CARF

CENTER FOR
ADVANCED
RESEARCH
IN FINANCE

C A R F   W o r k i n g   P a p e r

CARF-F-546

Heterogeneous Risk Attitudes and Waves of Infection

Daisuke Fujii
University of Tokyo

Taisuke Nakata
University of Tokyo

Takeshi Ojima
Soka University

November 6, 2022

CARF Working Papers can be downloaded without charge from:
https://www.carf.e.u-tokyo.ac.jp/research/

Heterogeneous Risk Attitudes and Waves of Infection [*]

Daisuke Fujii[†]
University of Tokyo

Taisuke Nakata[‡]
University of Tokyo

Takeshi Ojima[§]
Soka University

November 6, 2022

Abstract

Many countries have experienced multiple waves of infection during the COVID-19 pandemic. We propose a novel extension of the SIR model, a CSIR model, that can endogenously generate waves. In the model, cautious individuals take appropriate prevention measures against the virus and are not exposed to infection risk. Incautious individuals do not take any measures and are susceptible to the risk of infection. Depending on the size of incautious and susceptible population, some cautious people lower their guard and become incautious—thus susceptible to the virus. When the virus spreads sufficiently, the population reaches "temporary" herd immunity and infection subsides thereafter. Yet, the inflow from the cautious to the susceptible eventually expands the susceptible population and leads to the next wave.

Keywords: COVID-19, epidemiology, SIR model, waves, periodic solution

JEL classification: C61, C65, E71, I18

[†]Email: dfujii@e.u-tokyo.ac.jp
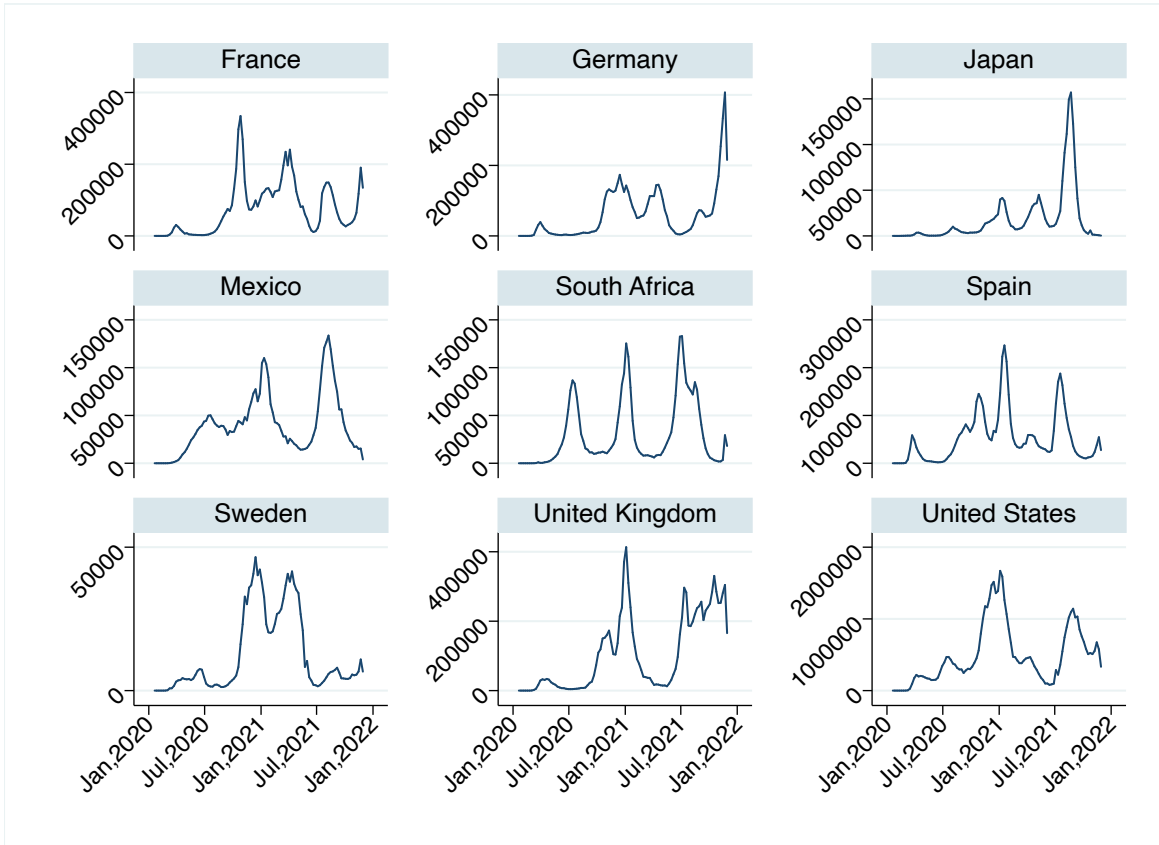[‡]Email: taisuke.nakata@e.u-tokyo.ac.jp
[§]Email: ojima@soka.ac.jp

Figure 1: Weekly number of new COVID-19 cases between January 2020 and November 2021 (Source: WHO Coronavirus Dashboard)

1 Introduction

The COVID-19 pandemic caused by a coronavirus SARS-CoV-2 rapidly spread around the globe posing unprecedented health and economic challenges to mankind. Since its outbreak, the pandemic has given rise to waves of infection in many countries, as shown in in Figure 1.[1] The susceptible-infectious-removed (SIR) model, one of the most fundamental and widely-used epidemic models, cannot generate waves of infection in its purest form.

In this paper, we propose a novel modification to the SIR model, a "CSIR" model, to describe infection waves with constant population. The structure of our model is straightforward, yet it generates complex dynamics of infection. Key features of our CSIR model are that people have different risk attitudes toward the virus and that these attitudes vary over time. In the standard SIR model, those who have not been infected are homogeneous in terms of their risk attitudes.

Our model starts when a new virus emerges, all members of society learn about it, and

---

[1]A notable exception is China, which seems to have contained the infection in the early phase of pandemic with strong lockdown policies.

some individuals become "cautious". Those cautious individuals (compartment C) take appropriate measures against the virus—such as staying at home and wearing masks—and are not subject to infection risk. Susceptible individuals are less vigilant about the virus and vulnerable to infection. Depending on the size of the susceptible population, some cautious people lower their guard and become less vigilant against, and hence susceptible to, the virus. When the virus spreads, it eventually reaches "local" herd immunity and subsides thereafter. Nevertheless, the inflow from the cautious to the susceptible regenerates the expansion of the susceptible population leading to the next wave of infection. This repeated process creates a finite number of infection waves.

The multiplicity of waves in the CSIR model before reaching the steady state means that the size of each wave is smaller than the size of the wave in the SIR model when the parameter of the CSIR model is calibrated to match the final epidemic size in the SIR model. During the COVID-19 pandemic, the peak of many infection waves was substantially lower than what the standard SIR would predict, even after taking into account factors such as vaccination and NPIs. Our model provides a theory of why that was the case that relies on heterogeneity in risk attitudes and their time-variation.

In addition to its ability to generate waves, the CSIR model possesses other intriguing features that are distinct from those of the standard SIR model. For example, we show via numerical analysis that a higher transmission rate leads to lower peaks of early infection waves and higher peaks for late infection waves in the CSIR model. In the SIR model, a higher transmission rate leads a higher peak. We also show, analytically, that a higher transmission rate implies a lower epidemic final size in the CSIR model, whereas it implies a higher final size in the SIR model.

As reviewed in Section 5, the existing literature has provided a variety of extensions to—and modifications of—the standard SIR model that can generate waves—or oscillations in the terminology often used in the literature. Examples include reinfection, age-structure, fear effects, seasonality, and network structures, among others. At the same time, beyond the academic literature, there are several factors that might have contributed to the generation of waves during the COVID-19 pandemic, such as the periodic emergence of a more infectious variant of coronavirus SARS-CoV-2 and government lockdown policies. Our CSIR model featuring heterogeneous risk attitudes and their time-variation provides the literature with a novel mechanism for multiple waves as well as a plausible factor that might have contributed to the emergence of multiple waves during the COVID-19 pandemic.

This paper is organized as follows. The next section reviews the standard SIR model and shows why waves do not occur in its purest form. Section 3 presents the CSIR model

and conditions to generate infection waves. Analytic properties are also discussed. Section 4.1 highlights the difference between SIR and CSIR models when a more infectious virus emerges. Section 6 concludes.

2 Review of the SIR Model

This section reviews the standard SIR model introduced in Kermack and McKendrick (1927) and explains why it does not explain the waves of infection. Consider the following SIR model

$$
\frac{dS(t)}{dt} = -\beta S(t) I(t)
$$
$$
\frac{dI(t)}{dt} = \beta S(t) I(t) - \gamma I(t)
$$
$$
\frac{dR(t)}{dt} = \gamma I(t),
$$

where S, I, and R are susceptible, infectious, and removed populations respectively. Let $N(t) = \beta S(t) I(t)$ denote the number of new cases at time t. The removed population can also be divided into recovered and dead classes, but we do not focus on the difference in this paper since $R(t)$ does not affect the dynamics of the system. The parameter β is called a transmission rate, and governs the infectious force. γ denotes the removal rate, where $1/\gamma$ is the average duration of infectiousness. We normalize total population to one

$$
S(t) + I(t) + R(t) = 1
$$

so that S, I, and R are corresponding population shares. Total population is preserved at any time since $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$. The basic reproduction number is defined as $\Re_0 = \frac{\beta}{\gamma}$. From the expression of $\frac{dI}{dt}$, if $\Re_0 < 1$, then $I(t)$ is decreasing regardless of $S(t)$ and there is no outbreak of infection. If $\Re_0 > 1$, an outbreak occurs. In that case, it is useful to focus on the effective reproduction number $\Re_t = \Re_0 \times S(t)$. If $\Re_t > 1$, infection keeps spreading whereas if $\Re_t < 1$, the infectious population will decrease. Since $S(t)$ is monotonically decreasing, when it crosses $\frac{1}{\Re_0} = \frac{\gamma}{\beta}$, the pandemic will converge to an end. Hence, $\frac{1}{\Re_0}$ is called the herd immunity threshold.

We discretize the model and interpret time at a daily frequency. Throughout this paper, we assume $\gamma = 0.05$ implying that the average infectious period is 20 days. Figure 2 depicts the time paths of S, I, R and N (new cases) for the initial values $[S(0), I(0), R(0)] =$
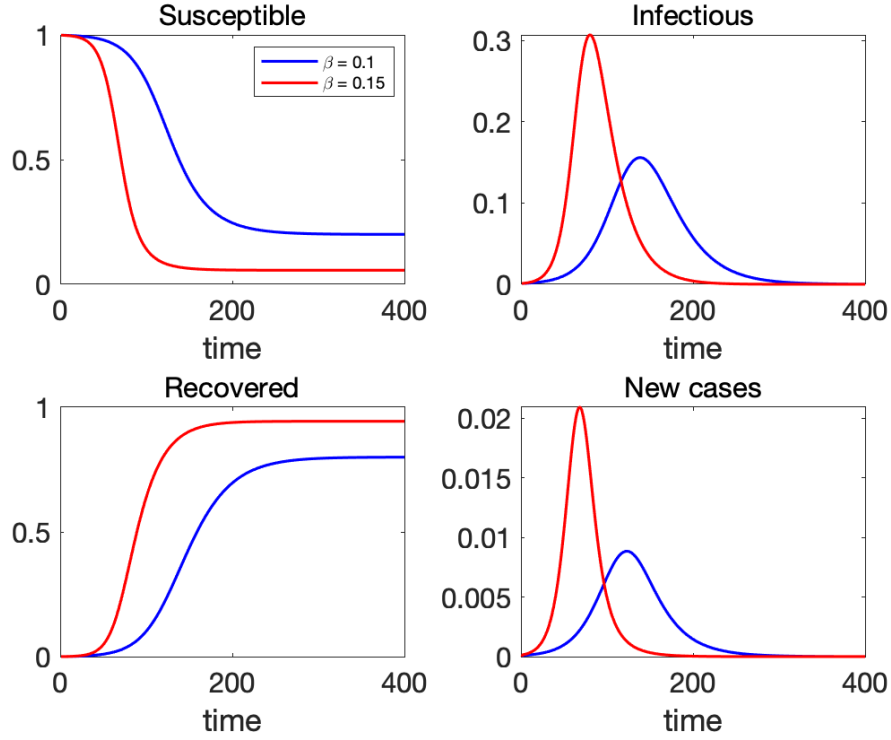
Figure 2: Time paths of each population in the SIR model with $\gamma = 0.05$ and $\beta = 0.1$ (blue) and $\beta = 0.15$ (red)

$[0.999, 0.001, 0]$ and two values of transmission rate: $\beta = 0.1$ and $\beta = 0.15$.[2] In the graphs of I and N, we observe only one peak of infection in both cases. If the virus is more contagious (red line), the peak of infection is higher and comes earlier. Also, the cumulative number of infected people ($R(\infty)$) will be higher if β is higher.

Figure 3 shows the phase diagram of S and I in the SIR model. The dynamics start from the bottom right corner. Since $\frac{\beta}{\gamma} > 1$ and the number of susceptible people is large enough, infection spreads. $I(t)$ goes up and $S(t)$ goes down. When $S(t)$ crosses $\frac{\gamma}{\beta}$, $I(t)$ starts decreasing and the system will converge to the steady state with $I(\infty) = 0$. Since $S(t)$ is monotonically decreasing, after crossing the herd immunity threshold, it cannot generate the second wage of infection.

If we linearize the dynamic system around the steady state $I(t) = 0$, we obtain

$$
\begin{bmatrix} S(t+1) \\ I(t+1) \end{bmatrix} = \begin{bmatrix} 1 & -\beta S \\ 0 & 1 + \beta S - \gamma \end{bmatrix} \begin{bmatrix} S(t) \\ I(t) \end{bmatrix}.
$$

The eigenvalues of the Jacobian matrix are $\lambda_1 = 1$ and $\lambda_2 = 1 + \beta S - \gamma$. Since they are not

---

[2]We discretize a differential equation of a variable X by replacing $\frac{dX(t)}{dt}$ with $X(t+1) - X(t)$.
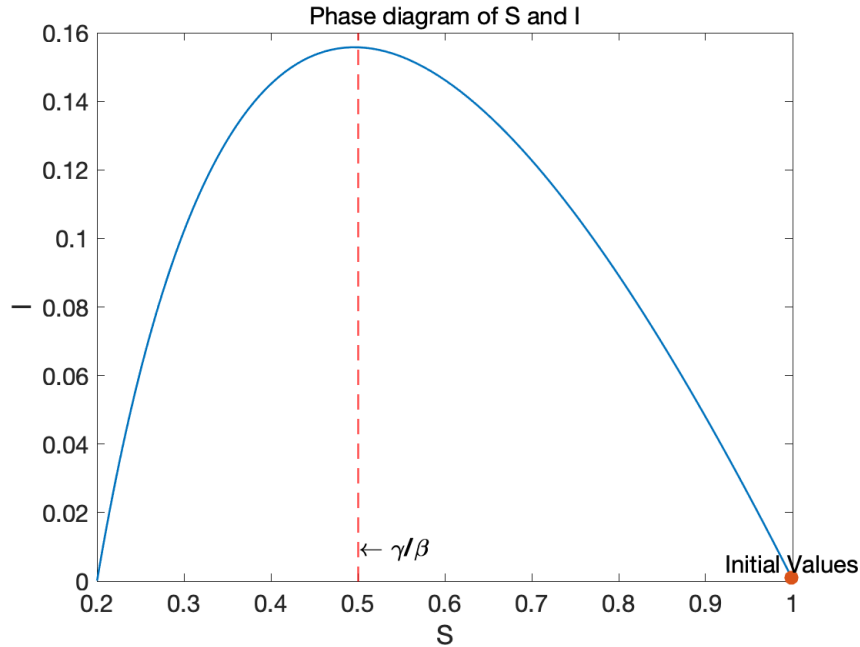
Figure 3: Phase diagram of the SIR model with $\beta = 0.1$ and $\gamma = 0.05$

complex conjugate, we know that this system generates only one peak of infection.

As reviewed in Section 5, the existing studies have shown that waves can emerge in modified versions of the standard SIR model. In the sections that follow, we will present a novel modification that features heterogeneity in risk attitudes and their time-variation.

3   CSIR Model and Waves of Infection

3.1   Model Description

In this section, we present the CSIR model and its properties. As described in Section 1, we introduce a new class of population, the cautious people, to the standard SIR model. We will denote the groups of cautious people by "C." The dynamics are described by the following system of differential equations:
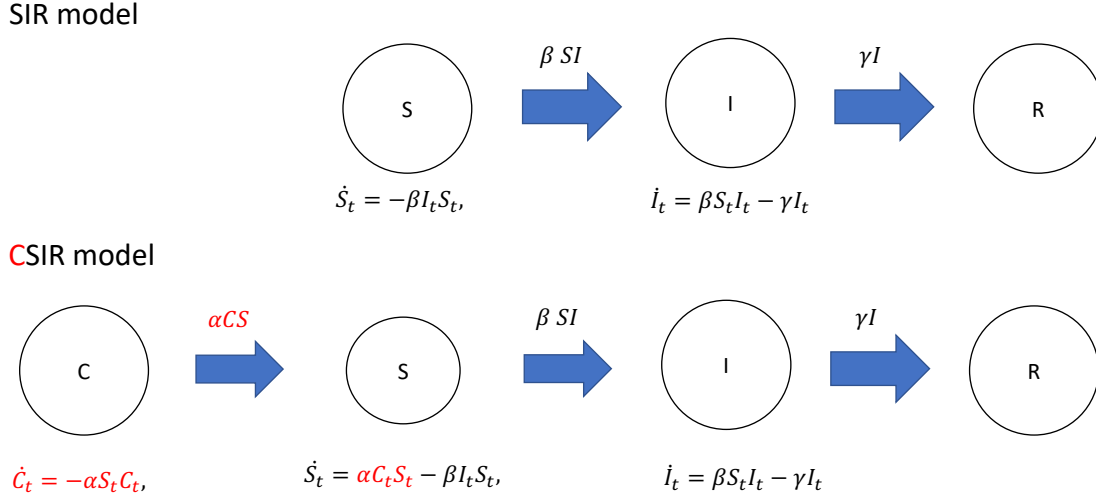
SIR model

$$\dot{S}_t = -\beta I_t S_t, \qquad\qquad \dot{I}_t = \beta S_t I_t - \gamma I_t$$

CSIR model

$$\dot{C}_t = -\alpha S_t C_t, \qquad \dot{S}_t = \alpha C_t S_t - \beta I_t S_t, \qquad \dot{I}_t = \beta S_t I_t - \gamma I_t$$

Figure 4: Flow chart of SIR and CSIR models

$$\frac{dC(t)}{dt} = -\alpha S(t) C(t) \tag{1}$$

$$\frac{dS(t)}{dt} = \alpha S(t) C(t) - \beta S(t) I(t) \tag{2}$$

$$\frac{dI(t)}{dt} = \beta S(t) I(t) - \gamma I(t) \tag{3}$$

$$\frac{dR(t)}{dt} = \gamma I(t). \tag{4}$$

In the model, cautious individuals (compartment C) take appropriate prevention measures against virus such as staying at home and are not exposed to infection risk. Susceptible individuals are incautious about virus and might get infected. In every period, some fraction of the cautious people lower their guard and become susceptible. The transition from C to S is modeled by $\alpha S(t) C(t)$. There is a peer effect of S on the outflow of C. The more incautious people go out into the town, the more cautious people relax their efforts and become incautious. We call α "a rate of slackening". Unlike the SIR model, there are both inflow and outflow of $S(t)$, which allow the volume of the susceptible to expand and shrink. Flow charts of the SIR and CSIR models are summarized in Figure 4.

3.2 Waves of Infection

Figure 5 depicts the time paths of the four variables with initial values $C(0) = 0.95$, $S(0) = 0.0495$, $I(0) = 0.0005$, $R(0) = 0$, and parameters $[\alpha, \beta, \gamma] = [0.007, 1, 0.05]$ for
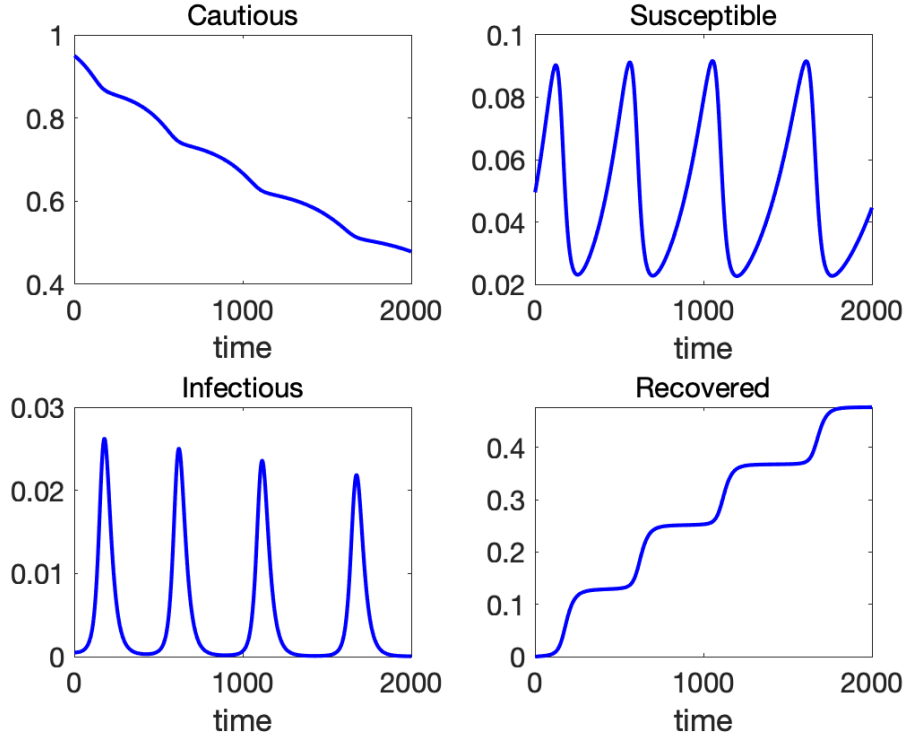
Figure 5: Time paths of C, S, I and R with $\alpha = 0.007$, $\beta = 1$, and $\gamma = 0.05$

$T = 2000$. We assume that the news of an outbreak is shared by all individuals immediately so that 95% of people are cautious about the virus in the initial period. Around five percent of population does not fear the risk of infection. Gradually, some of the cautious people loosen infection control measures and flow into the susceptible class. Infection spreads among the susceptible and reaches its first peak around the 200th day (bottom-left panel of Figure 5). At this point, there are not enough susceptible people to spread infection further, and the volume of infected population starts to decline leading to "local" or "temporary" herd immunity. When the number of the infected remains at a low level for a while, the susceptible population expands because the inflow from the cautious class outweighs the outflow to the infected class. This expansion of the susceptible population results in the next wave of infection. The above process repeats over time and generates multiple waves of infection. In Appendix A, we consider other specifications of $\frac{dC(t)}{dt}$ and $\frac{dS(t)}{dt}$, and show that the generation of waves is a general feature of the CSIR model.

In the early phase of the COVID-19 pandemic, infection begun to drop at a much lower level than the herd immunity threshold predicted by the SIR model in many countries. It is widely accepted that non-pharmaceutical interventions (NPIs) such as stay-at-home or-
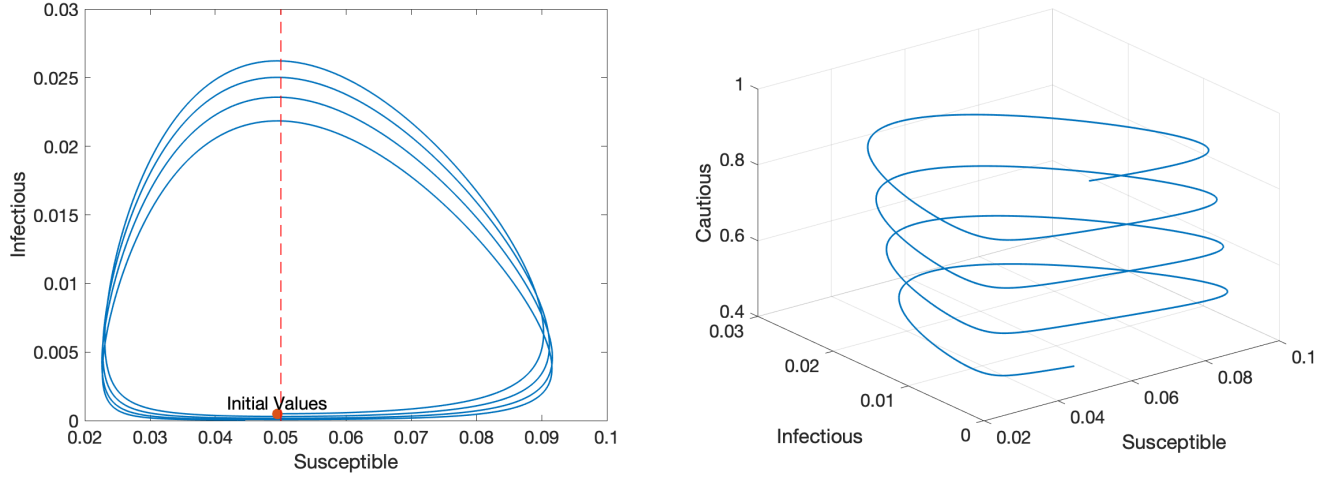
8

Figure 6: Phase diagrams of a CSIR model

ders and city-wide lockdowns played a key role to level off infection. We provide an alternative mechanism to explain the earlier and lower peak of infection. In the CSIR model, the spread of infection starts from a certain fraction of population who are incautious about virus, and once enough of them are infected, infection starts to decline. This threshold is much lower than that of the SIR model which assumes all agents are homogeneous and equally mixed. Yet, because risk attitudes of people change over time, the CSIR model generates multiple waves. It is necessary that $C(0)$ is sufficiently large and $C(t)$ gradually declines over time for the CSIR model to generate infection waves.[3] This trend is related to the misperception of infection risks considered in the SIR-macro model of Hamano et al. (2020). In their model, people are overly cautious in the early periods of pandemic but gradually let their guard down as more information becomes available to explain the actual path of infection in Japan.

Figure 6 illustrates the phase diagrams of the CSIR model. The left panel is the phase diagram on a S-I plane. The dynamics start at the red dot and move counterclockwise. From the equation (2), we know that S is increasing when $I < \frac{\alpha}{\beta}C$, and vice versa. Likewise from equation (3), I is increasing when $S > \frac{\gamma}{\beta}$. At the initial point, I is so small that S starts increasing. Once S becomes larger than $\frac{\gamma}{\beta}$, I starts increasing (southeast part of the graph). When I becomes sufficiently large ($I > \frac{\alpha}{\beta}C$), S starts decreasing (northeast part of the graph). Once S crosses the temporary herd immunity threshold $\frac{\gamma}{\beta}$ depicted by a vertical red dotted line, I starts decreasing (northwest part of the graph). When I shrinks sufficiently, S starts increasing again (southwest part of the graph). This cycle repeats and

---

[3]In Appendix **??**, we examine the number of waves for different values of $C(0)$ with the same parameters. When $C(0)$ is below 0.7, multiple waves do not occur within 2000 periods. The extreme case of $C(0) = 0$ collapses to the SIR model, so there would not be any waves as shown in the previous section.
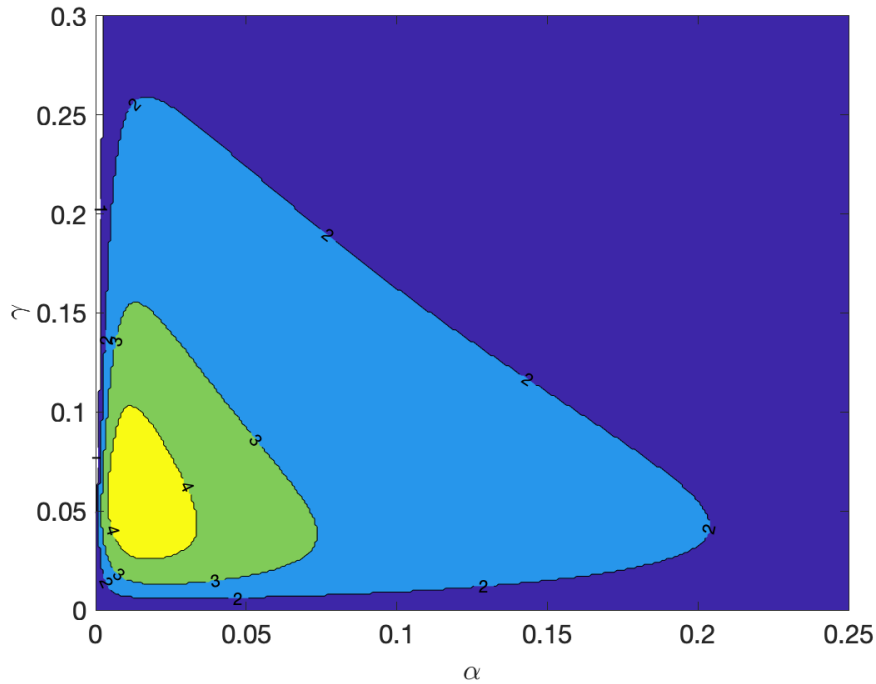
Figure 7: Contour plot for the number of waves for $T = 2000$ with different values of α and γ. Initial values are $C(0) = 0.95$, $S(0) = 0.0495$, $I(0) = 0.0005$, $R(0) = 0$ and $\beta = 1$.

gives rise to the waves of infection. The right panel of Figure 3 is a 3-D phase diagram in a S-I-C space. The dynamics start at the top and move down a helix. The left panel corresponds to an upper view of this 3-D phase diagram.

Given an initial condition, the number of waves generated by the CSIR system depends on parameter values. Figure 7 displays a contour plot for the number of waves within 2000 periods with different values of α and γ holding $\beta = 1$ and the same initial condition as in Figure 5.[4] If α and γ are relatively small, four waves occur. When α is large, only one peak is observed because the inflow from C to S is too rapid that there will not be enough C left to generate the next wave after the first peak. As α becomes large, the CSIR model approaches to the SIR model since most of the population starts from the state S. A larger value of γ also implies a lower number of waves because the volume of I shrinks too quickly, and it will take time to pile up enough volume of S to generate the next wave.

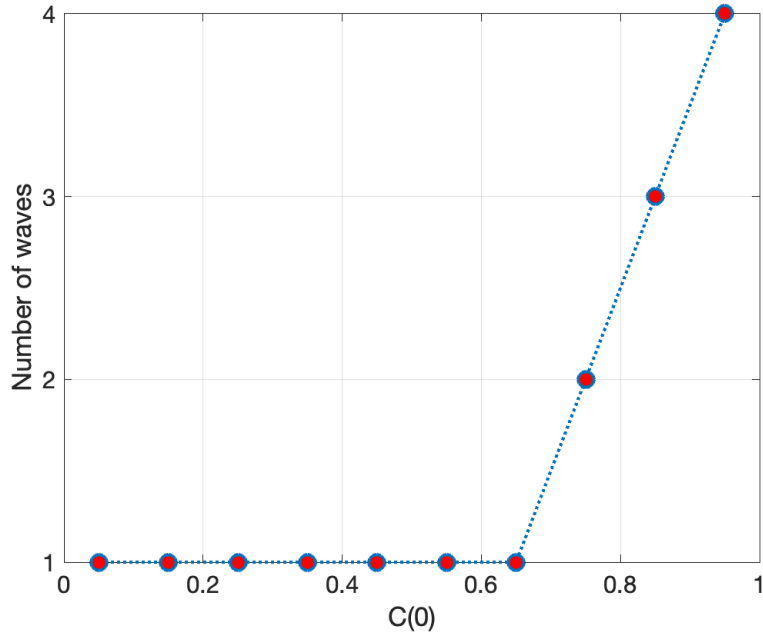[4]Waves are defined as the number of peaks of I.

Figure 8: Initial population of the cautious $C(0)$ and the number of waves

The above analysis also reveals that the initial population of the cautious group must be large enough to generate multiple waves. Figure 8 plots the number of infection waves in the CSIR model for different initial values of $C(0)$ ranging from 0.05 to 0.95. The parameters are $[\alpha, \beta, \gamma] = [0.05, 1, 0.07]$ and $T = 2000$. For each $C(0)$, other initial values are set as $S(0) = [1 - C(0)] \times 0.99$, $I(0) = [1 - C(0)] \times 0.01$, and $R(0) = 0$. If $C(0)$ is below 0.7, the model cannot generate any waves.

3.3 Steady-State Paths

Unlike the SIR model, the CSIR model described by equations (1) - (4) does not reach to the steady-state level of S. Nonetheless, further analytical investigations are possible by considering a new variable, and they provide more insights into the CSIR model.

Let $K_t = \frac{I_t}{C_t}$.[5] The law of motion of this new variable is given by

$$\frac{dK_t}{dt} = (\alpha + \beta) K_t S_t - \gamma K_t.$$

---

[5]Time subscripts are used to simplify notations.

11

$$\frac{dI_t}{I_t} < \frac{dC_t}{C_t} \quad \underline{\text{Infection shrinks}} \qquad \frac{dI_t}{I_t} > \frac{dC_t}{C_t} \quad \underline{\text{Infection spreads}}$$

Cautious

Inflow to S $<$ outflow from S

$\frac{\alpha}{\beta}$

Inflow to S $>$ outflow from S
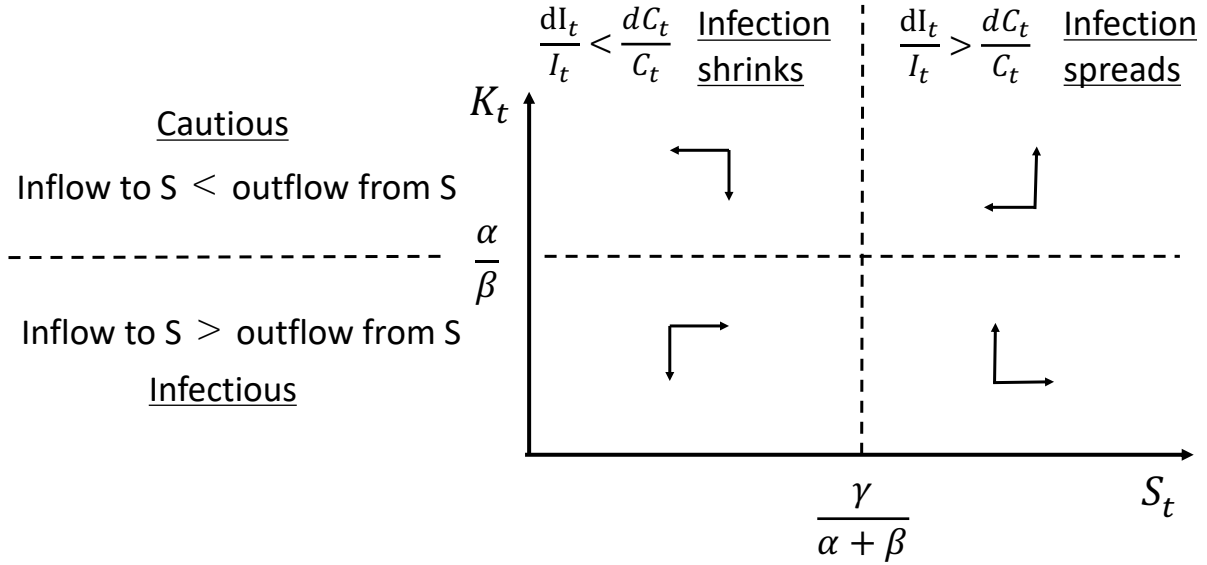
Infectious

$\frac{\gamma}{\alpha+\beta}$

Figure 9: Phase diagram of S and $K = \frac{I}{C}$

[6] The differential equation for S_t can be written as

$$\frac{dS_t}{dt} = (\alpha - \beta K_t)\, S_t C_t$$

The system of two differential equations $\frac{dK_t}{dt}$ and $\frac{dS_t}{dt}$ has an equilibrium point characterized by the following conditions

$$K^* \left[(\alpha + \beta)\, S^* - \gamma \right] = 0$$
$$S^* \left[\alpha - \beta K^* \right] = 0$$

Thus, the equilibrium point is given by

$$K^* = \frac{\alpha}{\beta} \text{ and } S^* = \frac{\gamma}{\alpha + \beta}.$$

The phase diagram of this system is described in Figure 9. When K_t is larger than $\frac{\alpha}{\beta}$, S_t is decreasing, and vice versa. When S_t is larger than $\frac{\gamma}{\alpha+\beta}$, K_t is increasing, and vice versa. At

---

[6]We have

$$dK_t = d\left(\frac{I_t}{C_t}\right) = \frac{dI_t C_t - I_t dC_t}{C_t^2} = \frac{(\beta S_t I_t - \gamma I_t)\, C_t - I_t\, (-\alpha S_t C_t)}{C_t^2} dt$$
$$= \frac{(\beta S_t I_t - \gamma I_t) + \alpha I_t S_t}{C_t} dt = \left[\beta \left(\frac{I_t}{C_t}\right) S_t - \gamma \left(\frac{I_t}{C_t}\right) + \alpha \left(\frac{I_t}{C_t}\right) S \right] dt.$$
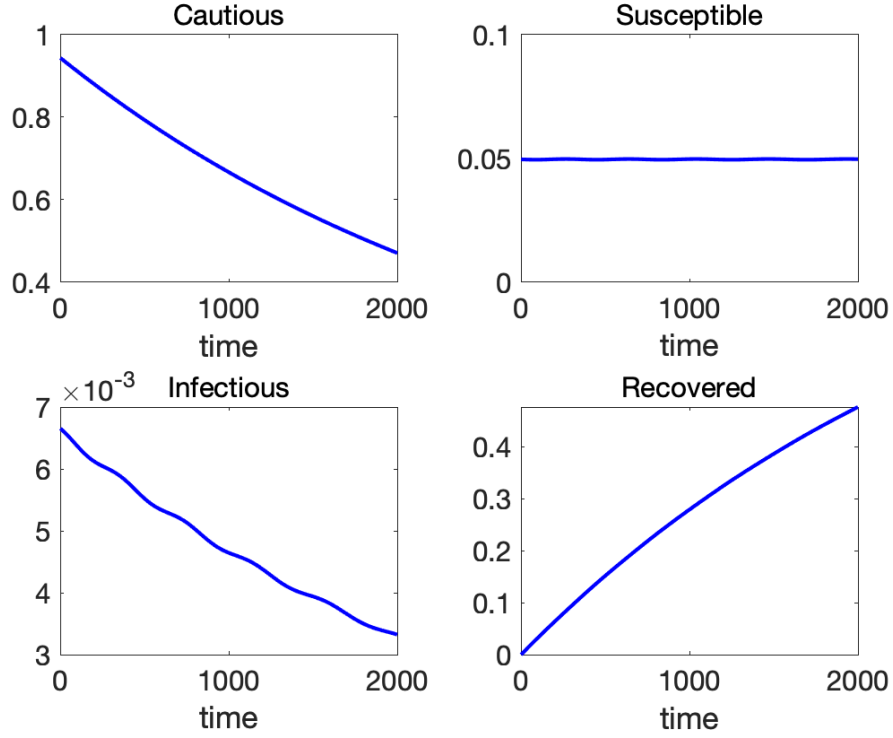
Figure 10: Steady-state paths of the CSIR model when $\frac{I_0}{C_0} = \frac{\alpha}{\beta}$ and $S_0 = \frac{\gamma}{\alpha+\beta}$.

the steady state $(S^*, K^*) = \left(\frac{\gamma}{\alpha+\beta}, \frac{\alpha}{\beta}\right)$, S^* remains constant and C_t and I_t are decreasing at the same rate so that $\frac{I_t}{C_t}$ is constant. Figure 10 illustrates the steady-state paths of the CSIR model when the initial conditions correspond to the equilibrium point. We see that S_t is constant over time, and C_t and I_t are declining at the same rate. On the steady-state paths, the differential equation of C_t can be solved analytically as

$$C_t^* = C_0 e^{-\frac{\alpha\gamma}{\alpha+\beta}t} \tag{5}$$

Analogously, the path of I_t^* is also given by $I_t^* = I_0 e^{-\frac{\alpha\gamma}{\alpha+\beta}t}$.

3.4 CSIR Model as the SIR Formulation

Using the analytical results of the steady state, we can describe the CSIR model as an extended SIR model.

Proposition 1. *The dynamics of S_t, I_t, and R_t in equations (1) - (4) can be written as in the*

13

following SIR formulation

$$\frac{dS_t}{dt} = \alpha S_t I_t^{-\frac{\alpha}{\beta}} C_0 \left(I_0 e^{-\gamma t} \right)^{\frac{\alpha}{\beta}} - \beta S_t I_t \tag{6}$$

$$\frac{dI_t}{dt} = \beta S_t I_t - \gamma I_t \tag{7}$$

$$\frac{dR_t}{dt} = \gamma I_t \tag{8}$$

Proof: See Appendix B.

A critical difference from the standard SIR model is the first term of equation (6), which is complex. Because of this term, the CSIR model allows the susceptible population to increase and decrease depending on the value of I_t. It can be verified that $\frac{dS_t}{dt} = 0$ when I_t is declining at a constant rate $-\frac{\alpha \gamma}{\alpha+\beta}$ as shown in equation (5).

If we linearize the system around the steady state $(S^*, K^*) = \left(\frac{\gamma}{\alpha+\beta}, \frac{\alpha}{\beta} \right)$, we have

$$\begin{bmatrix} \dot{S}_t \\ \dot{K}_t \end{bmatrix} = \begin{bmatrix} 0 & -\frac{\beta\gamma}{\alpha+\beta} \left(\frac{\beta}{\alpha} K_0 \right)^{\frac{\alpha}{\alpha+\beta}} C_0 e^{-\frac{\alpha\gamma}{\alpha+\beta}t} \\ \frac{\alpha(\alpha+\beta)}{\beta} & 0 \end{bmatrix} \begin{bmatrix} S_t - S^* \\ K_t - K^* \end{bmatrix}.$$

Because diagonals of the Jacobian matrix are zero, eigenvalues are complex conjugate and this system is characterized by a periodic solution. However, because of the term $e^{-\frac{\alpha\gamma}{\alpha+\beta}t}$, as $t \to \infty$, the force to generate waves will slow down and approach to zero resulting in a finite number of waves.

4 Comparison between SIR and CSIR models

4.1 Implications of a More Infectious Virus

This section analyzes implications of a more infectious virus in the CSIR model and contrast them to those in the SIR model.

Figure 11 presents the time paths of the CSIR model with two different values of β. Red lines correspond to a 1.5 times higher transmission rate. Other parameter values and initial conditions are the same as in the previous section. We extend the time horizon to 6000 periods to show long-run implications.

As can be seen in the bottom-left panel, a larger β does not necessarily imply higher peaks of infection. In the first 2000 periods, red lines (higher β) exhibit lower peaks of infection. When β is higher, the speed of infection spread is fast and the susceptible
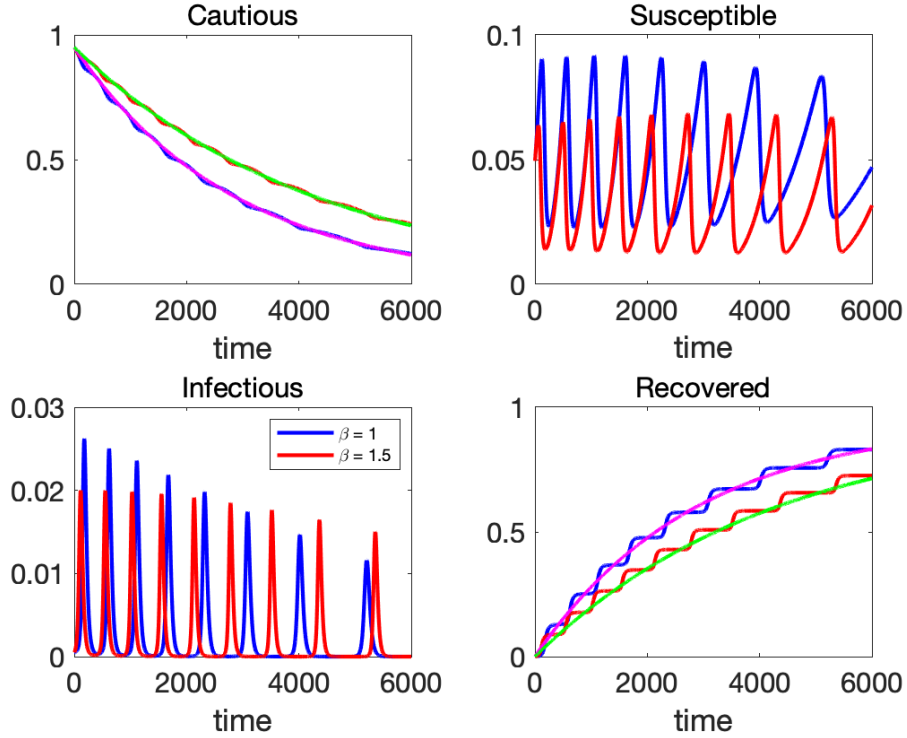
Figure 11: Time path of the CSIR model with $\beta = 1$ (blue) and $\beta = 1.5$ (red)

population is exhausted sooner compared to the case of a lower β because the inflow from C cannot keep up the rapid outflow from S. Therefore, temporary herd immunity is achieved sooner at a lower peak. However, the height of peaks will be reversed as time passes. From the top-left panel of Figure 11, we see that the higher β scenario depletes C at a slower pace. Eventually, this will strengthen the power to pump up S, so the height of infection peaks remains elevated.

This feature—a higher beta lowers the peak values of earlier waves but increases the peak of later waves—is in contrast to what we observe in the SIR model. As we saw in Figure 2, the peak of infection is higher if β is larger in the standard SIR model.

Another interesting contrast between the CSIR and SIR models lies in the epidemic final size, $R(\infty)$ given an initial condition. In the SIR model described in Section 2, we obtain $\frac{dI}{dS} = -1 + \frac{\gamma}{\beta S}$, which yields $I(t) = -S(t) + \frac{\gamma}{\beta} \ln S(t) + b$ where b is a constant. Using $I(\infty) = 0$ and assuming $S(0) \approx 1$, we obtain $S(\infty) = e^{-\frac{\beta}{\gamma}[1-S(\infty)]}$, or equivalently

$$R(\infty) = 1 - e^{-\frac{\beta}{\gamma}R(\infty)}.$$

The final size of cumulative infected population is increasing in β. This can be verified in

15

Figure 2 as well.[7]

In contrast, the opposite turns out to be true in the CSIR model.

Proposition 2. *On the steady-state paths of the CSIR model,*

$$\frac{dR^*(t)}{d\beta} < 0,$$

so more infectious virus leads to a smaller size of cumulative infected population at any given time.

Proof: See Appendix C.

Holding α and γ, an increase in β leads to smaller cumulative infection as can be seen in the bottom-right panel of Figure 11. As discussed above, a highly infectious virus (a higher β) exhausts the susceptible population too quickly compared to the rate of inflow from the cautious group, and hence, does not result in more cumulative infection. In the bottom-left panel of the same figure, the area below the infection curves determines to cumulative infection. The infection wave of a higher β exhibits steeper shapes and a shorter period (more waves within a fixed time window), so the area under the red curve is smaller than that under the blue curve.

In Figure 11, we overlay the steady-state paths of C and R as purple and green lines. They are computed from the equations shown in subsection 3.3 with the same parameter set, but initial values are replaced with the steady-state values. Given a set of parameters, the paths of C and R can be approximated by the steady-state paths, and the waves of I can be considered as a perturbation from those paths. If the initial conditions are close to the steady state values $(S^*, K^*) = \left(\frac{\gamma}{\alpha+\beta}, \frac{\alpha}{\beta}\right)$, the system generates many waves. This explains the contour plot of the number of waves in Figure 7. With the initial values $C(0) = 0.95$, $S(0) = 0.0495$, $I(0) = 0.0005$, and $\beta = 1$, when $\alpha = \frac{0.0005}{0.95} = 0.000526$ and $\gamma = S(0) \times (\alpha + \beta) = 0.0495 \times 1.000526 = 0.049526$, the initial conditions perfectly coincide the steady state values. Thus, around those parameter values of α and γ, we observe many waves.

4.2 A Large Wave versus Small Multiple Waves

Another contrast between SIR and CSIR models is the dynamics of infection over time. In Figure 12, the paths of the infectious and recovered are depicted for SIR and two cases of CSIR models ($\beta = 1$ and $\beta = 1.5$). Other parameters are set so that the final size will be

---

[7]See Miller (2012) for detailed discussions on the derivation of epidemic final sizes.
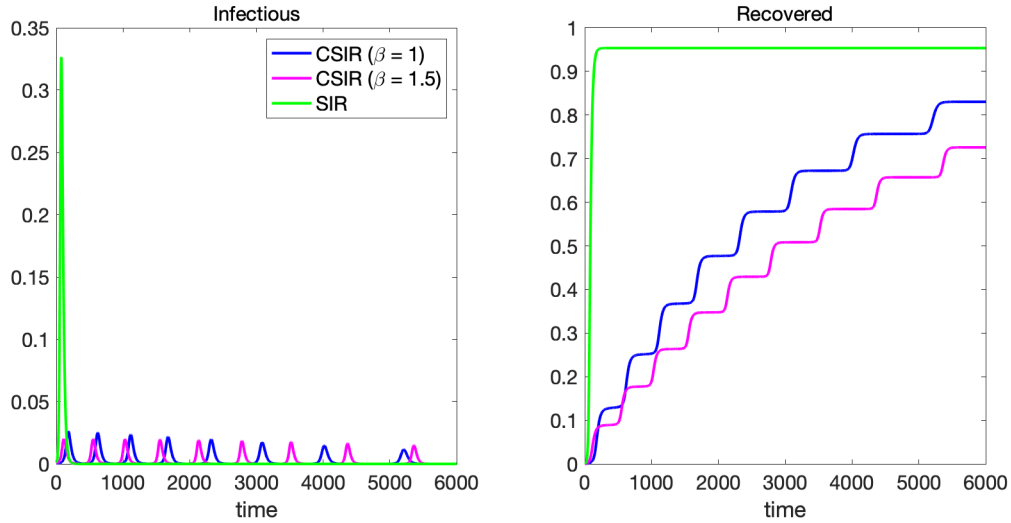
Figure 12: The paths of the infectious and recovered in CSIR and SIR models

equal for all models.[8] The SIR model generates a large wave of infection in the early phase of pandemic, and quickly reaches to the herd-immunity steady state. In this parameter specification, the number of new cases reaches more than 30% of total population at its peak, and about 95% of total population get infected in the very early phase. On the other hand, the CSIR model generates many small waves of infection over time, which is closer to reality as shown in Figure 1.

5 Other Theories of Waves

In this section, we review other theories of waves in SIR-like compartment models to be precise about our contribution.[9]

Some theories rely on people's behavioral changes to explain infection waves. As the number of infections and deaths increases, people restrain their activity or take appropriate prevention measures so as to reduce the risk of infection. Accordingly, when the number of infections or deaths increases sufficiently, infection starts declining. As infection or death counts fall, people start relaxing preventive measures, which eventually leads to a rise in infection. Models in which the transmission rate inversely depends on infection or deaths are consistent with this story, and some of these models can generate

---

[8]The final size of the CSIR model is determined by $C_0 + I_0$ as described in Appendix C. In the figure, we show the time paths in the first 6000 periods.

[9]See also Kuniya (2021), which reviews some of the papers in this section in detail. Beyond the SIR model, some have shown that network models of infection—such as scale-free network and small-world—can generate multiple waves. See, for example, Ohsawa and Tsubokura (2020) .

waves.[10] Some theories emphasize the role of government policies in promoting people's behavioral changes. When medical resources are limited, a government may need to impose NPIs to avoid collapse of medical system when infection rises and lift NPIs when infection declines and medical pressures loosen. This cycle can also create waves (Fujii and Nakata (2021), and Kurahashi (2021)).[11]

Models with the possibility of reinfection can also generate waves. One class of models with reinfection is the SIRS model in which recovered people probabilistically lose their immunity and become susceptible again. The existing studies have shown that SIRS models can generate waves under certain conditions, though they do not necessarily generate waves.[12] Another class of models with reinfection is age-structured model in which the effect of vaccines wane over time. Anderson (1995) , Franceschetti et al. (2012) , and Kuniya (2019) , among others, have studies conditions under which periodic solutions emerge.

A more straightforward way to generate waves is to exogenize the transmission rate and make it seasonal (Heathcote and Levin (1989) and Nakata and Kuniya (2010)). In a similar vein, periodic appearance of a more infectious variant—which can be captured by making the transmission rate increases over time in a step function—is likely to be able to generate waves.

Our model differs from these existing theories of waves because it features heterogeneity in risk attitudes and their time-variation. Waves during the COVID-19 pandemic are likely to reflect a mixture of these various forces.[13] Which forces were more dominant in reality likely depends on time and place. It would be useful to quantitatively investigate such a question in future research.

6   Conclusion

Many countries have experienced multiple waves of infection during the COVID-19 pandemic. In this paper, we have presented a novel extension of the SIR model, the CSIR

---

[10]See Cochrane (2020) and Heathcote and Levin (1989) show that it is possible to generate wave in a model in which the current force of infection depends on the past level of infected population. Xiao and Ruan (2007) and Capasso and Serio (1978) show that it is not possible to generate waves in SIR models in which the transmission rate continuously depends on I (t) in certain functional forms. Wang (2020) show that it is possible to generate waves if the transmission rate is a jump function of I (t).

[11]Empirical literature distinguishes voluntary and forced lockdown/social-distancing, with some arguing that the former was more important than the latter during the COVID-19 pandemic. See, for example, and Watanabe and Yabu (2021).

[12]See Good and Hawkes (2020) and Kassa et al. (2020) for the applications of SIRS models to COVID-19.

[13]If new agents are continuously born, it is possible for the SIR model to generate periodic solutions, but this does not account for what we have observed during the COVID-19 pandemic because multiple waves occurred within a year or two.

model, that can endogenously generate such multiple waves. Key features of our model are heterogeneity in risk attitudes among those who have not been infected and time-variation in these risk attitudes. We think that the mechanism in our model is a plausible contributing factor to the emergence of multiple waves during the COVID-19 pandemic and that it complements other theories of waves.

It would be interesting to examine the role of factors that are likely to have been important during the COVDI-19 pandemic, such as the role of vaccines and NPIs, in our CSIR model framework. It would be also interesting to extend the model to analyze the joint dynamics of infection and economy—as done in the economics literature using macro-SIR models—and examine the effects of various policy interventions on both infection and economy.[14] We leave these interesting extensions for future research.

---

[14]See, for example, Atkeson, Acemoglu et al. (2021) Alvarez et al. (2021), Eichenbaum et al., 2021, Far-boodi et al. (2020) , Fujii and Nakata (2021), and Kaplan et al., 2020, among many others.

References

ACEMOGLU, D., V. CHERNOZHUKOV, I. WERNING, AND M. D. WHINSTON (2021): "Optimal Targeted Lockdowns in a Multi-Group SIR Model," *American Economic Review: Insights*, forthcoming.

ALVAREZ, F., D. ARGENTE, AND F. LIPPI (2020): "A Simple Planning Problem for Covid-19 Lockdown," *Covid Economics, Vetted and Real-Time Papers*, 14, 1–32.

ANDREASEN, V. (1995): *Instability in an SIR-model with age-dependent susceptibility*, Wuerz publ., 3–14.

EICHENBAUM, M. S., S. REBELO, AND M. TRABANDT (2021): "The Macroeconomics of Epidemics," *The Review of Financial Studies*, forthcoming.

FARBOODI, M., G. JAROSCH, AND R. SHIMER (2020): "Internal and External Effects of Social Distancing in a Pandemic," *Covid Economics, Vetted and Real-Time Papers*, 9, 25–61.

FRANCESCHETTI, A., A. PUGLIESE, AND D. BREDA (2012): "Multiple endemic states in age-structured SIR epidemic models," *Mathematical Biosciences and Engineering*, 9, 577–599.

FUJII, D. AND T. NAKATA (2021): "COVID-19 and output in Japan," *The Japanese Economic Review*, 72, 609–650.

GOOD, M. F., M. T. HAWKES, AND A. SHER (2020): "The Interaction of Natural and Vaccine-Induced Immunity with Social Distancing Predicts the Evolution of the COVID-19 Pandemic," *mBio*, 11, e02617–20.

GOOLSBEE, A. AND C. SYVERSON (2021): "Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020," *Journal of Public Economics*, 193, 104311.

HAMANO, M., M. KATAYAMA, AND S. KUBOTA (2020): "COVID-19 Misperception and Macroeconomy," *Working Paper*.

HETHCOTE, H. W. AND S. A. LEVIN (1989): *Periodicity in Epidemiological Models*, Berlin, Heidelberg: Springer Berlin Heidelberg, 193–211.

KAPLAN, G., B. MOLL, AND G. L. VIOLANTE (2020): "The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S." *Working Paper*.

KASSA, S. M., J. B. NJAGARAH, AND Y. A. TEREFE (2020): "Analysis of the mitigation strategies for COVID-19: From mathematical modelling perspective," *Chaos, Solitons and Fractals*, 138, 109968.

KERMACK, W. O., A. G. MCKENDRICK, AND G. T. WALKER (1927): "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115, 700–721.

KUNIYA, T. (2019): "Hopf bifurcation in an age-structured SIR epidemic model," *Applied Mathematics Letters*, 92, 22–28.

——— (2021): "Structure of epidemic models: toward further applications in economics," *The Japanese Economic Review*, 72, 581–607.

KURAHASHI, S. (2021): "Estimating the Effectiveness of Vaccinations and Circuit Breakers: Tokyo's Forecast of Infected People," Tech. rep., Cabinet Office: COVID-19 AI and Simulation Project.

MILLER, J. C. (2012): "A note on the derivation of epidemic final sizes," *Bulletin of mathematical biology*, 74, 2125–2141.

NAKATA, Y. AND T. KUNIYA (2010): "Global dynamics of a class of SEIRS epidemic models in a periodic environment," *Journal of Mathematical Analysis and Applications*, 363, 230–237.

OHSAWA, Y. AND M. TSUBOKURA (2020): "Stay with your community: Bridges between clusters trigger expansion of COVID-19," *Plos one*, 15, e0242766.

WANG, X. (2020): "Studying social awareness of physical distancing in mitigating COVID-19 transmission," *Mathematical Biosciences and Engineering*, 17, 7428–7441.

WATANABE, T. AND T. YABU (2021): "Japan's Voluntary Lockdown," *PLos ONE*, 16, e0252468.

XIAO, D. AND S. RUAN (2007): "Global analysis of an epidemic model with nonmonotone incidence rate," *Mathematical Biosciences*, 208, 419–429.

A Alternative Specifications

The flow between the cautious and susceptible determines the dynamics of the CSIR model. In the main text, we present a particular specification, but infection waves can occur in other specifications. The laws of motion for C and S can be described by the following equations

$$\frac{dC(t)}{dt} = -FlowCS(t)$$
$$\frac{dS(t)}{dt} = FlowCS(t) - \beta S(t) I(t).$$

Consider the following cases where $FlowCS(t)$ are characterized by

$$Model\ 1: \alpha C(t) S(t)$$
$$Model\ 2: \alpha C(t) S(t) - \theta S(t) I(t)$$
$$Model\ 3: \alpha C(t) S(t) - \theta I(t)$$
$$Model\ 4: \alpha C(t)$$
$$Model\ 5: \alpha C(t) - \theta S(t)$$
$$Model\ 6: \alpha C(t) - \theta C(t) S(t)$$
$$Model\ 7: \alpha C(t) - \theta S(t) I(t).$$

Model 1 corresponds to our baseline specification presented in Section 3. Models 2 and 3 allow a reverse flow from S to C, which is increasing in the current infectious populations. These models intend to capture the "fear" effect, by which a certain fraction of the susceptible become cautious again when infection spreads in the society. Models 4 through 7 assume that a constant fraction of the cautious individuals become susceptible over time regardless of the current susceptible population (no peer effect). Models 5 through 7 allow a reverse flow from S to C in different forms.

Figures 13 and 14 show that infection waves can be generated in all models.[15] In Model 2, a temporary increase of the cautious can be observed, which implies a monotonic decline of C is not a necessary condition to generate the waves of infection. Of course, the generation of waves are not guaranteed for any parameter values, but we can find a set of parameters to generate waves in all specifications.

---

[15]Note that these figures show that there exist a set of parameter values that generates multiple waves; they do not necessarily show that waves emerge under any parameter values, as in the baseline model.
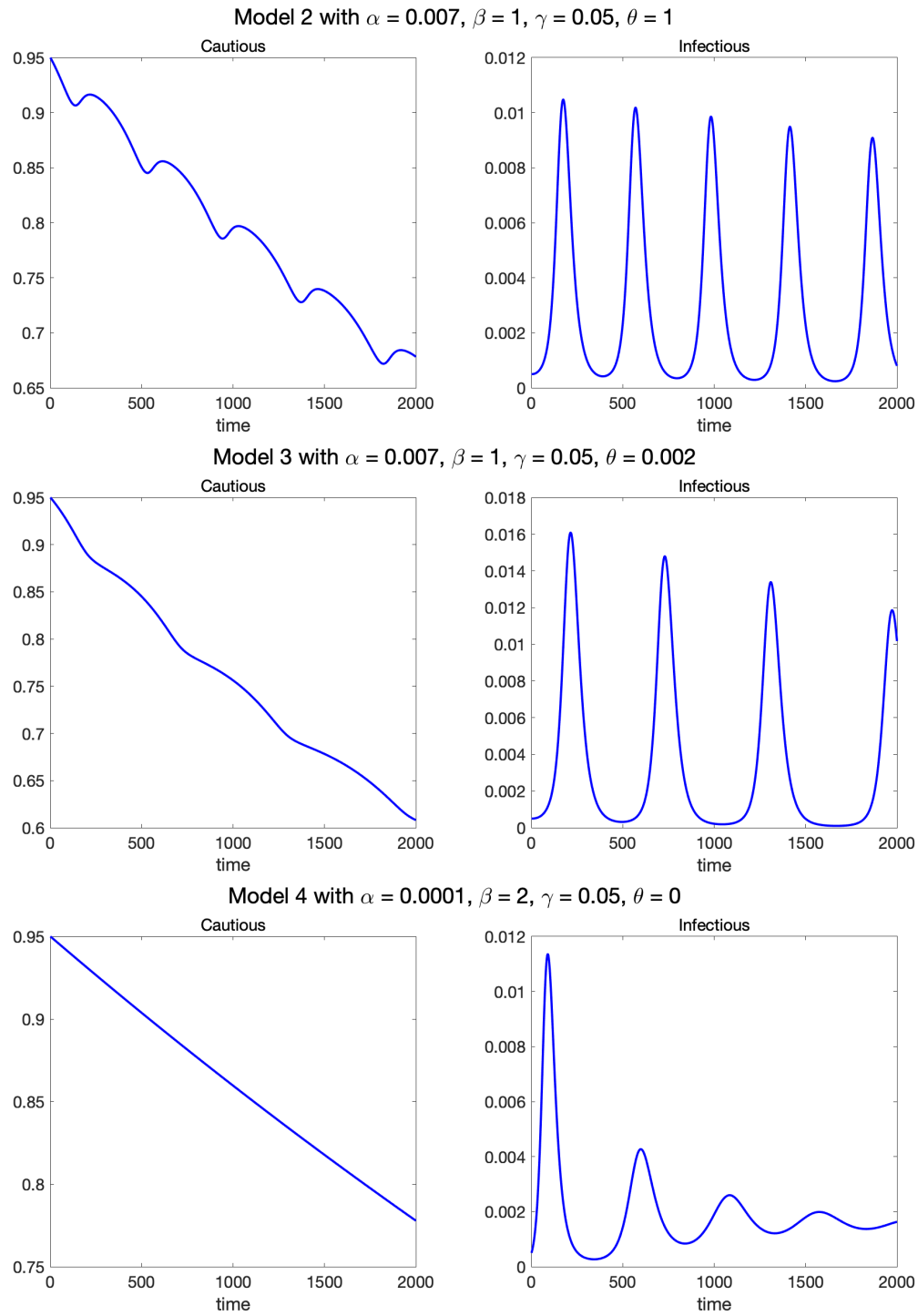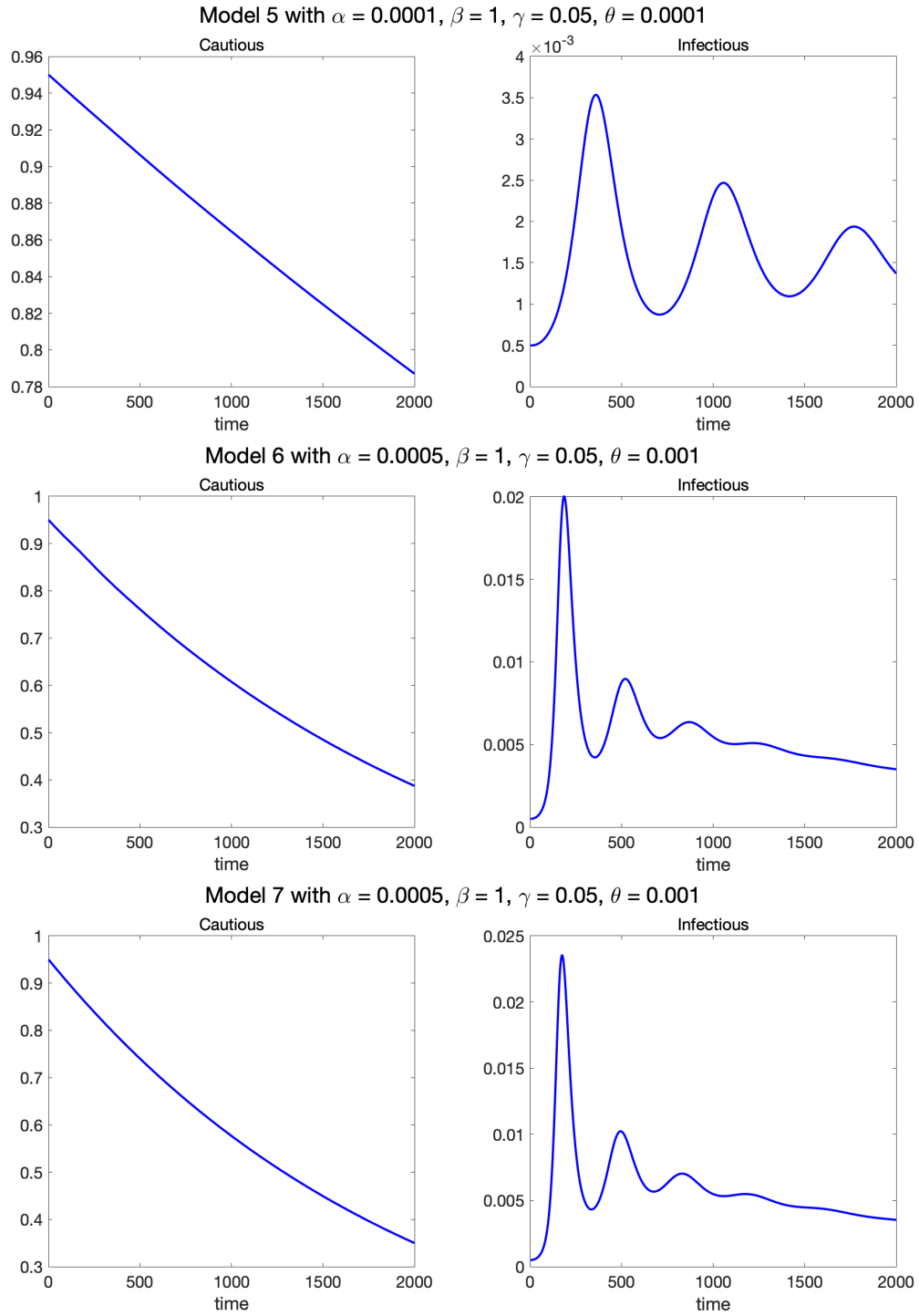
Figure 13: Infection waves in Models 2, 3, 4

Figure 14: Infection waves in Models 5, 6, 7

B Proof of Proposition 1

At the steady state paths,

$$C_t = C_t^*, S^* = \frac{\gamma}{\alpha + \beta}, K^* = \frac{\alpha}{\beta}.$$

Define a deviation of C_t from the steady state path as $\varepsilon_t = \frac{C_t}{C_t^*}$. Then,

$$d\varepsilon_t = d\left(\frac{C_t}{C_t^*}\right) = \frac{dC_t C_t^* - C_t dC_t^*}{C_t^{*2}} = \frac{(-\alpha S_t C_t) C_t^* - C_t \left(-\frac{\alpha\gamma}{\alpha+\beta} C_t^*\right)}{C_t^{*2}} dt = \left(-\alpha S_t \frac{C_t}{C_t^*} + \frac{\alpha\gamma}{\alpha+\beta} \frac{C_t}{C_t^*}\right) dt,$$

which can be written as

$$\frac{d\varepsilon_t}{dt} = \alpha\left(\frac{\gamma}{\alpha+\beta} - S_t\right)\varepsilon_t.$$

Hence, we obtain the following system

$$\frac{d\varepsilon_t}{dt} = \alpha\left(\frac{\gamma}{\alpha+\beta} - S_t\right)\varepsilon_t$$

$$\frac{dS_t}{dt} = (\alpha - \beta K_t) C_0 \varepsilon_t S_t e^{-\frac{\alpha\gamma}{\alpha+\beta}t}$$

$$\frac{dK_t}{dt} = [(\alpha+\beta)S_t - \gamma]K_t$$

We have $\frac{dK_t}{\gamma K_t dt} = \frac{(\alpha+\beta)}{\gamma}S_t - 1$ and $\frac{d\varepsilon_t}{\alpha\varepsilon_t dt} = \frac{\gamma}{\alpha+\beta} - S_t$. Then,

$$\frac{\alpha+\beta}{\alpha\gamma}\frac{d\varepsilon_t}{\varepsilon_t dt} = 1 - \frac{(\alpha+\beta)}{\gamma}S_t = -\frac{dK_t}{\gamma K_t dt}$$

$$\Rightarrow \frac{\alpha+\beta}{\alpha\gamma}\ln\varepsilon_t = -\frac{1}{\gamma}\ln K_0 + C \Rightarrow 0 = \frac{\alpha+\beta}{\alpha\gamma}\ln\varepsilon_0 = -\frac{1}{\gamma}\ln K_0 + C$$

$$\Rightarrow C = \frac{1}{\gamma}\ln K_0 = \frac{1}{\gamma}\ln\frac{I_0}{V_0} \Rightarrow \frac{\alpha+\beta}{\alpha\gamma}\ln\varepsilon_t = -\frac{1}{\gamma}\ln K_t + \frac{1}{\gamma}\ln K_0$$

$$\Rightarrow \frac{\alpha+\beta}{\alpha}\ln\varepsilon_t = \ln\frac{K_0}{K_t} \Rightarrow \varepsilon_t^{\frac{\alpha+\beta}{\alpha}} = \frac{K_0}{K_t}$$

Thus,

$$\varepsilon_t = \left(\frac{K_0}{K_t}\right)^{\frac{\alpha}{\alpha+\beta}}$$

Because $\varepsilon_t = \frac{C_t}{C_t^*}$,

$$C_t = C_t^{* \frac{\alpha+\beta}{\beta}} \left(\frac{I_0}{C_0}\right)^{\frac{\alpha}{\beta}} I_t^{-\frac{\alpha}{\beta}} = \left(C_0 e^{-\frac{\alpha\gamma}{\alpha+\beta}t}\right)^{\frac{\alpha+\beta}{\beta}} \left(\frac{I_0}{C_0}\right)^{\frac{\alpha}{\beta}} I_t^{-\frac{\alpha}{\beta}}$$

$$= C_0 \left(\frac{I_0}{I_t} e^{-\gamma t}\right)^{\frac{\alpha}{\beta}}$$

The differential equation of S_t can be written as

$$\frac{dS_t}{dt} = \alpha S_t C_t - \beta I_t S_t = \alpha S_t C_0 \left(\frac{I_0}{I_t} e^{-\gamma t}\right)^{\frac{\alpha}{\beta}} - \beta I_t S_t$$

$$= \alpha S_t I_t^{-\frac{\alpha}{\beta}} C_0 \left(I_0 e^{-\gamma t}\right)^{\frac{\alpha}{\beta}} - \beta S_t I_t$$

which is the expression in Proposition 1.

C   Proof of Proposition 2

On the steady-state path,C_t^* and I_t^* evolve as

$$C_t^* = C_0 e^{-\frac{\alpha\gamma}{\alpha+\beta}t}$$

$$I_t^* = \frac{\alpha}{\beta} C_t^* = \frac{\alpha}{\beta} C_0 e^{-\frac{\alpha\gamma}{\alpha+\beta}t}$$

Because the differential equation of R_t is given by

$$\frac{dR_t}{dt} = \gamma I_t,$$

at the steady state, it can be written as

$$\frac{dR_t^*}{dt} = \gamma \frac{\alpha}{\beta} C_0 e^{-\frac{\alpha\gamma}{\alpha+\beta}t}$$

Integrating both sides yields

$$R_t^* = -\gamma I_0 \frac{\alpha+\beta}{\alpha\gamma} \left(e^{-\frac{\alpha\beta}{\alpha+\beta}t} - 1\right) = I_0 \frac{\alpha+\beta}{\alpha} \left(1 - e^{-\frac{\alpha\gamma}{\alpha+\beta}t}\right)$$

26

Since $I_0 = \frac{\alpha}{\beta} C_0$, we can also write the above equation as

$$R_t^* = (C_0 + I_0) \left(1 - e^{-\frac{\alpha\gamma}{\alpha+\beta}t} \right)$$

Differentiating R_t^* with respect to β yields

$$\frac{dR_t^*}{d\beta} = -\frac{\alpha\gamma}{(\alpha+\beta)}t\,(C_0 + I_0)\,e^{-\frac{\alpha\gamma}{\alpha+\beta}t} < 0,$$

which concludes the proof.