcome from our model while the black/continuous line to the non-optimization standard SIR model. As in most figures that follow, the top chart shows the activity level $\{x_t\}$, the middle chart the mass of susceptible people, and the bottom chart the mass of infected people. Figure 1 shows the first 100 periods of the simulation (about 4 years).

Before the disease the two models have the same activity level, normalized to unity. But whereas without optimization the activity level remains at that level throughout the epidemic, in the decentralized simulation it drops a little when the ε infections are first introduced and then drops dramatically and very quickly to a number below 0.5. This drop cuts the effective reproductive number in the decentralized equilibrium to below $1.2S_t$. The middle panel plots the stock of susceptible people, initially normalized to one for both models. In the SIR model with the high activity level continuing, the stock of susceptible people drops very quickly to the new steady state level. After about ten periods (20 weeks) only 3.7% of the susceptible people avoid the infection, which virtually corresponds to the end of the diseases ($I_t = 0$), reached when 3.6% avoid it. Adjustments are much slower in the case of the optimizing solution. As people drop the level of activity, the stock of susceptible people falls smoothly and gradually throughout the epidemic. After 10 periods, 72.5% of the population are still disease-

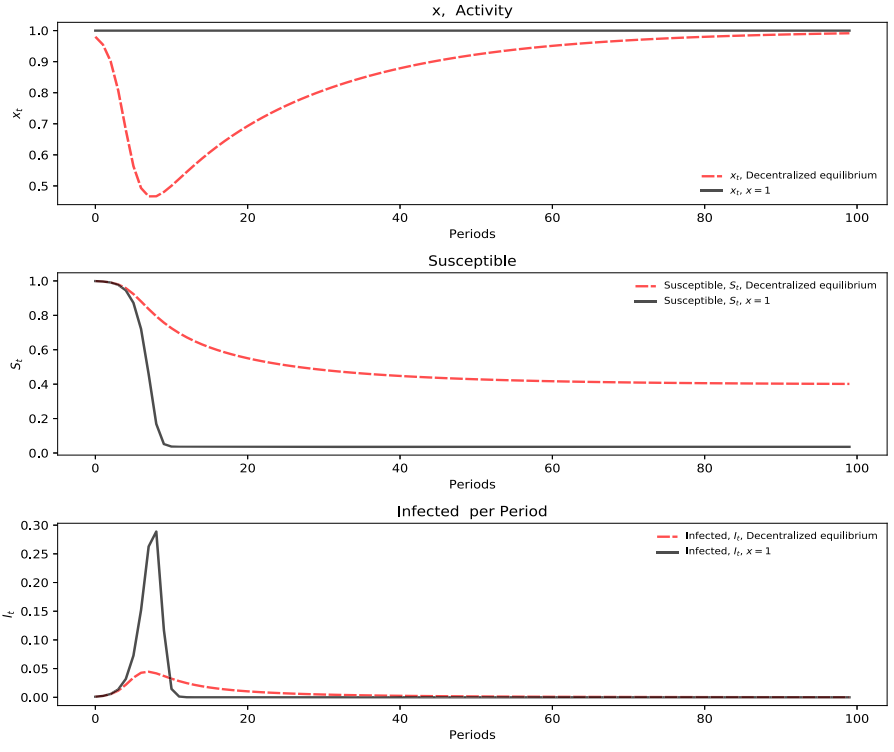

Fig. 1 Standard SIR versus optimizing SIR: 100 periods

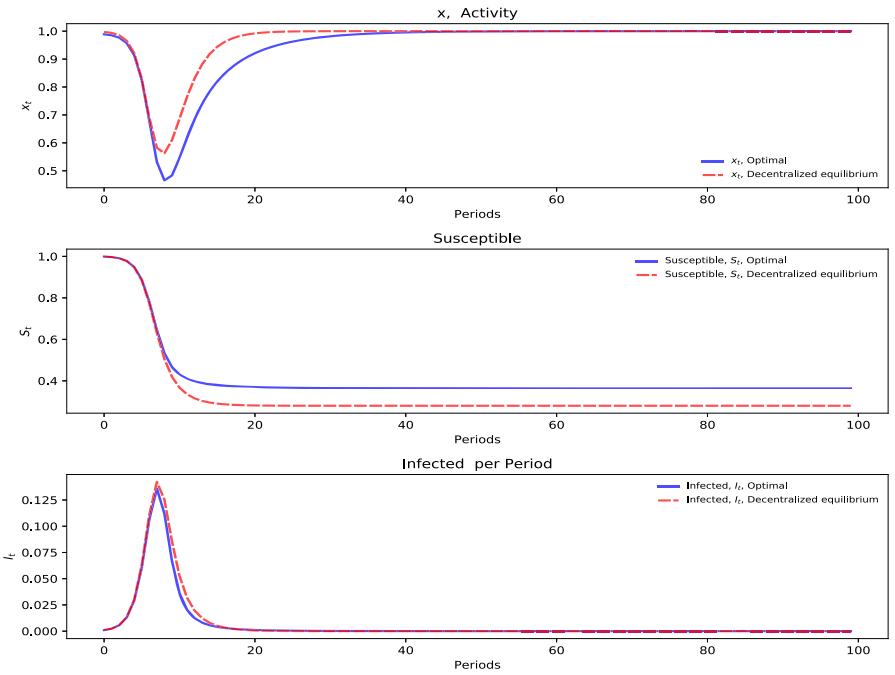

Fig. 2 Matching-SIR: planner versus market, low δ_0

free. We will discuss this in more detail below. In the bottom panel of Fig. 1 we plot the stock of infected people. At its peak it is around 10 times higher in the SIR model without social distancing compared with the optimizing SIR model.

6.3 The planner's solution

In this subsection we will first compare the planner's solution and the market solution in our main simulations with $\delta_1 = 0$ and $\alpha = 1$.[23]

We first consider a situation with low cost of attracting the disease, $\delta_0 = .017$. Figure 2 shows the activity level and the numbers of susceptible and infected individuals along the decentralized equilibrium path and the planner's optimal path. The dotted (red) line refers to the decentralized outcome while the continuous (blue) line refers to the central planner solution. From Fig. 2 we see that the planner tends to impose more social distancing than the decentralized equilibrium, particularly just after the peak of the disease (when I_t is at its maximum) and during the recovery.

This contrasts sharply with the impact of higher medical costs. For a fixed cost $\delta_0 = 0.0083$, Fig. 3 shows that there is a race to shield; the planner shields less than

[23] As a consistency check, we have simulated the model using different methods: the first order conditions of the planner defined above, first order conditions derived using Lagrange's method, and a more plain method where we search for a maximum over a grid. The computationally most efficient way of simulating the model is by using the Lagrangean. The simulations are done by Per August Moen, research fellow and Ph.D. student at the Department of Mathematics, University of Oslo.

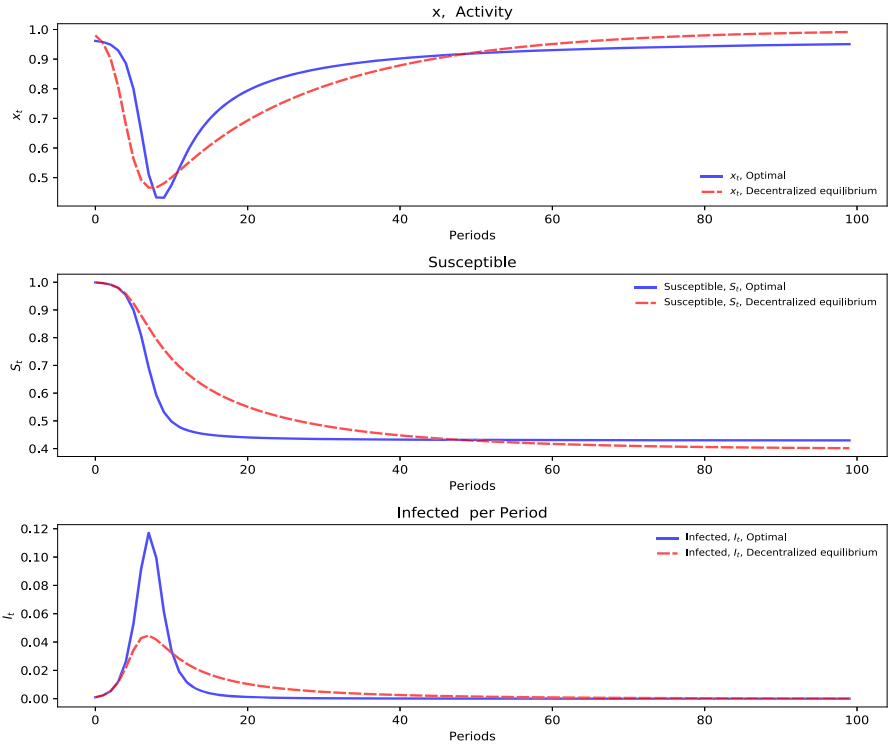

Fig. 3 Matching-SIR: planner versus market, high δ_0

the agents in the decentralized solution. In the first fifty periods of the epidemic (except briefly at the very top) the planner wants to raise the social activity level chosen in the decentralized solution, because of a strong immunity externality. Private agents shield too much when the medical costs are high, delaying the transition to herd immunity. However, towards the end of the epidemic the planner shields more than the agents in the decentralized solution, so that the aggregate number of infected people over the course of the epidemic is higher in the decentralized solution than in the planner's solution.[24]

Recall that the net dynamic externality Z_t captures the effect on W_t^V of having one more person infected and one less susceptible. With our parametrization, it follows that Eq. (19) can be written as

$$Z_{t+1} = -\frac{1 - x_{t+1}}{I_{t+1}} + \beta(1 - p_{t+1})\frac{1 - x_{t+2}}{I_{t+2}} + \beta(1 - p_{t+2})(1 - p_{t+1})Z_{t+2}. \quad (30)$$

[24] Note that except in the very beginning, the number of infected and susceptible individuals along the decentralized solution path and the planner's path differ, implying that one should be cautious when comparing the graphs of the planner's solution and the decentralized solution.

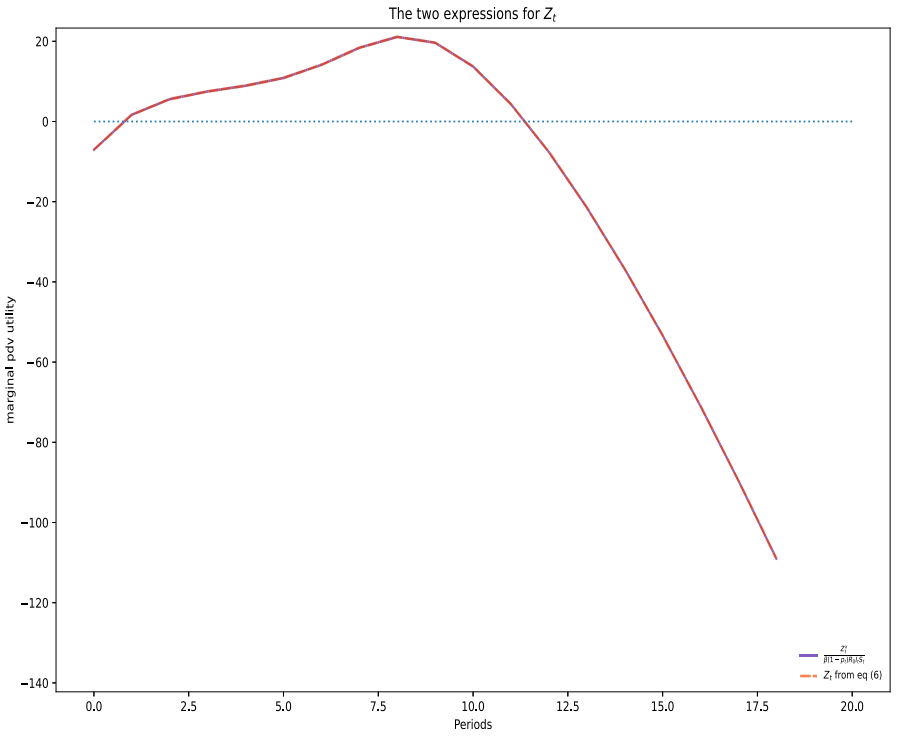

Fig. 4 Z_t first 20 periods

Recall that a negative Z_t implies that the contagion externality dominates and the social planner wants to impose more social distancing, whereas a positive one implies the opposite, that the balance of the dynamic externalities is for less social distancing by the planner.

The first 20 periods of Z_t are plotted in Fig. 4, and the first 150,000 periods in Fig. 5. Z_t is negative and increasing in the very beginning, then turns positive and peaks around period 10, and then decreases again to converge to Z_∞, as defined in Eq. (20) (horizontal line in Fig. 5). However, full convergence of Z_t takes a very long time (thousands of years).[25]

The dynamics of Z_t can easily be understood by inspecting (30) and the simulated values of I_t shown in Fig. 3. We know that $Z_\infty < 0$ (and large in absolute terms). After their peaks, I_t and x_t are both decreasing in t, $\frac{1-x_{t+1}}{I_{t+1}} < \frac{1-x_{t+2}}{I_{t+2}}$. If the relative difference between the two terms is bigger than the factor $\beta(1-p_{t+1})$, the sum of the two first terms is positive. In this case, as we move backwards in time, Z_{t+1} will be strictly greater than Z_{t+2} as long as Z_{t+2} is negative (or positive but not too large) since the factor in front of Z_{t+2} is less than 1. This is also what we see in the plot of Z_t. We also see that Z_t switches sign around period 12.

[25] Each of the figures actually shows two plots, Z_t calculated by using (18) and Z_t calculated with (30). The two are so close that they cannot be distinguished from each-other. Since Z_t defined by (30) holds only if the first order condition is satisfied, this is a consistency check of our results.

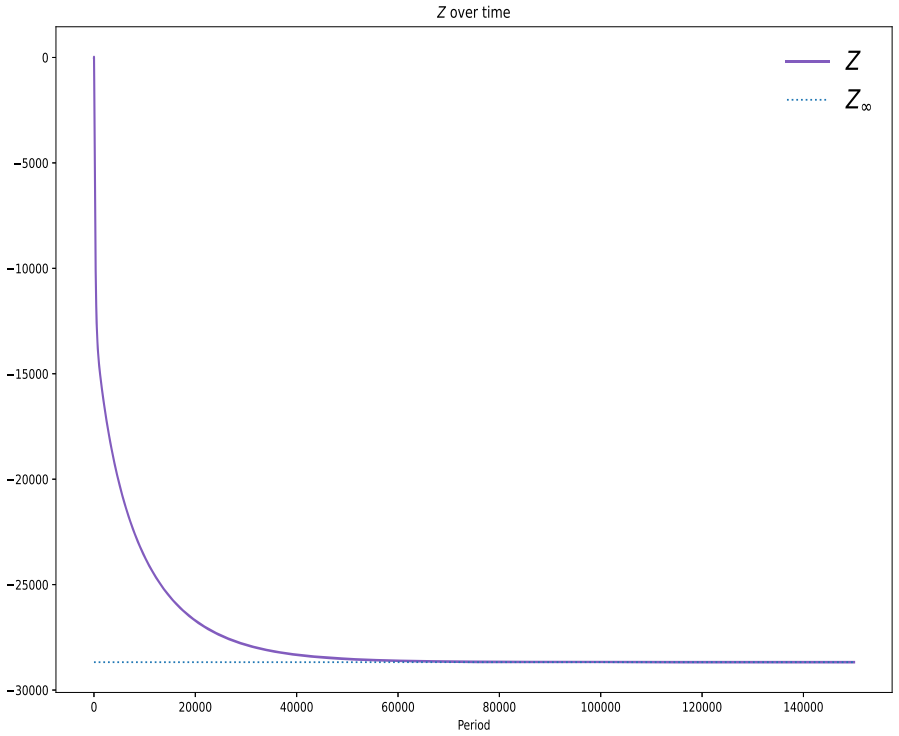

Fig. 5 Z_t first 150,000 periods

However, at the early stages of the epidemic, before I_t and x_t peak, they are both increasing in t. Hence the sum of the two first terms is negative, and Z_{t+1} will surely start to decrease again as we go back in time and t decreases. From Fig. 4 we see that Z_t becomes negative as t is approaching zero.

Our results beg the question: why is the planner closing down more than the market when medical costs (δ_0) are low, and less when they are high?

The epidemic ends when the number of infected agents converges to zero. i.e.,as I_t approaches 0. Furthermore, x_t converges to 1 because the infections risk is zero. At this stationary point, the effective reproduction number must be lower than 1. Hence, in the limit as $t \to \infty$, $R_{0t} S_t = km(1) S_t \leq 1$. Therefore, the supremum of the number of susceptible individuals who never catch the disease, S_∞, is given by $S^{\max} = 1/km(1)$, which, with our parameters, is equal to 0.417 (i.e., a maximum 41.7% of the population can avoid the epidemic).

With awareness of the requirement that herd immunity will bring an end to the disease with $S_\infty \leq S^{\max}$, the social planner faces a trade-off. Speed up the infection process to get to the end quickly and save the costs of deviations of x_t from its optimum, or slow down the process, bear more costs from lower activity levels, but reach the end with a bigger S_∞.

Compared with the decentralised solution, and for a low δ_0, the planner does much better when it comes to the aggregate number of infections. At $T = 40$ (approximately

after 1.5 years) the epidemic is for all practical purposes over both in the planner's solution and in the decentralized solution. At this stage the number of susceptible individuals is .36 in the planner's solution and .28 in the market solution. It follows that the planner is much closer to the maximum achievable number of susceptible individuals consistent with herd immunity. Note also that in the decentralized solution, S is 0.37, well below $S^{\max}$, already in period 10. Since curbing the epidemic takes time, as infected individuals will infect others before the epidemic dies out, the planner will cut back on activity relative to the decentralized solution early on in the epidemic in order to reduce the total number of infections before herd immunity is reached.

For a high δ_0, this is different. In the decentralized solution, the epidemic now moves much more slowly. After 40 periods, the fraction of susceptible individuals is still .45, well above $S^{\max} = .417$. It takes 60 periods before S falls below $S^{\max}$, and in the limit (after 300,000 periods), $S = .40$. The number of susceptible individuals is also very high for the planner. At $T = 40$, S is equal to .45 (the same number as in the market solution). In the limit (after 300,000 periods) S in the planner's solution is equal to .41663, extremely close to $S^{\max}$. The decentralized equilibrium also moves slowly in the beginning, after 10 periods the value of S is still .73.

It follows that with the high δ_0, the scope for the planner to do better than the decentralized solution in terms of a lower total number of individuals being infected is much less than with the low δ_0. The agents in the decentralized solution cut back more on x_t in the early phases of the epidemic than the planner needs to do in order to obtain a total number of infections close to the minimum. With the need to cut back to preserve lives removed, the dominant preoccupation of the planner becomes the long time needed to get to herd immunity. She shortens that time by choosing more social interaction early on, as dictated by the immunity externality. In other words, although private agents do well to shield more when medical costs are high, they overdo it when they exceed a certain level.

6.4 Implications of a vaccine

As we explained above, the expectation of a vaccine is likely to increase shielding by both the planner and the private agents, as the value of remaining susceptible increases. But because the planner responds to both the contagion and the immunity externality, and the latter is expected to weaken by more, we would expect the planner to reduce social activity by more than private agents. These expectations are confirmed in our simulations, for a fairly high, but not completely implausible, vaccine arrival rate. We select $\lambda = 0.05$, implying an expected duration of the disease of 20 periods (about 40 weeks). Although the COVID-19 vaccines were invented and approved very fast (on December 8, 2020, Britain became the first country to approve the Phizer-BioNTech vaccine, about 37 weeks after the first COVID-19 lockdown), there are some important differences between our assumptions about vaccines and COVID-19. The COVID vaccines took some time to implement, offered partial protection and needed renewal every few months. See below for more discussion of the

impact of less effective vaccines like the ones for COVID-19 and the annual influenza ones.

To see the significance of this parameter, refer to Fig. 3. Without a vaccine, the expected length of time before the end of the epidemic is in excess of 40 periods in decentralized equilibrium, whereas the social planner wants to reduce it to about 20. The corresponding trajectories with a vaccine are shown in Fig. 6. The optimal policy is to enforce much more shielding until the vaccine arrives, and much more than the agents in the decentralized solution, who change their behaviour very little. Of course, with this very cautious policy, if there are delays in the arrival of a vaccine, herd immunity is much harder to achieve. Even after 100 periods without a vaccine, the fraction of susceptible individuals in the optimal solution is around .8, and the epidemic is still far from ending. With our assumptions, the probability of reaching 100 periods without a vaccine arriving is around $1/2$ percent.

A concern for the planner before was to reduce the length of time that society needed to get to herd immunity. This concern is now less relevant, since the epidemic is likely to end as a result of a vaccine arriving. For a sufficiently high value for the arrival probability λ, and a correspondingly weakened immunity externality, the planner's action is dominated by the contagion externality. If the vaccine is expected to arrive quickly, avoiding infections even more than before is socially a good outcome.

We have focused on the extreme version of a vaccine, a fully effective one that ends the epidemic as soon as it arrives. Although this analysis illustrates well the implications of a vaccine, it is not what we see in practice. Vaccines for epidemics such as influenza or COVID usually offer substantial but not 100% protection, or they reduce the severity of an infection. We do not fully formalize such a scenario but our analysis so far gives good indications of the dynamics of the epidemic when vaccines are not fully effective. We reinterpret our formalization as one in which on arrival the vaccine reduces medical costs to zero. Consider the more realistic alternative, that on arrival infections can still happen, but the medical cost before recovery is substantially reduced. We have shown in our basic simulations that lower medical costs have a bigger impact on the immunity externality than on the contagion one. It follows that less effective vaccines that reduce medical costs still have the externality effects that we studied in the case of fully effective vaccines, but in weaker form. The immunity externality is weaker, but might still be sufficiently strong to offset the contagion one, depending on how much medical costs are reduced by the vaccine.

6.5 Medical externalities

We consider the impact of the medical congestion externalities in the absence of a vaccine, so our results in this subsection should be compared with those in Fig. 3. Figure 3 shows a high peak in the number of infected people in the optimal solution. With medical costs increasing in I_t due to congestions in the health sector, the social planner will want to avoid the congestion by flattening the infections curve. This is shown in Fig. 7. Recall that $\delta(I) = \delta_0 e^{\delta_1 I_t}$. We set $\delta_0 = 0.0058$

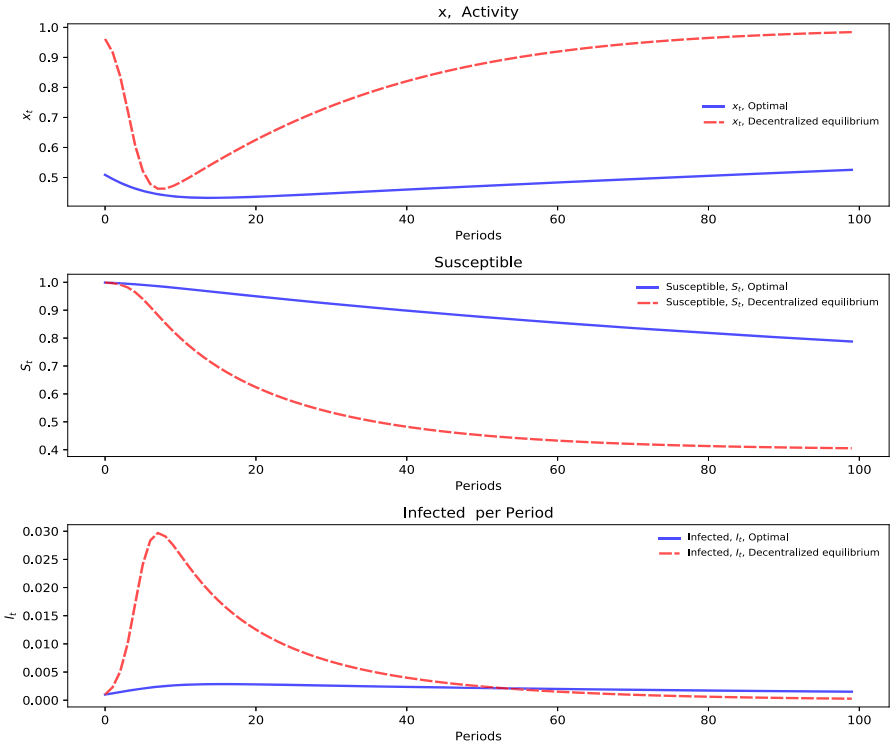

Fig. 6 Planner versus market with vaccine, $\lambda = 0.05$

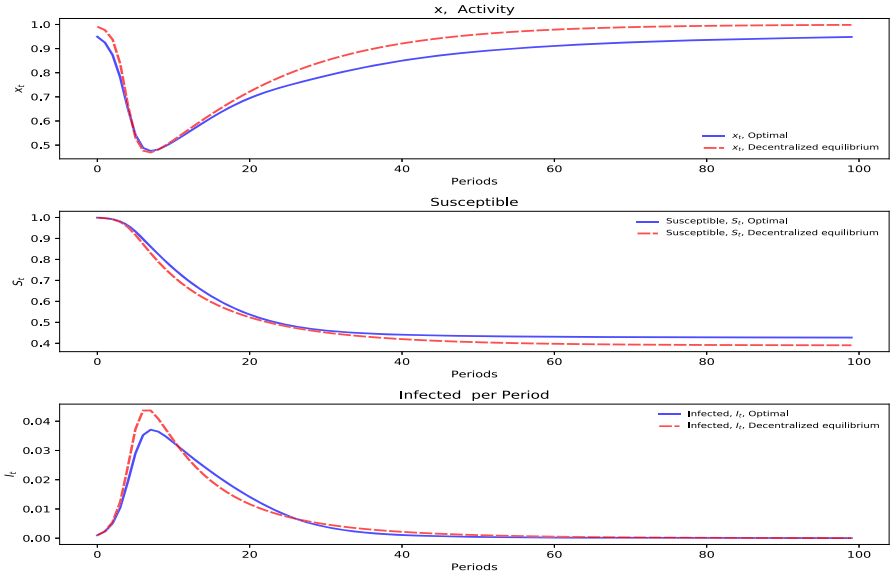

Fig. 7 Planner versus market with medical externalities

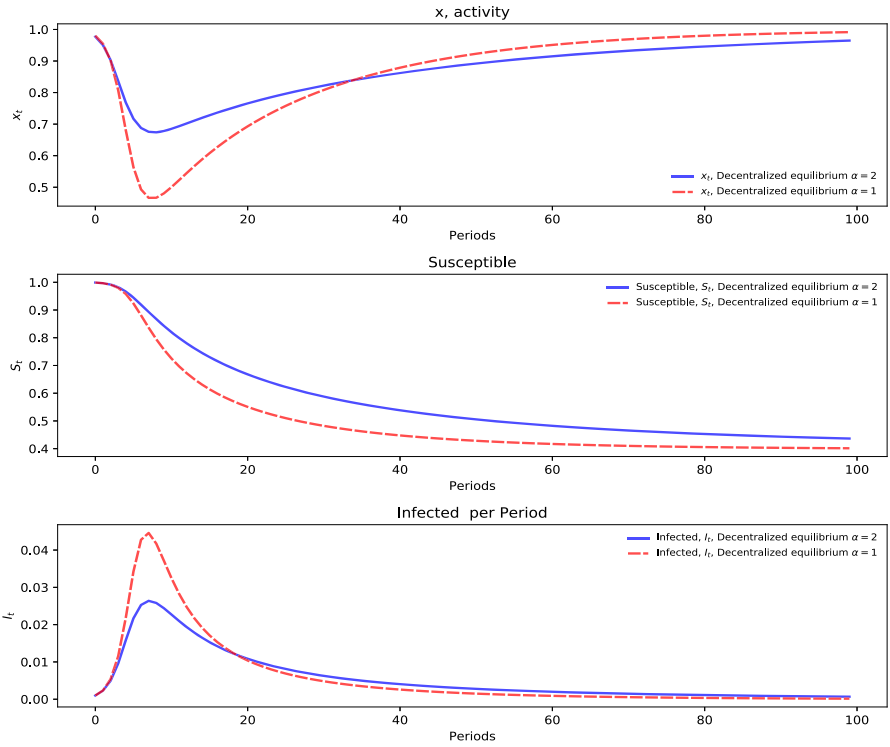

Fig. 8 Decentralized equilibrium with constant and increasing returns to scale, high δ_0, no vaccine or medical externalities

(35/600%) and $\delta_1 = 6$, so that medical costs increase fast with I_t and the mean overall cost between $I_t = 0$ and I_t at its maximum in the absence of the medical externality is approximately 0.0833, as in the case of the high fixed medical costs.

In the comparison between Figs. 3 and 7, we see that the social planner has reduced the high point of infections much more than the decentralized choices. The adjustments made to the social action that the planner chooses in order to achieve the flattening of the infections curve are such that the paths of the planner become very close to the decentralized paths, but this is a coincidence due to the particular parameter set that we have simulated.

In the absence of medical congestion externalities, agents shield more than the social planner because of the rat race to shield. The medical congestion externality gives rise to a force in the opposite direction, as agents in the decentralized solution shield less because they ignore the impact of their illness on the medical costs. With our parametrization the two effects cancel each-other out.

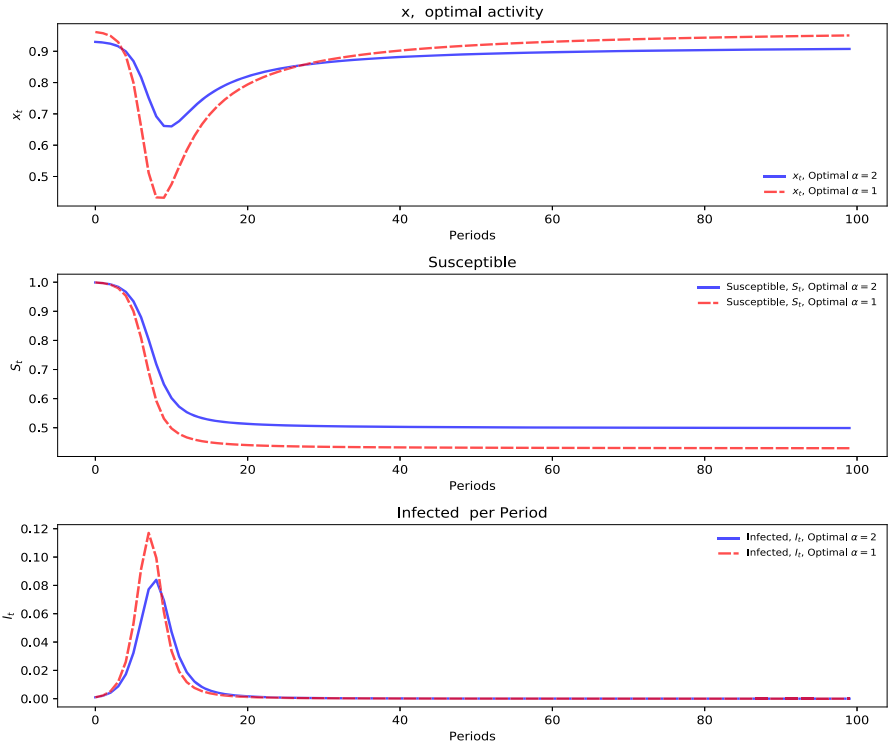

Fig. 9 Central planner equilibrium with constant and increasing returns to scale

6.6 Increasing returns to scale in contacts

So far the simulations of this section focused on the dynamic externalities, which is the main new contribution of this paper. The static externalities were shut out by imposing constant returns to scale in the contact technology. We now relax this assumption and derive the implications of increasing returns.

We simulate two equilibrium paths, one with linear technology as before and one with quadratic technology; i.e., we now write the contact function: $km(x_t) = R_0 x_t^2$. The rest of the parameters are as in our main simulation. With a quadratic contact function it is straightforward to derive, by substituting the quadratic expression for $m(x_t)$ into (11) and (13), that for private agents $p'(x_t^p) = R_0 x_t^p I_t$ whereas for the planner, $p'(x_t^s) = 2R_0 x_t^s I_t$. We argued that because of this difference and ignoring the dynamic externalities, the planner will want to impose more social distancing. On the other hand, dynamic externalities are present, and may dominate the externalities caused by the convexity of the contact function.[26]

[26] It follows readily from (19) that for any α, Z_{t+1} is given by

$$Z_{t+1} = -\frac{1 - x_{t+1}}{I_{t+1}}\alpha^{-1} + \beta(1 - p_{t+1})\frac{1 - x_{t+2}}{I_{t+2}}\alpha^{-1} + \beta(1 - p_{t+2})(1 - p_{t+1})Z_{t+2}.$$

In Fig. 8 we compare the paths of the decentralized equilibrium with $\alpha = 1$ and $\alpha = 2$. Since the elasticity of the contact function with respect to social activity is α, the contact function is twice as sensitive to changes in the activity level with $\alpha = 2$ than with $\alpha = 1$. Hence reducing the activity level is much more effective in reducing the contact rate with a quadratic contact function. As a result, the activity level during the midst of the epidemic is reduced by less, and at the same time the number of infected individuals is lower at its peak. The aggregate number of infections over the life span of the epidemic is also lower with $\alpha = 2$.

The effects of increasing returns on the social solution, shown in Fig. 9, are similar. However, along the optimal path, the activity level is kept slightly below 1 for a much longer period of time than when $\alpha = 1$, keeping the aggregate number of infections below $1 - S^{\max}$ for an extended period of time. As before, adjustment to the herd immunity state is much faster in the social planning solution in the relatively early stages of the epidemic, because of the race to shield and the immunity externality, which are strong with the fixed medical cost of $\delta_0 = 0.833$.

7 Conclusions

We have shown that the differences between three alternative paths from the onset of an epidemic to its eradication can be large and complex. The first path, commonly found in the early epidemiological literature, exposes large numbers of individuals to the disease early on and reaches herd immunity quickly. The other two paths, chosen by private agents in a decentralized equilibrium or by a social planner, favour much longer adjustment paths by restricting the number of infections.

In the absence of a vaccine, such a large number of susceptible individuals when infections end, produces a "race to shield," especially if the costs of becoming ill are high. With high costs, agents shield too much in the midst of the disease to increase the chances that they will be the lucky ones who will avoid the disease altogether. Obviously not everyone can succeed in this race; a low infection probability simply increases the time that it takes to end the disease through herd immunity, at a higher economic cost. The social planner will want to avoid this cost by choosing more social activity to end the epidemic faster.

The expectation of a vaccine in the near future makes a large difference to the results. Although private agents may still ignore the impact of their actions on the time it takes to reach herd immunity, the social planner is much less concerned about the potential higher economic cost, because the arrival of the vaccine will end the disease faster than herd immunity will. Private agents also ignore the fact that if they contract the disease they will infect others in the future, and this is something that the planner cares about. So with the expectation of a vaccine, the social planner imposes more social distancing, to reduce infections before the vaccine arrives. The ranking result that we derived with a race to shield reverses.

Hence for $\alpha = 2$, the first two terms are divided by 2 relative to the situation with a constant-returns to scale contact function.

We illustrate our results with simulations but as more data become available the model should be taken to the data. Our model applies generally to epidemiological models in the SIR tradition of Kermack and McKendrick (1927) and shows that quantitatively some parameters make a large difference to the simulated paths, such as the lifetime cost of the disease, the elasticity with which contacts respond to changes in social action and the vaccine arrival probability. Information on their relative magnitudes is still scant and may be disease-dependent. As our assumptions are consistent with many of the features of COVID-19, data for this epidemic could shed light on these magnitudes. Another feature that needs to be taken to the data is the economic cost of shielding. In this paper we have assumed that it is a simple convex function of a single variable, our measure of social action, but as many authors have shown, it depends on a multitude of features of occupations and economic structures.

Finally, our analysis is based on the assumption that recovered individuals stay immune permanently. If recovered individuals lose their immunity as time goes by, this will change our model in several ways. If the time that previously infected agent are immune is short, herd immunity may never be reached, and the new steady state will be characterized by equal flows of individuals moving out of and into the susceptible state. Society "lives with the disease" as it does with new variants of influenza, and we conjecture that without a vaccine the activity level never recovers to its first best level. Generally, when immunity is temporary, social distancing will be larger. In addition, the immunity externality weakens, so the planner increases social distancing more than the agents do in the decentralized solution. These are topics that should be explored in future work.

Appendix 1: The infections technology

In this Appendix we illustrate the technology of infections with a particular example that is derived from first principles and satisfies the main properties of our general function in the text. It is based on the urn-ball game that has been used in labour theory and it brings out the contrast between the labour matching function and the epidemiology transmission function.[27]

Suppose that a social activity is performed in one of N social spaces. By social space we mean a place where performing an activity requires contact with at least one other person. For a given population size $S + I + R$, N is a measure of the density of the community, with smaller N indicating a more dense community. Social distancing is also related to this measure: bigger N makes social distancing easier to achieve. A contact in social space between a susceptible and an infected individual infects the susceptible individual with probability $k \in [0, 1]$.

We now interpret social spaces as urns. Infected individuals hold white balls and susceptible individuals hold black balls. They all place $\bar{x}_t$ balls each in randomly selected urns. To simplify the exposition we assume that there is no memory of where a previous ball was placed, so each person places each one of its balls in a randomly selected urn out of the N available. A susceptible person gets infected with probability k if any one of the urns that she selected for her black balls contains one or more white

[27] See Petrongolo and Pissarides (2001) for discussion of the use of this game in labour theory.

balls.[28] We are interested in deriving the probability that there will be a white ball in at least one of the x_t urns selected by the $i \in S$ individual, given the $\bar{x}_t$ selected by the infected individuals.

Our assumption of no memory makes the problem equivalent to placing $\bar{x}_t I_t$ white balls at random in urns. The probability that an urn avoids a given white ball is $1-1/N$, so, since there are $\bar{x}_t I_t$ white balls, the probability that an urn contains no white balls after all have been placed is,

$$\begin{aligned} h_t &= \left(1 - \frac{1}{N}\right)^{\bar{x}_t I_t} \\ &= e^{-\bar{x}_t I_t/N}. \end{aligned} \tag{31}$$

Because of large numbers, h_t is also the fraction of social spaces that are infection-free (healthy).

We consider now how the choices of a single susceptible agent influence the probability that the agent will get infected. The agent selects x_t urns to place black balls. The probability that a single ball avoids an urn containing a white ball is given by (31). So the probability that all x_t black balls avoid an urn containing a white ball is

$$h_{it} = \left(e^{-\bar{x}_t I_t/N}\right)^{x_t}. \tag{32}$$

It follows that $1-h_{it}$ is the probability that a single agent meets an infected individual and so the probability that this person gets infected in period t is,

$$p_{t+1} = k\left(1 - e^{-x_t \bar{x}_t I_t/N}\right). \tag{33}$$

Differentiation of p_{t+1} yields,

$$\frac{\partial p_{t+1}}{\partial x_t}\frac{x_t}{p_{t+1}} = \frac{x_t \bar{x}_t I_t}{N}\frac{e^{-x_t \bar{x}_t I_t/N}}{1 - e^{-x_t \bar{x}_t I_t/N}}. \tag{34}$$

This is a number less than 1, in contrast to the formulation in the text, which gives unit elasticity. For small x_t and I_t it is approximately equal to 1.[29]

From (33), and since there are S_t susceptible individuals who choose $x_t = \bar{x}_t$, the aggregate infections function is

$$M_{t+1} = kS_t\left(1 - e^{-x_t^2 I_t/N}\right). \tag{35}$$

[28] Here is the biggest contrast with the labour matching function. A job vacancy (read infected person) can remove at most one unemployed worker from the pool of unemployment (read, susceptible). An infected person can remove any number that comes into contact with them.

[29] An alternative formulation replicates the method used in the text exactly. From (31) the fraction of black balls (susceptible people) which are placed in urns that contain at least one white ball is approximately $\bar{x}_t S_t(1 - \exp(\bar{x}_t I_t/N))$. A single individual supplies a fraction $x_t/\bar{x}_t S_t$ of black balls, so the probability that this person is infected is $kx_t(1 - \exp(\bar{x}_t I_t/N))$, giving the proportionality between p_{t+1} and x_t.

For fixed N, this function exhibits increasing returns to scale in S_t and I_t. This result is saying that bigger social spaces that have proportionally more susceptible and more infectious individuals do not have a higher infections rate, but a proportional increase in susceptible and infectious individuals in a given social space does lead to proportionally more infections. Analogous to this result is the role of N. As it measures density, a lower N indicates a more dense community and a higher infections rate for given S_t and I_t.

Another property of significance is the dependence of the aggregate infections rate on the square of the social activities of susceptible and infected individuals. For small numbers of $x_t I_t/N$, as satisfied by our model, the elasticity of infections with respect to social action is approximately 2 but for larger numbers it is lower.

Appendix 2

Proof of Proposition 1

In an arbitrary period t, let (S_t^o, I_t^o, p_t^o) be an arbitrary, feasible triple of values of S_t, I_t, and p_t, with $p_t^o = \frac{I_t^o}{S_t^o + I_t^o}$. From this arbitrary starting point, let the sequence $\left\{S_z^o, I_z^o, p_z^o, x_z^o\right\}_{z=t}^{\infty}$ solve the planner's problem.

Consider first an increase in the initial condition I_t from I_t^o to $I_t^o + \rho$, while S_t and p_t stay fixed at S_t^o and p_t^o, respectively.[30]

Define the function $x^I(\rho)$ implicitly by the function $km(x^I(\rho))(I_t^o + \rho)S_t^o = I_{t+1}^o$. This function exists on an interval $[-\bar{\rho}_I, \bar{\rho}_I]$ for some $\bar{\rho}_I > 0$. Clearly $x^I(0) = x_t^o$. It follows that if the economy starts at $I_t^o + \rho$, S_t^o, and the activity level is $x^I(\rho)$, then $S_{t+1} = S_{t+1}^o$, $I_{t+1} = I_{t+1}^o$, and $p_{t+1} = p_{t+1}^o$. Furthermore,

$$\frac{dx^I(0)}{d\rho} = -\frac{m(x_t^o)}{m'(x_t^o)I_t^o} \tag{36}$$

Now define the function $\tilde{W}^V(S_t^o, I_t, p_t^o)$ (where S_t^o and p_t^o are fixed, so this is a function of I_t only) for $I^t \in (I_t^o - \bar{\rho}_I, I_t^o + \bar{\rho}_I)$ as

$$\begin{aligned}\tilde{W}^V(S_t^o, I_t, p_t^o) = \phi(x^I(I_t - I_t^o)) + \beta p_t^o W^R(1 - \delta(I_t)) \\ +\beta(1 - p_t^o)W_{t+1}^V(S_{t+1}^o, I_{t+1}, p_{t+1}^o)\end{aligned} \tag{37}$$

$W^V(S_t^o, I_t, p_t^o)$ is the pay-off in optimum, and $x^I(0) = x_t^o$, it follows that

[30] Along any path, we must have that $p_t = \frac{I_t}{S_t + I_t}$ for all t. However, mathematically, the planner's maximization problem is well defined for also for initial values S_t, I_t, p_t such that $p_t \neq \frac{I_t}{S_t + I_t}$. The dynamic equations ensure that $p_z = \frac{I_z}{S_z + I_z}$ for any $z > t$. Recall further that the effect of a change in I_t and S_t through p_t is captured by the term $\frac{\partial W^V}{\partial p_t}$.

$$\begin{aligned}\tilde{W}^V(S_t^o, I_t, p_t^o) &\le W^V(S_t^o, I_t, p_t^o) \quad \forall I_t \in (I_t^o - \bar{\rho}_I, I_t^o + \bar{\rho}_I)\\ \tilde{W}^V(S_t^o, I_t^o, p_t^o) &= W^V(S_t^o, I_t^o, p_t^o)\end{aligned} \tag{38}$$

Hence the Benveniste–Scheinkman theorem applies, and we know that

$$\frac{\partial \tilde{W}^V(S_t^o, I_t^o, p_t^o)}{\partial I_t} = \frac{\partial W^V(S_t^o, I_t^o, p_t^o)}{\partial I_t} \tag{39}$$

From this equation, and (36) and (37), it follows that

$$\frac{\partial W^V(S_t^o, I_t^o, p_t^o)}{\partial I_t} = -\phi'(x_t^o)\frac{m(x_t^o)}{m'(x_t^o)I_t^o} - \beta p_t \delta'(I_t^o)V^R \tag{40}$$

Since the starting point S_t^o, I_t^o, p_t^o is arbitrary, this shows (17).
Next, consider an increase in S_t from S_t^o to $S_t^o + \rho$, while I_t and p_t stay fixed at I_t^o, p_t^o. Define $x^S(\rho)$ implicitly by the equation $(S_t^o + \rho) - km(x^S(\rho))I_t^o(S_t^o + \rho) = S_{t+1}^o$. This function exists on an interval $(-\bar{\rho}_S, \bar{\rho}_S)$ for some $\bar{\rho}_S > 0$. Taking derivatives gives

$$\frac{dx^S(0)}{d\rho} = \frac{1 - kI_t^o m(x_t^o)}{km'(x_t^o)S_t^o I_t^o} = \frac{1 - p_{t+1}^o}{p_{t+1}^{o\prime} S_t^o} \tag{41}$$

By definition we have that $S_{t+1} = S_{t+1}^o$, $I_{t+1} = I_{t+1}^o + \rho$, and $p_{t+1} = km(x^S(\rho)))I_t^o$. Now define the function $\hat{W}^V(S_t, I_t^o, p_t^o)$ for $S_t \in (S_t^o - \bar{\rho}_S, S_t^o + \bar{\rho}_S)$ as

$$\begin{aligned}\hat{W}^V(S_t, I_t^o, p_t^o) = {} & \phi(x^S(S_t - S_t^o)) + \beta p_t W^R(1 - \delta(I_t^o))\\ & + \beta(1 - p_t^o)W_{t+1}^V(S_{t+1}^o, I_{t+1}^o + S_t - S_t^o, km(x^S(S_t - S_t^o))I_t^o)\end{aligned} \tag{42}$$

Now I_t^o and p_t^o are fixed, so this is a function of S_t only. Again it follows by construction that

$$\begin{aligned}\hat{W}_t^V(S_t, I_t^o, p_t^o) &\le W_t^V(S_t, I_t^o, p_t^o) \quad \forall S_t \in (S_t^o - \bar{\rho}_S, S_t^o + \bar{\rho}_S)\\ \hat{W}_t^V(S_t^o, I_t^o) &= W_t^V(S_t^o, I_t^o)\end{aligned} \tag{43}$$

so that the Benveniste–Scheinkman theorem applies. Hence

$$\frac{\partial \hat{W}^V(S_t^o, I_t^o, p_t^o)}{\partial S_t} = \frac{\partial W^V(S_t^o, I_t^o, p_t^o)}{\partial S_t} \tag{44}$$

Using this and (42) we get that

$$\frac{\partial W^V(S_t^o, I_t^o, p_t^o)}{\partial S_t} = \frac{dx^S(0)}{d\rho}\phi'(x_t^o) + \beta(1-p_t^o)\frac{\delta W_{t+1}^V}{\delta I_{t+1}} + \beta(1-p_t^o)p_{t+1}^{o\prime}\frac{dx^S(0)}{d\rho}\frac{\delta W_{t+1}^V}{\delta p_{t+1}} \quad (45)$$

Together with (41) this shows (18).

Deriving (19)

To show (19), first insert from Eqs. (17) and (18) to get that

$$\begin{aligned} Z_{t+1} &\equiv \frac{\partial W_{t+1}^V}{\partial I_{t+1}} - \frac{\partial W_{t+1}^V}{\partial S_{t+1}} \\ &= \frac{\partial W_{t+1}^V}{\partial I_{t+1}} - \beta(1-p_{t+1})\frac{\partial W_{t+2}^V}{\partial I_{t+2}} \\ &\quad -\frac{1-p_{t+2}}{p'_{t+2}S_{t+1}}\left[\phi'(x_{t+1}) + \beta(1-p_{t+1})p'_{t+2}\frac{\partial W_{t+2}^V}{\partial p_{t+2}}\right] \end{aligned}$$

From Eq. (15) it follows that

$$\begin{aligned} \phi'(x_{t+1}) + \beta(1-p_{t+1})p'_{t+2}\frac{\partial W_{t+2}^V}{\partial p_{t+2}} &= -\beta(1-p_{t+1})\left(\frac{\partial W_{t+2}^V}{\partial I_{t+2}}\frac{\partial I_{t+2}}{\partial x_{t+1}} + \frac{\partial W_{t+2}^V}{\partial S_{t+2}}\frac{\partial S_{t+2}}{\partial x_{t+1}}\right) \\ &= -\beta(1-p_{t+1})km'(x_{t+1})I_{t+1}S_{t+1}Z_{t+2} \quad (46) \end{aligned}$$

It follows that

$$\begin{aligned} Z_{t+1} &= \frac{\partial W_{t+1}^V}{\partial I_{t+1}} - \beta(1-p_{t+1})\frac{\partial W_{t+2}^V}{\partial I_{t+2}} + \frac{1-p_{t+2}}{p'_{t+2}S_{t+1}}\beta(1-p_{t+1})km'(x_{t+1})I_{t+1}S_{t+1}Z_{t+2} \\ &= -\phi'(x_{t+1})\frac{m(x_{t+1})}{m'(x_{t+1})I_{t+1}} + \beta(1-p_{t+1})\phi'(x_{t+2})\frac{m(x_{t+2})}{m'(x_{t+2})I_{t+2}} \\ &\quad +\beta(1-p_{t+1})(1-p_{t+2})Z_{t+2} \end{aligned}$$

where we used that $p\prime_{t+2} = km'(x_{t+1})I_{t+1}$ to obtain the last equation.

Deriving (20)

To calculate Z_∞, consider the limit economy with $S = S_\infty$, $I = 0$. The optimal trajectory prescribes that $x = 1$ in all periods, and $W^V = W^R$. From this starting point, suppose I increases from 0 to dI. Along the optimal path, behavioural changes are of second order and can therefore be ignored.[31]

[31] At this point herd immunity is obtained, and the system is stable around $I_t = 0$. This is not the case initially.

A person that is infected in period t is expected to infect $km(1)S_\infty$ persons in the next period. Using the law of iterated expectations, it follows that in period $t+2$, the expected number of infected people is $(km(1)S_\infty)^2$, and using the law repeatedly, it follows that in period $t+k$ the expected number of infected people is $(km(1)S_\infty)^k$ people are expected to be infected and so on. With $I=0$, $W^V = W^R$. It follows that the total social cost of increasing the number of infected people from 0 to dI is

$$dI\left(1+\beta km(1)\bar{S}+(\beta km(1)\bar{S})^2++\right)\beta^2\delta_0 W^R,$$

where $W^R=\phi(1)(1-\beta)^{-1}$. The first term is the private cost of obtaining the illness, while the rest of the terms are the external effects. Hence the external effects sum to

$$-\frac{\beta^2 km(1)\bar{S}}{1-\beta km(1)\bar{S}}\delta_0 W^R.$$

The cost is only carried by the susceptible individuals. Recall that W^V, equal to W^S when $I=0$, captures the NPV utility of the representative susceptible individual. To obtain Z_∞ (the cost per individual), we divide by $\bar{S}$, and we get that

$$Z_\infty = -\frac{\beta^2 R_0}{1-\beta R_0\bar{S}}\delta_0 W^R. \tag{47}$$

Derivation of dynamic externalities with vaccine, Eq. (29)

Consider first an increase in I_t. If a vaccine has arrived, no-one is infected, hence from Eq. (22) it follows that

$$\frac{\partial \overline{W}^{VV}}{\partial I_t} = -\beta p_t \delta'(I_t) W^R \tag{48}$$

This happens with probability λ. With the complementary probability, a vaccine does not arrive. In this case we apply the exact same argument as when deriving $\frac{\partial W^V}{\partial I_t}$ above, Hence, parallel with (40) we get that

$$\frac{\partial W^{VN}(S_t, I_t, p_t)}{\partial I_t} = -\phi'(x_t)\frac{m(x_t)}{m'(x_t)I_t} - \beta p_t \delta'(I_t) V^R \tag{49}$$

Here, and below, x_t refers to optimal activity levels (as x_t^o in the proofs above). It thus follows that

$$\frac{\partial W^{VV}(S_t, I_t, p_t)}{\partial I_t} = -(1-\lambda)\phi'(x_t)\frac{m(x_t)}{m'(x_t)I_t} - \beta p_t \delta'(I_t) W^R \tag{50}$$

Then consider $\frac{\partial \bar{W}_t^{VV}}{\partial S_t}$. First note that $\bar{W}_t^{VV}$ is independent of S_t: If a vaccine arrives between period $t-1$ and t, no-one is infected from period t and onwards, and S_t is irrelevant. If a vaccine does not arrive, we can use the exact same procedure as in the case with no vaccine in order to find the effect on W^{VN}. Parallel with (45) we have that

$$\frac{\partial W^{VV}(S_t, I_t, p_t)}{\partial S_t} = (1-\lambda)\frac{dx^S(0)}{d\rho}\phi'(x_t) + (1-\lambda)\left\{\beta(1-p_t)\frac{\delta W^{VV}_{t+1}}{\delta I_{t+1}} + \beta(1-p_t)p^{o\prime}_{t+1}\frac{dx^S(0)}{d\rho}\frac{\delta W^{VV}_{t+1}}{\delta p_{t+1}}\right\} \tag{51}$$

where $x^S(0)$ is defined as in Eq. (45). Inserting this expression gives (29).

Simulations

The most efficient way to solve the planner's problem numerically is to use the first order conditions obtained by using Lagrange's method. The planner's objective can be written non-recursively as (with $\delta = \delta_0$ constant)[32]

$$\sum_{t=0}^{\infty}\beta^t\left((S_t + I_t)\phi(x_t) + \beta I_t W^I\right) \tag{52}$$

where $W^I = \beta(1-\delta_0)W^R$ independently of t. This is maximized given the constraints

$$S_{t+1} = S_t - km(x_t)I_t S_t \tag{53}$$

$$I_{t+1} = km(x_t)I_t S_t \tag{54}$$

The initial condition is that $S_0 + I_0 = 1$, and that $I_0 = \epsilon$.

The first order conditions can be written as

$$\frac{\phi'(x_t)}{1-p_t} = -\beta p'(x_t)(\lambda^I_{t+1} - \lambda^S_{t+1}) \tag{55}$$

$$\lambda^S_t = \phi(x_t) + \beta(1-p_{t+1})\lambda^S_{t+1} + \beta p_{t+1}\lambda^I_{t+1} \tag{56}$$

$$\lambda^I_t = \phi(x_t) + \beta^2(1-\delta_0)W^R - \beta km(x_t)S_t(\lambda^S_{t+1} - \lambda^I_{t+1}) \tag{57}$$

We solve this system nummerically in the high-δ scenario, and plug the solution into the planner's first order condition, which is satisfied with with extreme precision. The difference between the left-hand side and the right-hand side of (16) is minuscule (substantially less than 10^{-11}).

[32] It is easy to show that this is equivalent to maximize W^V_0.

References

Acemoglu, D., Chernozhukov, V., Werning, I., Whinston, M.D.: A multi-risk sir model with optimally targeted lockdown. Am. Econ. Rev.: Insights **3**(4), 487–502 (2021)

Alfaro, L., Faia, E., Lamersdorf, N., Farzad, S.: Social interactions in pandemics: fear, altruism and reciprocity. NBER Working Paper N. 27134 (2020)

Alvarez, F., Argente, D., Lippi, F.: A simple planning problem for COVID-19 lockdown. Am. Econ. Rev.: Insights **3**(3), 367–82 (2021)

Berger, D., Herkenhoff, K., Mongey, S.: A seir infectious disease model with testing and conditional quarantine. NBER Working Paper N. 27059 (2020)

Bosi, S., Carmen, C., Desmarchelier, D.: Optimal lockdown in altruistic economies. J. Math. Econ. **93**, 859–872 (2010)

Brotherhood, L., Kircher, P., Santos, C., Terlit, M.: An economic model of the COVID-19 epidemic: the importance of testing and age-specific policies. CEPR Discussion Paper N. 14695 (2020)

Chakraborty, S., Papageorgiou, C., Sebastian, F.P.: Diseases, infection dynamics, and development. J. Math. Econ. **57**, 859–872 (2010)

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

Diamond, P.: Aggregate demand management in search equilibrium. J. Polit. Econ. **90**(5), 881–994 (1982)

Diamond, P., Maskin, E.: An equilibrium analysis of search and breach of contracts, I: steady states. Bell J. Econ. **10**, 282–316 (1979)

Eichenbaum, M.S., Rebelo, S., Trabandt, M.: The macroeconomics of epidemics. Rev. Financ. Stud. **34**, 5149–5187 (2021)

Farboodi, M., Jarosch, G., Shimer, R.: Internal and external effects of social distancing in a pandemic. J. Econ. Theory **196**, 105293 (2021)

Favero, C., Ichino, A., Rustichini, A.: Restarting the economy while saving lives under COVID-19. CEPR Discussion Paper N. 14695 (2020)

Federico, S., Ferrari, G.: Taming the spread of an epidemic by lockdown policies. J. Math. Econ. **93**, 102453 (2021)

Federico, S., Ferrari, G., Torrente, M.-L.: Optimal vaccination in a sirs epidemic model. Econ. Theory (2022). https://doi.org/10.1007/s00199-022-01475-9

Garibaldi, P., Moen, E.R., Pissarides, C.A.: Static and dynamic inefficiencies in an optimizing model of epidemics. CEPR Working Paper N. 15439; IZA Discussion Paper 13844 (2020)

Garibaldi, P., Moen, E.R., Pissarides, C.A.: The SAM approach to epidemic models. In: Polachek, S.W., Tatsiramos, K. (eds.) Research in Labour Economics, 50th Celebratory Volume, pp. 1–23. Emerald Publishing Limited (2023)

Geoffard, P.-Y., Philipson, T.: Rational epidemics and their public control. Int. Econ. Rev. **37**(3), 603–624 (1996)

Goenka, A., Lin, L.: Infectious diseases and endogenous fluctuations. Econ. Theory **50**, 125–149 (2012)

Goenka, A., Liu, L., Nguyen, M.: Modelling optimal lockdowns with waning immunity. Econo. Theory (2022). https://doi.org/10.1007/s00199-022-01468-8

Greenwood, J., Kircher, P., Santos, C., Tertil, Mt.: An equilibrium model of the African hiv/aids epidemic. Econometrica **87**(4), 1081–113 (2019)

Gregory, V., Menzio, G., Wiczer, D.: Pandemic recession: L or v shaped? J. Math. Econ. **40**, 1–28 (2020)

Hall, R.H., Charles, J.I., Klenow, P.J.: Trading off consumption and COVID-19 deaths. Fed. Reserv. Bank Minneap. Q. Rev. **42**(1), 2–13 (2020)

Hosios, A.J.: On the efficiency of matching and related models of search and unemployment. Rev. Econ. Stud. **57**(5), 279–298 (1990)

Hritonenko, N., Yatsenko, Y.: Analysis of optimal lockdown in integral economic-epidemic model. Econ. Theory (2022). https://doi.org/10.1007/s00199-022-01469-7

Jones, C., Philippon, T., Venkateswaran, V.: Optimal mitigation policies in a pandemic: social distancing and working from home. Rev. Financ. Stud. **34**(11), 5188–5223 (2020)
Kapicka, M., Rupert, P.: Labor markets during pandemics. Working Paper (2020)
Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Serie A Contain. Pap. Math. Phys. Charact. **115**(772), 700–721 (1927)
Kremer, M.: Integrating behavioral choice into epidemiological models of AIDS. Q. J. Econ. **111**(2), 549–573 (1996)
La Torre, D., Liuzzi, D., Marsiglio, S.: Epidemics and macroeconomic outcomes: social distancing intensity and duration. J. Math. Econ. **93**, 102473 (2021)
Lebeau, L.: Social engagement and the spread of infectious diseases. Working Paper (2020)
Makris, M.: COVID and social distancing with a heterogenous population. Econ. Theory (2021). https://doi.org/10.1007/s00199-021-01377-2
Nævdal, E.: Epidemics and increasing returns to scale on social distancing. COVID Econ.: Vetted Real Time Pap. **39**, 139–148 (2020)
Petrongolo, B., Pissarides, C.A.: A survey of the matching. J. Econ. Lit. **39**(2), 390–431 (2001)
Pissarides, C.A.: Equilibrium Unemployment Theory. MIT Press, Cambridge (2000)
Quercioli, E., Smith, L.: Contagious matching games. Working Paper (2006)
Rachel, L.: An analytical model of covid-19 lockdowns. Working Paper (2020)
Reluga, T.C.: Game theory of social distancing in response to an epidemic. PLOS Comput. Biol. **6**(5), 1–9 (2010)
Rowthorn, R., Toxvaerd, F.: The optimal control of infectious diseases via prevention and treatment. CEPR Discussion Paper N. 8925 (2012)
von Thadden, E.-L.: A simple, non-recursive model of the spread of covid-19 with applications to policy. SSRN Working Paper N. 3580767 (2020)