A Two-Phase Dynamic Contagion Model for COVID-19

Zezhun Chen

London School of Economics

Angelos Dassios

London School of Economics

Valerie Kuan

University College London

Jia Wei Lim

Brunel University London

Yan Qu*

University of Warwick

Budhi Surya

Victoria University of Wellington

Hongbiao Zhao

Shanghai University of Finance and Economics

10th June 2020

Abstract

In this paper, we propose a continuous-time stochastic intensity model, namely, *two-phase dynamic contagion process* (2P-DCP), for modelling the epidemic contagion of COVID-19 and investigating the lockdown effect based on the dynamic contagion model introduced by Dassios and Zhao (2011). It allows randomness to the infectivity of individuals rather than a constant reproduction number as assumed by standard models. Key epidemiological quantities, such as the distribution of final epidemic size and expected epidemic duration, are derived and estimated based on real data for various regions and countries. The associated time lag of the effect of intervention in each country or region is estimated. Our results are consistent with the incubation time of COVID-19 found by recent medical study. We demonstrate that our model could potentially be a valuable tool in the modeling of COVID-19. More importantly, the proposed model of 2P-DCP could also be used as an important tool in epidemiological modelling as this type of contagion models with very simple structures is adequate to describe the evolution of regional epidemic and worldwide pandemic.

---

*Correspondence Email: y.qu3@lse.ac.uk

1 Introduction

In the early stages of epidemic modelling, the spread of diseases was formulated as a deterministic process. The classical deterministic model of *susceptible-infectious-recovered* (SIR) was introduced in the seminal paper of Kermack and McKendrick (1927). It models the spread and ultimate containment of an infection in a setting where those who recover are immune to the disease and thus the susceptible population declines over time. Many epidemic models are variations of the SIR model, see Brauer et al. (2008); Keeling and Rohani (2008); Diekmann et al. (2013) and Martcheva (2015). For example, during the outbreak of COVID-19 since December 2019, a commonly adopted approach for predicting the number of infections is the *susceptible-exposed-infectious-recovered* (SEIR) model, which adds an exposed period to the SIR model for accounting the reported incubation period of COVID-19 during which individuals are not yet infectious, e.g. Berger et al. (2020); Liu et al. (2020) and Tian et al. (2020). More recently, Acemoglu et al. (2020) develop a multi-risk SIR model, which takes into account that different subpopulations have different risks and is applied to analysing optimal lockdown.

However, the random nature of the epidemics spread in our real world suggests that a stochastic model is needed. A continuous-time stochastic counterpart of SIR model was first proposed by McKendrick (1925), and then a variety of stochastic models were studied in the literature, e.g. Bartlett (1949, 1956) and Bailey (1950, 1953, 1957). For more recent developments on stochastic epidemic models in general, see Daley and Gani (1999); Andersson and Britton (2000); Allen (2008) and Fuchs (2013). In particular, many researchers adopted branching processes. Ball (1983) used the birth-and-death process for constructing a sequence of general stochastic epidemics, and Ball and Donnelly (1995) used branching processes to approximate the early stages of epidemic dynamics, see also Britton (2010) and Ball et al. (2016).

In this paper, we propose a continuous-time stochastic epidemic model, namely, the *two-phase dynamic contagion process* (2P-DCP), for modelling the epidemic contagion. It is a branching process, and can be considered as a generalisation of dynamic contagion process (Dassios and Zhao, 2011), which is an extension of the classical Hawkes process (Hawkes, 1971a,b). In fact, Hawkes process and its various generalisations were originally used for modelling earthquakes in seismology, and recently become extremely popular for modelling financial contagion in economics, see Bowsher (2007); Large (2007); Embrechts et al. (2011); Bacry et al. (2013a,b); Aït-Sahalia et al. (2015); Dassios and Zhao (2017a,b) and Qu et al. (2019). Analogously, we advocate that they are also applicable to epidemiology. As not all individuals are equally infectious in reality, the main

advantage of this Hawkes-based approach is that, it allows randomness to the infectivity of individuals, rather than a constant reproduction number (the average number of subsequent infections of an infected individual) in standard models. In this paper, we adopt the 2P-DCP as a more realistic and parsimonious example of Hawkes-based models for modelling the current progression of COVID-19 and investigating the lockdown effect. Key epidemiological quantities, such as the distribution of final epidemic size and expected epidemic duration, have been derived and estimated based on real data. Pandemics have largely shaped the history of human being as described in the popular book by McNeill (1976), and have made huge impacts to our society and and economy. However, mathematical models developed in epidemiology and economics don't talk to each other much until the current outbreak of COVID-19, which needs urgent calls (e.g. from The Royal Society) for researchers across disciplines to work together and jointly support the scientific modelling for epidemics, see recent intensive interplays between the two fields, e.g. Acemoglu et al. (2020); Alvarez et al. (2020); Atkeson (2020); Eichenbaum et al. (2020) and Guerrieri et al. (2020). Our paper also responds to the calls by introducing the Hawkes-based approach as a potentially very valuable tool for epidemic modelling.

This paper is organised as follows: Section 2 offers the preliminaries including an introduction and formal mathematical definitions for our stochastic epidemic model, a two-phase dynamic contagion process. Key distributional properties, such as the distribution of final epidemic size and expected epidemic duration, are provided in Section 3. In Section 4, our model is implemented based on real data, and the associated time lag of the effect of intervention in each country or region is estimated. Finally, Section 5 draws a conclusion for this paper, and proposes some issues for possible further extensions and future research.

2 Two-Phase Dynamic Contagion Process

In this section, we introduce a *two-phase dynamic contagion process (2P-DCP)* for modelling the dynamics of COVID-19 contagion. The unobservable *effective time* that aggregated government interventions (e.g. lockdown of a city or country) came into effect is denoted by $\ell > 0$, which divides the COVID-19 epidemic dynamics into two phases. Note that, the time point ℓ is different from the exact timing of intervention that can be observed. The cumulated number of infected individuals is described by a counting process N_t with $N_0 = 0$, and it is modelled by a two-phase dynamic contagion process defined as below.

Definition 2.1 (Two-Phase Dynamic Contagion Process (2P-DCP))**.** *A two-phase dynamic contagion process (2P-DCP) is a point process N_t with two phases:*

Phase 1 (Full Contagion): *For the first phase period $t \in [0, \ell)$, N_t follows a dynamic contagion process with stochastic intensity*

$$\lambda_t = \lambda_0 e^{-\delta t} + \sum_{k=1}^{N_t^*} Z_k e^{-\delta(t-T_k^*)} + \sum_{i=1}^{N_t} Y_i e^{-\delta(t-T_i)}, \qquad t \in [0, \ell], \qquad (2.1)$$

where

- $\lambda_0 > 0$ *is the initial intensity at time $t = 0$;*

- $\delta > 0$ *is the constant rate of exponential decay;*

- $N_t^* \equiv \{T_k^*\}_{k=1,\dots}$ *is a Poisson process of constant rate $\varrho > 0$ arriving in time $t \leq \ell$;*

- $\{Z_k\}_{k=1,\dots,N_\ell^*}$ *are i.i.d. externally-exciting jump sizes, realised at times $\{T_k^*\}_{k=1,\dots,N_\ell^*}$, with distribution $H(y)$;*

- $\{Y_i\}_{i=1,\dots,N_\ell}$ *are i.i.d. self-exciting jump sizes of the first phase, realised at times $\{T_i\}_{i=1,\dots,N_\ell}$, with distribution $G_1(y)$.*

Phase 2 (Self Contagion): *For the second phase period $t \in (\ell, \infty)$, N_t is a pure self-exciting Hawkes process with stochastic intensity*

$$\lambda_t = \lambda_\ell e^{-\delta(t-\ell)} + \sum_{i=N_\ell+1}^{N_t} Y_i e^{-\delta(t-T_i)}, \qquad t \in (\ell, \infty), \qquad (2.2)$$

where

- λ_ℓ *is the initial intensity of the second phase starting at the cutoff time point ℓ, which is the terminal intensity of the first phase;*

- $\{Y_i\}_{i=N_\ell+1,\dots}$ *are i.i.d. self-exciting jump sizes of the second phase, realised at times $\{T_i\}_{i=N_\ell+1,\dots}$, with distribution $G_2(y)$. Note that compared with the average mean of the self-exciting jump size for Phase 1, the mean of self-exciting jump in Phase 2 is smaller, which demonstrates that the COVID-19 becomes less contagious on average after time point ℓ.*

The point process N_t and its intensity process λ_t are illustrated in Figure 1. Overall, we can more compactly define our new pandemic model, a two-phase dynamic contagion process, as a counting process $N_t \equiv \{T_i\}_{i=1,\dots}$ with $N_0 = 0$ and stochastic intensity

$$\lambda_t = \lambda_0 e^{-\delta t} + \sum_{k=1}^{N_t^*} Z_k \mathbf{1}_{\{t \leq \ell\}} e^{-\delta(t-T_k^*)} + \sum_{i=1}^{N_t} Y_i e^{-\delta(t-T_i)}, \qquad t \geq 0, \qquad (2.3)$$

where
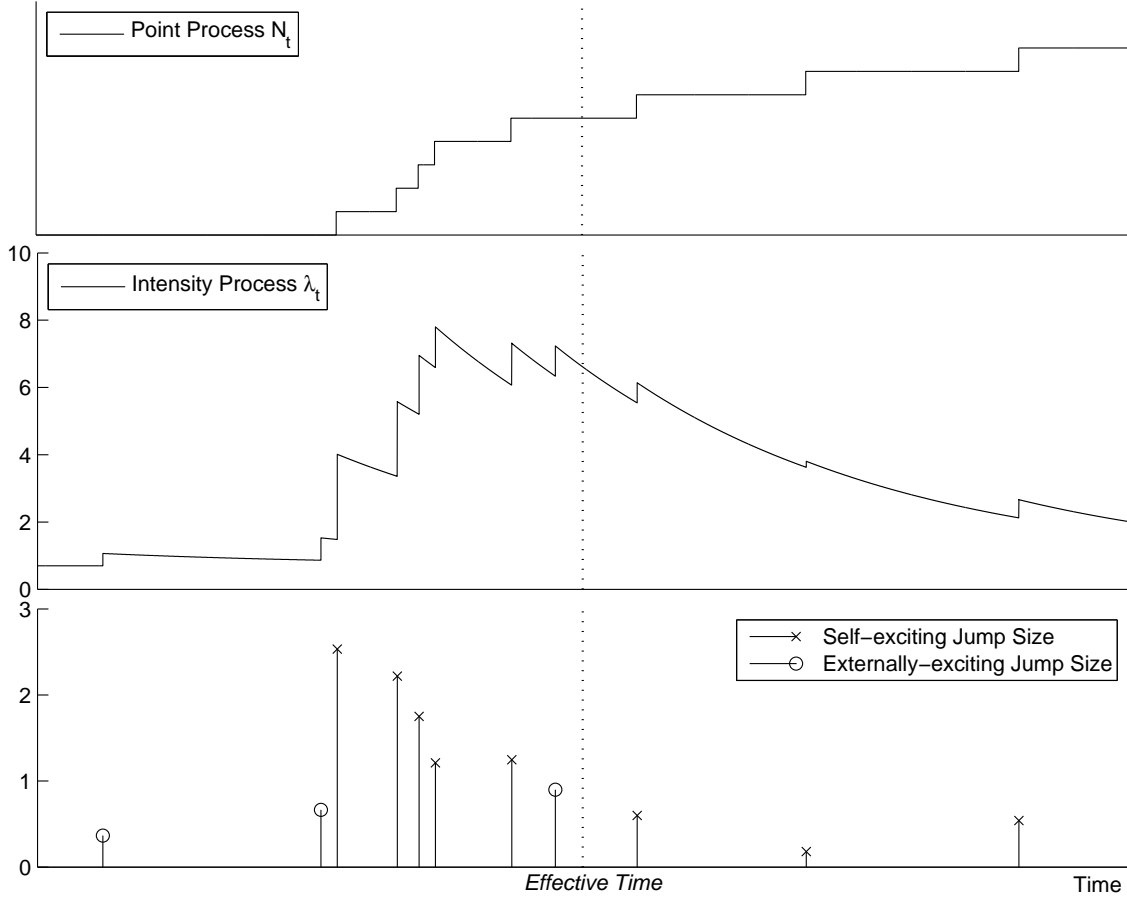
Figure 1: Two-phase dynamic contagion process

- $\{Y_i\}_{i=1,\ldots N_t}$ are i.i.d. *self-exciting* jump sizes with a two-phase distribution $G(y;t)$, i.e.,

$$G(y;t) = G_1(y)\mathbf{1}_{\{t\leq\ell\}} + G_2(y)\mathbf{1}_{\{t>\ell\}}. \tag{2.4}$$

- $\{Z_k\}_{k=1,\ldots N_t^*}$ are i.i.d. *externally-exciting* jump sizes with distribution $H(y)$.

This equivalent definition as a dynamic contagion process has an advantage: it can be viewed as a branching process and has a more intuitive interpretation with regard to a pandemic. The cluster-process presentation is provided as follows.

- The cumulated number of infected cases, N_t, is a cluster point process, which consists of two types of points: *outside-imported* cases and *inside-infected* cases.

- The arrivals of *outside-imported* cases follows a Cox process with shot-noise intensity

$$\lambda_0 e^{-\delta t} + \sum_{k=1}^{N_t^*} Z_k e^{-\delta(t-T_k^*)},$$

where externally-exciting jumps arrive as a Poisson process N_t^* at time points $\{T_k^*\}_{k=1,\ldots}$

5

with sizes (marks) $\{Z_k\}_{k=1,\ldots}$, and they disappear after time point ℓ when the interventions took effect, and there will be no any increase of imported cases in a long run.

- Each imported case may infect other individuals inside and thereby causes new cases, and each of these new cases would further infect others inside, and so on. The infection of any *new* cases caused by the *previous* infected cases follows a Cox process with exponentially decaying intensity $Y.e^{-\delta(t-T.)}$, where $Y.$ follows a two-phase distribution $G(y;t)$ and $T.$ is the infection time of the *previous* infected case. After the interventions took effect, the COVID-19 becomes less easy to spread on average. This is captured by our assumption of two-phase distribution (2.4) for $Y.$ here.

- Overall, the superposition of all these infected cases form a point process N_t, a two-phase dynamic contagion process with stochastic intensity (2.3).

3 Distributional Properties

In this section, we outline key distributional properties for the two-phase dynamic contagion process. We derived the conditional joint Laplace transform of λ_t and probability generating function of N_t, which are the key results to further derive the elimination probability of the epidemic and the distribution of the final epidemic size.

Joint Distribution of (λ_t, N_t)

Let $\{\mathcal{F}_t\}_{t\geq0}$ be the natural filtration of the point process N_t, i.e. $\mathcal{F}_t = \sigma\left(N_s, s \leq t\right)$ and assume that the intensity process $\{\lambda_t\}_{t\geq0}$ being $\mathcal{F}_t$-adapted. The joint Laplace transform and probability generating function for (λ_t, N_t) is provided in Theorem 3.1 as below.

Theorem 3.1. *For time $s \leq t$, the conditional joint Laplace transform and probability generating function for λ_t and the point process N_t is of the form*

$$\mathbb{E}\left[\theta^{N_t}e^{-v\lambda_t}|\mathcal{F}_s\right] = \begin{cases} \theta^{N_s}e^{c(s)}e^{-A(s)\lambda_s}, & 0 \leq s \leq t \leq \ell, \\ \theta^{N_s}e^{-A(s)\lambda_s}, & \ell < s \leq t, \end{cases} \tag{3.1}$$

where $A(s)$ is determined by the nonlinear ordinary differential equation (ODE)

$$A'(s) - \delta A(s) + 1 - \theta\hat{g}(A(s);s) = 0, \tag{3.2}$$

where the boundary condition is $A(t) = v$ with

$$\hat{g}(u; t) = \int_0^\infty e^{-uy} \mathrm{d}G(y; t),$$

and $c(t)$ is determined by

$$c(t) = \varrho \mathbf{1}_{\{t \le \ell\}} \int_0^t \left[1 - \hat{h}(A(u))\right] \mathrm{d}u, \tag{3.3}$$

with

$$\hat{h}(u) = \int_0^\infty e^{-uy} \mathrm{d}H(y).$$

Proof. For $0 \le s \le t \le \ell$, λ_t is the intensity process of the dynamic contagion process introduced in Dassios and Zhao (2011). The corresponding conditional joint Laplace transform, probability generating function for the process λ_t and the point process N_t is provided in Theorem 3.1 of Dassios and Zhao (2011). For $\ell < s \le t$, given $\mathcal{F}_s$, the infinitesimal generator of the dynamic contagion process (λ_s, N_s, s) acting on a function $f(\lambda, n, s)$ within its domain $\Omega(\mathcal{A})$ is given by

$$\mathcal{A}f(\lambda, n, s) = \frac{\partial f}{\partial s} - \delta\lambda\frac{\partial f}{\partial \lambda} + \lambda \left(\int_0^\infty f(\lambda + y, n + 1, s)\mathrm{d}G(y; s) - f(\lambda, n, s) \right). \tag{3.4}$$

Consider a function $f(\lambda, n, s)$ of form

$$f(\lambda, n, s) = \theta^n e^{-A(s)\lambda},$$

and substitute this into $\mathcal{A}f = 0$, we then have the ODE

$$A'(s) = \delta A(s) - 1 + \theta \hat{g}\big(A(s), s\big),$$

adding the boundary condition $A(t) = v$, gives the ODE in (3.2). □

The moments of λ_t and N_t can be obtained by differentiating the joint Laplace transform and probability generating function of λ_t and N_t, and the results are provided in Proposition 3.1 and 3.2 as below.

Proposition 3.1. *The conditional expectation of the process λ_t given $\mathcal{F}_s$ for $s \le t$ is given by*

$$\mathbb{E}\left[\lambda_t | \mathcal{F}_s\right] = \begin{cases} \frac{\varrho\mu_H\mathbf{1}_{\{t\le\ell\}}}{\kappa} + \left(\lambda_s - \frac{\varrho\mu_H\mathbf{1}_{\{t\le\ell\}}}{\kappa}\right)e^{-\kappa(t-s)}, & \kappa \ne 0, \\ \lambda_s + \varrho\mu_H\mathbf{1}_{\{t\le\ell\}}(t-s), & \kappa = 0, \end{cases} \tag{3.5}$$

and the conditional expectation of the point process N_t given $\mathcal{F}_s$ is of the form

$$\mathbb{E}\left[N_t|\mathcal{F}_s\right] = \begin{cases} N_s + \frac{\varrho\mu_H\mathbf{1}_{\{t\leq\ell\}}(t-s)}{\kappa} + \left(\lambda_s - \frac{\varrho\mu_H\mathbf{1}_{\{t\leq\ell\}}}{\delta-\mu_G}\right)\frac{1-e^{-\kappa(t-s)}}{\kappa}, & \kappa \neq 0, \\[2mm] N_s + \lambda_s(t-s) + \frac{1}{2}\varrho\mu_H\mathbf{1}_{\{t\leq\ell\}}(t-s)^2, & \kappa = 0, \end{cases} \quad (3.6)$$

where

$$\mu_H = \int_0^\infty y\,\mathrm{d}H(y), \quad \mu_G = \int_0^\infty y\,\mathrm{d}G_1(y)\mathbf{1}_{\{t\leq\ell\}} + \int_0^\infty y\,\mathrm{d}G_2(y)\mathbf{1}_{\{t>\ell\}},$$

and $\kappa = \delta - \mu_G$.

Proof. The result in (3.5) immediately follows Theorem 3.6 in Dassios and Zhao (2011). And since $N_t - N_s - \int_s^t \lambda_u\,\mathrm{d}u$ is a martingale, we have

$$\mathbb{E}\left[N_t - N_s|\mathcal{F}_s\right] = \int_s^t \mathbb{E}\left[\lambda_u|\mathcal{F}_s\right]\mathrm{d}u,$$

which directly implies the result in (3.6). $\qquad\square$

Proposition 3.2. *The conditional second moment of the process λ_t given $\mathcal{F}_s$ for $s \leq t$ is given by*

$$\mathbb{E}\left[\lambda_t^2|\mathcal{F}_s\right] = \begin{cases} \lambda_s^2 e^{-2\kappa t} + \frac{2\varrho\mu_H\mathbf{1}_{\{t\leq\ell\}}+\mu_{2G}}{\kappa}\left(\lambda_s - \frac{\varrho\mu_H}{\kappa}\right)\left(e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}\right) \\[2mm] \quad + \left(\frac{(2\varrho\mu_H+\mu_{2G})\varrho\mu_H\mathbf{1}_{\{t\leq\ell\}}}{2\kappa^2} + \frac{\varrho\mu_{2H}\mathbf{1}_{\{t\leq\ell\}}}{2\kappa}\right)\left(1 - e^{-\kappa(t-s)}\right), & \kappa \neq 0, \\[2mm] \lambda_s^2 + \lambda_s\mu_{2G}t + (2\lambda_0\mu_H + \mu_{2H})\varrho\mathbf{1}_{\{t\leq\ell\}}(t-s) \\[2mm] \quad + \left(\varrho^2\mu_H^2 + \frac{1}{2}\varrho\mu_H\mu_{2G}\right)\mathbf{1}_{\{t\leq\ell\}}(t-s)^2, & \kappa = 0, \end{cases} \quad (3.7)$$

and the conditional joint expectation of the process λ_t and the point process N_t given $\mathcal{F}_s$ for $s > \ell$ is of the form

$$\mathbb{E}\left[\lambda_t N_t|\mathcal{F}_s\right] = \begin{cases} \lambda_s N_s e^{-\kappa(t-s)} + \left(\lambda_s\mu_G + \frac{\lambda_s\mu_{2G}}{\kappa}\right)(t-s)e^{-\kappa(t-s)} \\[2mm] \quad + \left(\frac{\lambda_s^2}{\kappa} - \frac{\lambda_s\mu_{2G}}{\kappa}\right)(e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}), & \kappa \neq 0, \\[2mm] \lambda_s N_s + \left(\lambda_s^2 + \lambda_s\mu_G\right)(t-s) + \frac{1}{2}\lambda_s\mu_{2G}(t-s)^2, & \kappa = 0 \end{cases} \quad (3.8)$$

and the second moment of point process N_t given $\mathcal{F}_s$ for $s > \ell$ is of the form

$$\mathbb{E}\left[N_t^2|\mathcal{F}_s\right] = \begin{cases} \left(\frac{\lambda_s\mu_{2G}}{\kappa^3} - \frac{\lambda_s^2}{\kappa^2}\right)(1 - e^{-2\kappa(t-s)}) - \left(\frac{2\lambda_s\mu_G}{\kappa} + \frac{2\lambda_s\mu_{G2}}{\kappa^2}\right)(t-s)e^{-\kappa(t-s)} \\[2mm] \quad + \left(\frac{\lambda_s+2\lambda_s N_s}{\kappa} + \frac{2\lambda_s^2+2\lambda_s\mu_{2G}}{\kappa^2}\right)(1 - e^{-\kappa(t-s)}), & \kappa \neq 0, (3.9) \\[2mm] \lambda_s(\mu_G + 2N_s)(t-s) + (\lambda_s^2 + \lambda_s\mu_G)(t-s)^2 + \frac{1}{3}\lambda_s\mu_{2G}(t-s)^3, & \kappa = 0, \end{cases}$$

8

where

$$\mu_{2_H} = \int_0^\infty y^2 \mathrm{d}H(y), \quad \mu_{2_G} = \int_0^\infty y^2 \mathrm{d}G_1(y)\mathbf{1}_{\{t \le \ell\}} + \int_0^\infty y^2 \mathrm{d}G_2(y)\mathbf{1}_{\{t > \ell\}},$$

and $\kappa = \delta - \mu_G$.

Proof. These results immediately follows Lemma 3.1, Lemma 3.2 and Theorem 3.9 in Dassios and Zhao (2011). □

Probability of Elimination Time $\widetilde{T}$

After the government interventions come into effect, the contagion rate will dramatically decline and new cases will drop abruptly almost to nothing in the near future. It is therefore of great interests to calculate the probability of elimination time, i.e. the time that the last ever case arrives, after government interventions come into effect. Let $\widetilde{T}$ to be *elimination time* such that

$$\widetilde{T} := \inf \left\{ t > \ell : \forall s \ge t, \quad N_s - N_t = 0 \right\}. \tag{3.10}$$

The condition probability of the elimination time is provided in Proposition 3.3.

Proposition 3.3. *For $\ell \le s \le t$, the elimination probability is given by*

$$\mathbb{P}\left(\widetilde{T} \le t | \mathcal{F}_s\right) = e^{-A(s)\lambda_s}, \tag{3.11}$$

where $A(s)$ is determined by the ODE in (3.2) with boundary condition $A(t) = \frac{1}{\delta}$.

Proof. Given $\widetilde{T}$ being the timing of the last ever event, the event $\{\widetilde{T} \le t\}$ implies that $N_u = N_t$ for any $u \ge t$, which also lead to

$$\lambda_u = e^{-\delta(u-t)}\lambda_t.$$

Hence, we have

$$
\begin{aligned}
\mathbb{P}\left(\widetilde{T} \le t | \mathcal{F}_s\right) &= \mathbb{E}\left[\mathbf{1}_{\{\widetilde{T} \le t\}} \mid \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\exp\left(-\int_t^\infty \lambda_t e^{-\delta(u-t)}\mathrm{d}u\right) \mid \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\exp\left(-\frac{\lambda_t}{\delta}\right) \mid \mathcal{F}_s\right]
\end{aligned}
\tag{3.12}
$$

And according to Theorem 3.1, by setting $\theta = 1$ in (3.1), the result follows immediately. □

Joint Expectation of Epidemic Size N_t and Elimination Time $\widetilde{T}$

Given the last ever event $\{\widetilde{T} < t\}$, one could obtain the expected size of the epidemic at time t. The relevant details are presented in Corollary 3.1.

Corollary 3.1. *For $\ell \leq s \leq \widetilde{T} \leq t$, the conditional joint expectation of N_t and $\mathbf{1}_{\{\widetilde{T}\leq t\}}$ is of the following form*

$$\mathbb{E}\left[N_t \mathbf{1}_{\{\widetilde{T}\leq t\}} \mid \mathcal{F}_s\right] = \frac{\mathrm{d}}{\mathrm{d}\theta}\left\{\theta^{N_s}e^{-A(s)\lambda_s}\right\}\Bigg|_{\theta=1^-}, \tag{3.13}$$

where $A(s)$ satisfies the ODE in (3.2) with boundary condition $A(t) = \frac{1}{\delta}$.

Proof. According to Theorem 3.1 and Proposition 3.3, we have

$$\begin{aligned}\mathbb{E}\left[\theta^{N_t}\mathbf{1}_{\{\widetilde{T}\leq t\}} \mid \mathcal{F}_s\right] &= \mathbb{E}\left[\theta^{N_t}e^{-\frac{\lambda_t}{\delta}} \mid \mathcal{F}_s\right] \\ &= \theta^{N_s}e^{-A(s)\lambda_s}.\end{aligned} \tag{3.14}$$

Since $\mathbb{E}\left[N_t\mathbf{1}_{\{\widetilde{T}\leq t\}} \mid \mathcal{F}_s\right] = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}\left[\theta^{N_t}\mathbf{1}_{\{\widetilde{T}\leq t\}} \mid \mathcal{F}_s\right]\Big|_{\theta=1^-}$, the result immediately follows (3.13). $\qquad\square$

Distribution of Final Epidemic Size N_∞

The final epidemic size is one of the most important epidemiological quantities to study. In fact, under the two-phase dynamic contagion model, the final epidemic size is the value of the point process N_t when time goes to infinity. Conditional on $s > \ell$, since there are no externally-exciting jumps in the intensity, the distribution of N_∞ can be characterised by Proposition 3.4 as below.

Proposition 3.4. *For $\ell < s$, the probability generating function of N_∞ conditional on $\mathcal{F}_s$ is given as*

$$\mathbb{E}\left[\theta^{N_\infty} \mid \mathcal{F}_s\right] = e^{-v^*\lambda_s}, \tag{3.15}$$

where

$$v^* = \frac{1}{\delta}\left(1 - \theta\int_0^\infty e^{-v^*y}\mathrm{d}G_2(y)\right).$$

Proof. The result immediately follows Theorem 3.5 in Dassios and Zhao (2011). $\qquad\square$

While the government interventions come into effect, if we assume *i.i.d.* self-exciting jump sizes $Y_i \sim \text{Exp}(\beta)$ for $i = N_\ell + 1, ...$, then, we have an explicit expression for the probability

generating function of N_∞ as

$$\mathbb{E}[\theta^{N_\infty} \mid \mathcal{F}_s] = \exp\left(-\frac{\sqrt{(\delta\beta-1)^2 + 4\delta\beta(1-\theta)} - (\delta\beta-1)}{2\delta}\lambda_s\right), \quad s > \ell.$$

This implies that, the final epidemic size N_∞ conditional on $\mathcal{F}_\ell$ follows a mixed-Poisson distribution with the probability mass function

$$\mathbb{P}\left(N_\infty = k \mid \mathcal{F}_s\right) = \int\limits_0^\infty \frac{v^k e^{-v}}{k!} m(v)\mathrm{d}v, \qquad k = 0, 1, ...,$$ (3.16)

where $m(v)$ is the density function of the mixing distribution,

$$m(v) := \exp\left(\frac{\delta\beta-1}{2\delta}\lambda_s - \left(\frac{\delta\beta-1}{2\delta}\right)^2 \frac{\delta}{\beta}v - \frac{\frac{\beta}{2\delta}\lambda_s^2}{2v}\right)\frac{\sqrt{\frac{\beta}{2\delta}}\lambda_s}{\sqrt{2\pi}v^{\frac{3}{2}}},$$

which is an inverse Gaussian distribution with parameters $\frac{\beta}{\delta\beta-1}\lambda_s$ and $\frac{\beta}{2\delta}\lambda_s^2$.

4 Empirical Study

We provide a calibration scheme based on the daily increments of the two-phase dynamic contagion process N_t. Let us first denote the observations of the daily confirmed COVID-19 cases as $\{C_t\}_{t=0,1,2,...,T}$. The mean square error (MSE) between the expected daily increments of N_t and the actual reported daily confirmed COVID-19 cases is given as

$$\mathrm{MSE}(\alpha, \beta, \delta, \varrho, \ell) = \frac{1}{T}\sum_{t=0}^{T-1}\left(\mathbb{E}\left[N_{t+1} - N_t\right] - C_{t+1}\right)^2.$$ (4.1)

We consider the calibration based on minimising the MSE (4.1), i.e., we choose parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ such that

$$\mathrm{MSE}(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell}) := \min_{(\alpha,\beta,\delta,\varrho,\ell)} \mathrm{MSE}(\alpha, \beta, \delta, \varrho, \ell),$$

with $\alpha, \beta, \delta, \varrho \geq 0$ and $\ell \in \mathbb{N}^+$.

Without loss of generality, for simplicity, we assume $Z_k = 1$ for any k, $Y_i \sim \mathrm{Exp}(\alpha)$ for $i = 1, ..., N_\ell$ and $Y_i \sim \mathrm{Exp}(\beta)$ for $i = N_\ell + 1, ...,$ and $\lambda_0 = 0$ in (2.3) for model calibration. Other assumptions for $Z_k, \{Y_i\}_{i=1,...N_\ell, N_\ell+1,...}, \lambda_0$ can also be used if necessary. We provide two empirical examples. The first one concentrates on the COVID-19 pandemic in mainland China during the period from early January to late March 2020. The second one focuses on the worldwide COVID-19 pandemic during the period from mid February to early May 2020. The data we used

are publicly available. Datasets are mostly cited from the associated official government health department websites for non-European countries and from the European Centre of Disease Control (ECDC) for European countries.

COVID-19 Pandemic in Mainland China

The daily confirmed COVID-19 cases for regions in China can be obtained from daily reports of the National Health Commission of the PRC. We use the reported daily confirmed cases for several regions in China from 2020-01-19 to 2020-03-31 as examples for model calibration. The corresponding estimation results are illustrated in Table 1. We can see that the estimator $\hat{\varrho}$ are quite different from each other. In particular, these regions that are close to Hubei, namely Henan, Hunan, Anhui, have relatively larger intensities for externally-exciting jumps, which means that these regions experienced more external shocks from Wuhan and these external shocks can be originally infected individuals from Wuhan. Naturally, without taking into account Wuhan, the rest of Hubei has the largest intensity $\hat{\varrho}$ for externally-exciting jumps. The main government intervention established by the Chinese authority is the announcement of the completely lockdown of Wuhan and later the whole Hubei Province on 23 January 2020. One or two days later, all other regions enforced the quarantine and raised the alert of public health emergency. Setting the date 2020-01-19 as the initial time $t = 0$, then the government intervention took place when $t = 4$ and came into effect when $t = \hat{\ell}$. Since the delay period of the government interventions is the difference of the date when the government interventions came into effect and the date when the government introduced the restriction measures, we can therefore observe from Table 1 that the estimated delays of the government interventions for different regions therefore are between 5 and 15 days, which are consistent to the *incubation time* of COVID-19 for most people, e.g. as found in a highly cited medical study of Lauer et al. (2020).

The *branching ratio* (BR), which demonstrates the average infection rate, is determined by $\mathbb{E}[Y_i]/\delta$. In Table 2, we compare the estimated branching ratios before and after the government interventions came into effect, namely R_b and R_a, respectively. It is clear that the branching ratio for every region decreases significantly when the state changed, i.e. government interventions came into effect. One can also access the efficiency for when regions implemented the restriction packages introduced by the central government by comparing the corresponding branching ratios R_b and R_a. The comparison of R_b and R_a for regions in China are presented in Figure 2. We can see that the government restriction packages had been well-implemented for all regions in China. In particular, Hubei, where the strictest measure, i.e. the completely lockdown of Wuhan and Hubei,

Table 1: Calibration parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ for total confirmed COVID-19 cases from 2020-01-19 to 2020-03-31 for various regions in China.

Parameters / Regions	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\varrho}$	$\hat{\ell}$	$\frac{\text{MSE}(\hat{\alpha},\hat{\beta},\hat{\delta},\hat{\varrho},\hat{\ell})}{N}$
Heilongjiang	2.431123	8.036146	0.252324	0.934574	16	0.036414
Sicuan	5.509284	7.402851	0.21608	3.913496	11	0.023037
Shandong	5.303477	7.868065	0.340273	5.296465	19	0.033823
Jiangxi	2.936512	7.903677	0.285922	3.782104	15	0.070023
Anhui	3.644156	7.773938	0.339183	7.161244	18	0.037944
Hunan	4.467952	7.660323	0.293860	8.351008	15	0.053281
Zhejiang	1.398741	8.065125	0.260449	1.627822	09	0.144322
Henan	3.370932	7.867786	0.286357	6.669105	15	0.042004
Guangdong	2.316879	7.987904	0.256143	2.360507	12	0.073389
Hubei	4.468435	28.46780	0.173874	58.24001	16	0.654813

had been introduced, shows a dramatic drop of contagion rate after the interventions came into effect.

Table 2: The estimated branching ratio (BR) before and after the government interventions came into effect, namely R_b, R_a respectively, for regions in China.

| BR / Regions	R_b	R_a
Heilongjiang	1.630176	0.493166
Sicuan	0.840021	0.625153
Shandong	0.554130	0.373512
Jiangxi	1.191025	0.442510
Anhui	0.809039	0.379250
Hunan	0.761641	0.444234
Zhejiang	2.744989	0.476066
Henan	1.035957	0.443853
Guangdong	1.685052	0.488747
Hubei	1.287094	0.202028
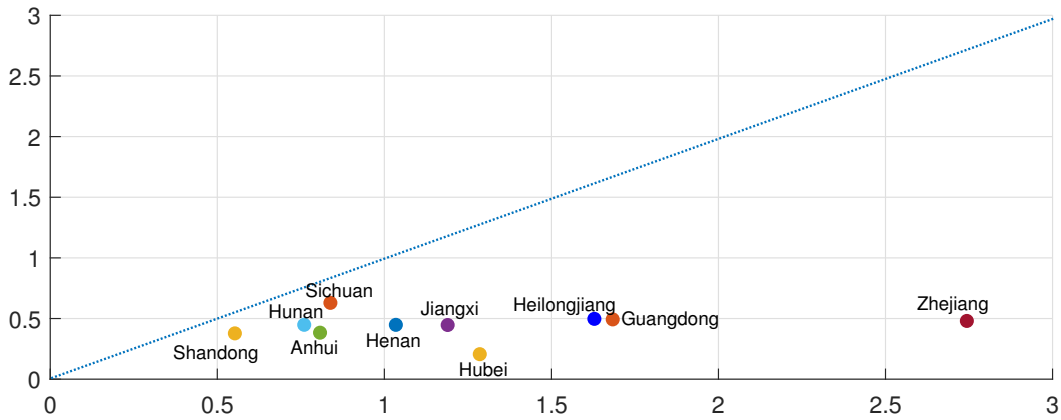
Figure 2: Comparison of the branching ratios *before* and *after* the government interventions came into effect for regions in China. The horizontal axis represents the branching ratio *before* the government interventions came into effect, namely R_b and the vertical axis represents the branching ratio *after* the government interventions came into effect, i.e. R_a.

Figure 3 and 4 demonstrate comparisons between the expected daily/total confirmed cases for the two-phase dynamic contagion model under the calibrated parameter $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ in Table 1 and the actual daily/total confirmed COVID-19 cases for the period 2020-01-19 to 2020-03-31. We observe that the model allows different shapes of trend before the interventions came into effect. All these regions indicate relatively smooth exponential decay of daily new cases after the peak. In addition, the estimated cumulative confirmed cases are very close to the actual total confirmed COVID-19 cases, which further confirms that our new model is a good candidate for describing the propagation process.
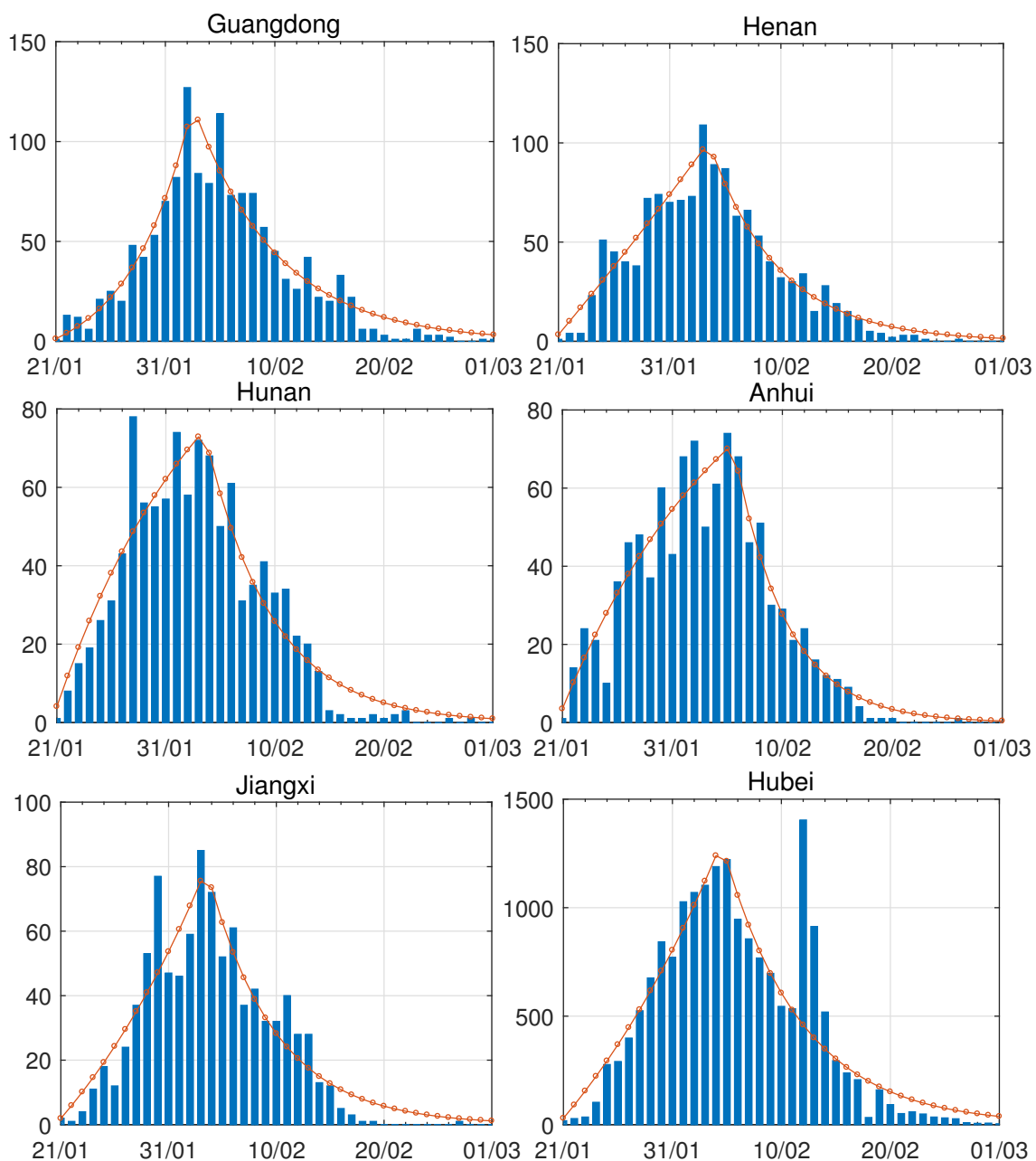
Figure 3: Model calibration comparisons between the expected daily confirmed cases under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ in Table 1 and actual daily confirmed COVID-19 cases from 2020-01-19 to 2020-03-31 for Guangdong, Henan, Hunan, Anhui, Jiangxi, and Hubei of China.
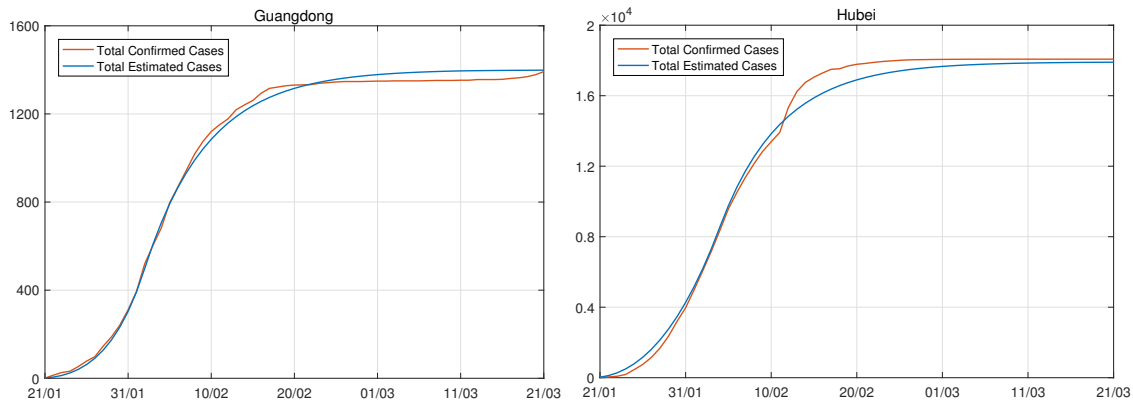
Figure 4: Comparisons between total confirmed COVID-19 cases and total estimated cases under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ in Table 1 from 2020-01-19 to 2020-03-31 for Guangdong and Hubei, China.

COVID-19 Pandemic for the World

From mid-February 2020, the COVID-19 started to spread in other countries around the world. At beginning, only a small number of initial cases were reported for some countries in Europe, South/East Asia and North American. However, lately, several large outbreaks were reported in South Korea, Italy, Iran, Spain, Japan and the total number of cases outside China quickly passed the China's. The WHO then recognized the spread of COVID-19 as pandemic on 2020-03-11. We could use this as a second example to confirm our observations from the last exercise. The calibration settings were the same as the previous one. We use the reported daily confirm cases for different regions and countries around world from mid-February to early May 2020. Note that, due to the fact that the pandemic reached each country or territory at different time and the corresponding government interventions also imposed and came into effect at different times, there is no sense to calibrate the model using the data within the same truncated time series. Table 3 presents the estimation results $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ of $(\alpha, \beta, \delta, \varrho, \ell)$ for various countries and territories. We notice that regions and countries like Italy, China, New York have much larger ϱ compared with other areas. This phenomena is reasonable as these areas have specific outbreak area which created external shocks to other part of the regions and countries and hence the number of confirmed cases increased more rapidly than other regions and countries.

In Table 3, we also presents the date of day 0, i.e. $Date_0$, and the date of government interventions imposed, i.e. $Date_G$. The delay period of government interventions came into effect therefore can be obtained given the estimated $\hat{\ell}$, with $Date_0$ and $Date_G$. The details for the delay periods of regions and countries are illustrated in Figure 5. We can see that the delay of the interventions for

15

Table 3: Calibration parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ for total confirmed COVID-19 cases from mid-February, 2020 to early May, 2020.

| Regions	$Date_0$	$Date_G$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\varrho}$	$\hat{\ell}$	$\frac{MSE(\hat{\alpha},\hat{\beta},\hat{\delta},\hat{\varrho},\hat{\ell})}{N}$
Australia	2020-02-27	2020-03-15	1.703807	3.316083	0.401550	0.489762	28	0.124391
Austria	2020-03-01	2020-03-10	2.860291	7.781922	0.222730	5.788434	24	0.326694
China(Mainland)	2020-01-19	2020-01-23	2.846236	6.534269	0.242773	83.160865	17	1.003152
Czech	2020-03-04	2020-03-10	3.464654	7.842862	0.174990	2.247588	25	0.359398
France	2020-02-25	2020-03-13	3.453459	6.912410	0.191890	15.492535	36	5.410160
Germany	2020-02-24	2020-03-12	3.246798	6.402097	0.201327	22.863727	32	3.464682
Greece	2020-02-26	2020-03-10	4.181831	7.778144	0.200886	1.337306	35	0.172176
Hong Kong	2020-03-01	2020-03-23	2.782369	8.054402	0.287048	0.512120	31	0.041714
Iceland	2020-02-28	2020-03-13	3.405901	7.908691	0.275378	1.892734	33	0.163572
Italy	2020-02-21	2020-03-05	3.068103	5.445646	0.211048	23.830431	30	0.911767
Latvia	2020-03-07	2020-03-13	4.227388	7.819275	0.208209	1.145524	22	0.099024
New York	2020-02-29	2020-03-12	2.449906	3.203563	0.356592	70.937604	33	2.926862
New Zealand	2020-03-12	2020-03-16	2.122999	8.438649	0.204667	0.551023	14	0.064829
Norway	2020-02-21	2020-03-12	4.496501	7.519573	0.195859	6.553447	29	0.303364
South Korea	2020-02-16	2020-02-20	1.309507	8.771968	0.244316	5.641225	13	0.418018
Switzerland	2020-02-25	2020-03-13	3.195829	7.046864	0.202615	14.083749	28	1.571950

different regions and countries is around $8 \sim 21$ days. In fact, the delay period can be considered as a criteria for evaluating the effectiveness of restrictions imposed by the authorities to prevent further spread of COVID-19. In general, most regions and countries with short delay periods normally took tougher restrictions or more effective measures to stop the spread of virus. New Zealand and South Korea are two typical examples. The authority of New Zealand introduced a nationwide lockdown by closing all borders and entry ports to all nonresidents. On contrary, the South Korea authority introduced one of the largest and best-organised epidemic control programs to screen the mass population for the virus with isolation, tracing, quarantine took place simultaneously without further lockdown. Therefore, the estimated delay periods for these two countries are only 9 days for New Zealand and 8 days for South Korea, which are much shorter than the average incubation time of COVID-19. For most European countries, due to the containment restriction measures such as quarantines and curfews were not strictly put into effect, the associated delay periods are relatively longer than the incubation time of COVID-19.

The estimated branching ratios before and after the government interventions came into effect for different countries and territories are reported in Table 4. And Figure 6 demonstrates a comparison between the BRs before the government interventions took effect and the BRs after the interventions worked, with a blue dash line that represents $R_b = R_a$. We can immediately see that for most regions and countries, the BR dropped dramatically after government interventions came into effect, which suggests that the restriction measures imposed by the authority indeed reduce the contagion/infection rate significantly.
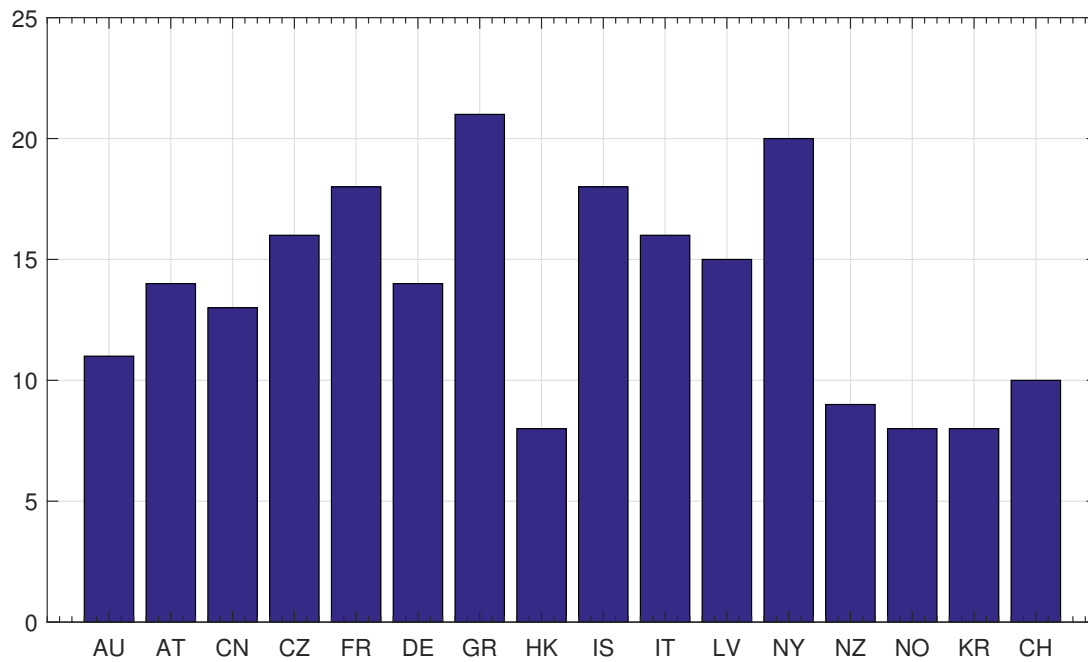
Figure 5: Comparison of the delay period for different regions and countries around the world. The horizontal axis represents the abbreviation of regions and countries listed in Table 4 and the vertical axis represents the number of days for the government interventions came into effect after the government announcement the relevant measures.
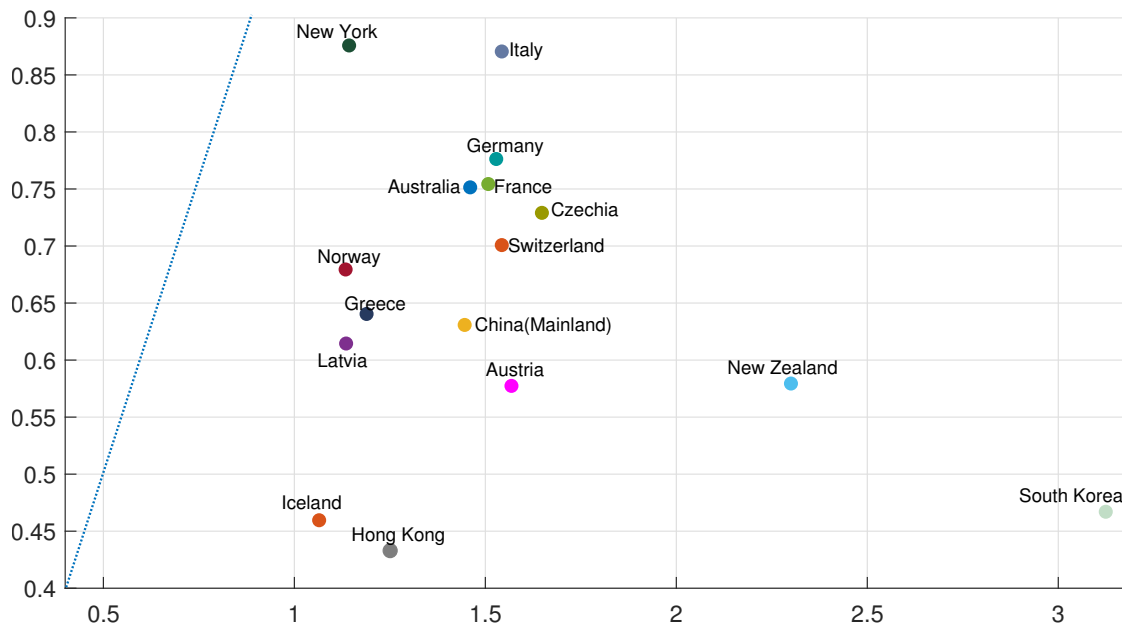
Figure 6: Comparison of the branching ratios *before* and *after* the government interventions came into effect for different regions and countries around the world. The horizontal axis represents the branching ratio *before* the government interventions came into effect, R_b, and the vertical axis represents the branching ratio *after* the government interventions came into effect, R_a.

Table 4: The estimated branching ratio (BR) *before* and *after* the government interventions came into effect, namely R_b, R_a respectively, for regions and countries around the would.

BR Regions	R_b	R_a
Australia	1.461638	0.750991
Austria	1.569681	0.576945
China(Mainland)	1.447200	0.630380
Czechia	1.649399	0.728637
France	1.509016	0.753908
Germany	1.529826	0.775845
Greece	1.190377	0.639993
Hong Kong	1.252077	0.432526
Iceland	1.066200	0.459162
Italy	1.544364	0.870102
Latvia	1.136798	0.614084
New York	1.144667	0.875377
New Zealand	2.301455	0.579001
Norway	1.135488	0.678991
South Korea	3.125643	0.466606
Switzerland	1.544349	0.700379

A comparison between the expected daily confirmed cases for the two-phase dynamic contagion model under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ is reported in Table 3, and the actual daily confirmed COVID-19 cases for different regions and countries over the period of mid-February to early May are presented in Figure 7. In general, we can see that the model can precisely catch the trend of infection, this further confirms that the two-phase dynamic contagion model is effective. Note that, we have smoothed the biggest jump of daily confirmed cases in China for better illustration and fitting purpose. This is due to a change in the confirmation standard established by the Chinese authority on 2020-02-12.

In Figure 8, we compare the estimated daily increment with the actual confirmed COVID-19 cases for France, Germany, Switzerland, and New York. The daily records of confirmed cases for these areas were not as smooth as those countries illustrated in Figure 7. The spikes within the graphs could be caused by many reasons such as the delay of reports, testing capacity, hospital capacity, diagnostic methods and etc. For instance, the daily confirmed cases for France, Germany, Switzerland and New York suddenly declined on a regular basis, which mostly happened during the weekends. Even so, we can see the model can still capture the trend of infectious evolution. In Figure 9, we also compare the actual total confirmed COVID-19 cases against the cumulative estimated cases with a confidence interval within two standard deviations[1] for some typical countries and territories. The black dash line in each graph of Figure 9 represents the end of data collection period for calibration. The red curve on the left of the black dash line shows the historical data used for calibration and the blue curve is the corresponding estimated result. The red and blue

---

[1] The standard deviations can be derived based on results in Proposition 3.1 and 3.2.
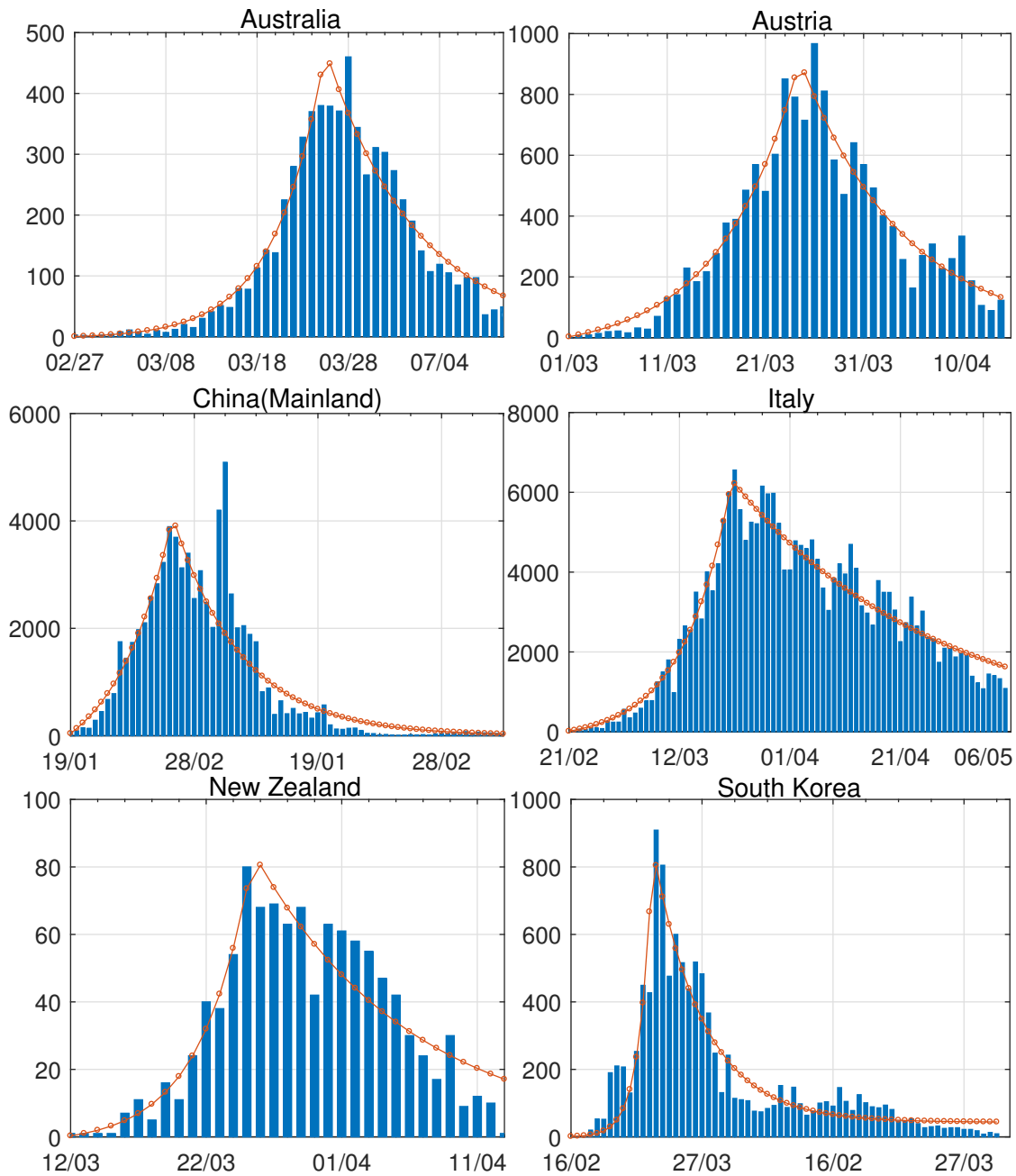
Figure 7: Model calibration comparisons between the expected daily confirmed cases under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ in Table 3 and actual daily confirmed COVID-19 cases for Australia, Austria, China(Mainland), Italy, New Zealand, and South Korea.

curves on the right of the black dash line demonstrate a comparison between the predicted and actual confirmed COVID-19 cases for countries and regions from the end of their data collection period to the end of May, 2020. We can see that the estimated curves of the number of confirmed infections under the two-phase dynamic contagion model well fitted the actual propagation process of the COVID-19. In addition, the forecasted infection cases in the coming weeks after the end of data collection period also well suited the up to date actual total confirmed COVID-19 cases.
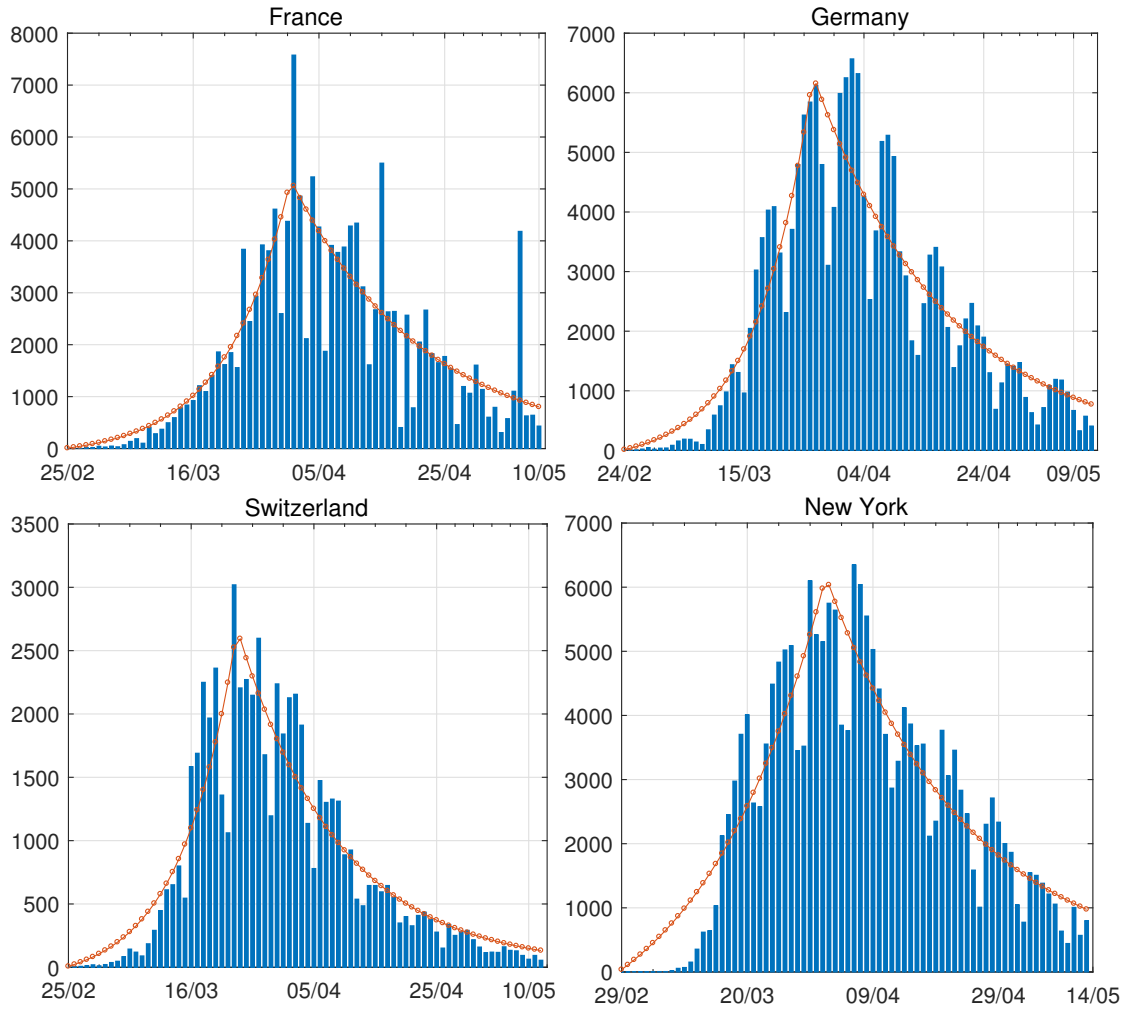
Figure 8: Model calibration comparisons between the expected daily confirmed cases under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ in Table 3 and actual daily confirmed COVID-19 cases for France, Germany, Switzerland, New York.

4.1 Elimination Probability and Final Epidemic Size

According to Proposition 3.3, one could obtain the elimination probability of the epidemic by numerically solving the ODE (3.2). Based on the calibration parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ provided in Table 3 for regions and countries, we could obtain the associated elimination probabilities. Figure 10 illustrates how $\mathbb{P}\left(\widetilde{T} \leq t | \mathcal{F}_{\hat{\ell}}\right)$ varies for different countries and territories. We can see that for regions and countries with effective restriction measures, the probability for a shorter period to observe the last ever event arrives after government interventions come into effect will be much higher. On contrary, for some regions and countries, longer periods are needed for elimination probabilities to be closed to 1. For instance, we can see that there is still a long way to go to end the COVID-19 pandemic for Italy.
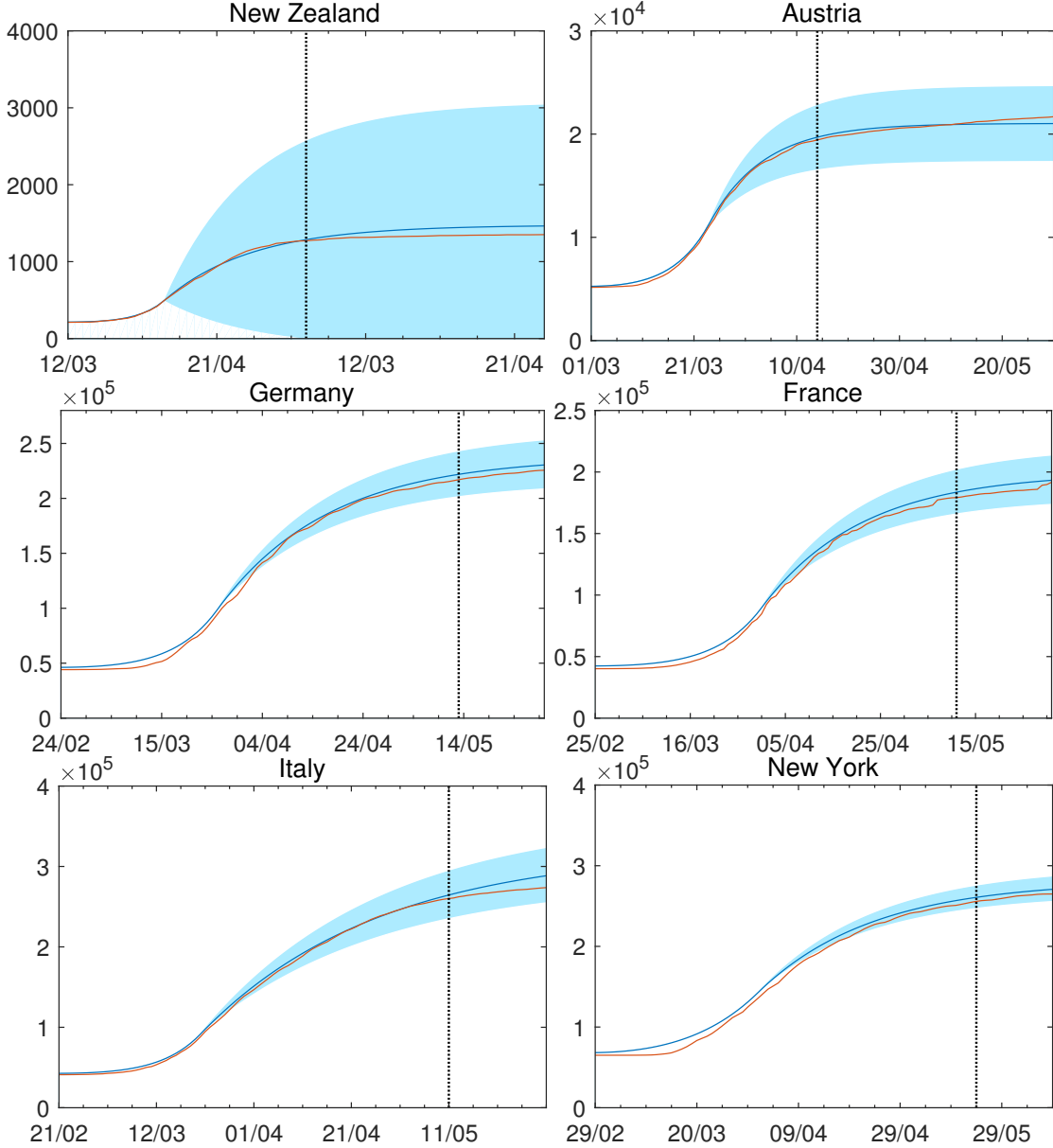
Figure 9: Comparisons between total confirmed COVID-19 cases and total estimated cases under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ in Table 3 from mid-February onward for New Zealand, Austria, Germany, France, Italy, and New York. The red curve represents total confirmed COVID-19 cases. The blue curve represents the total estimated cases and the left zone of the black dash line illustrates historical data used for calibration, and the right zone demonstrates the predicted estimated cases. The shadowed region plots the values within two standard deviations.

The elimination time of the pandemic depends on many decisive factors, such as the initial intensity of the externally-exciting jumps, the time needed for the government interventions to come into effect, the size of the branching ratio after the government interventions came into effect, etc. Figure 11, 12, and 13 illustrate comparisons between the estimated $\hat{\ell}$, $\hat{\varrho}$, R_a against $\mathbb{E}[\widetilde{T}|\mathcal{F}_{\hat{\ell}}]$, respectively. From Figure 11, we can see that for most countries and territories, the quicker the government interventions come into effect, the faster the pandemic will end. However, some places like Hong Kong and Iceland still have relatively fast elimination time even though it takes longer for the
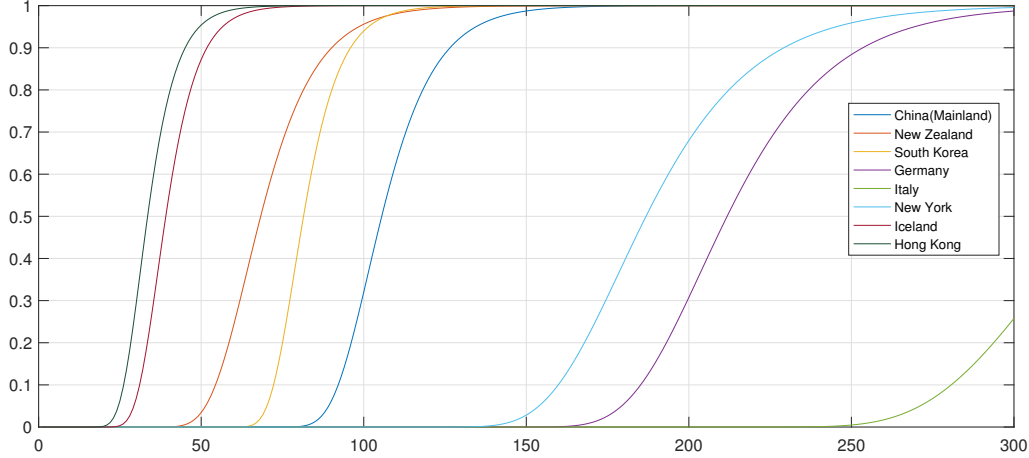
Figure 10: Conditional elimination probability $\mathbb{P}(\widetilde{T} \leq t | \mathcal{F}_{\hat{\ell}})$ for China (Mainland), New Zealand, South Korea, Germany, Italy, New York, Iceland and Hong Kong under the associated calibration parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ for these regions and countries suggested in Table 3.

government interventions to come into effect. This is probably because the restriction measures for these places were imposed so early that reporting procedures were not properly in place yet. Figure 12, and 13 clearly demonstrate that the externally-exciting jump intensity ϱ and the branching ratio after the government came into effect R_a are important factors that determine the extinction time of the pandemic. In general, to reduce the extinction time of the pandemic, the first priority for the authorities should be introducing restriction measures such as national/subnational lockdown to reduce the intensity of the external imported cases, and while the external imported cases are controlled and thereafter negligible, the governments should simultaneously introduce enforced restrictions to prevent further transmission. If the government intervention strategies were effectively implemented without a lack of civic spirit, the infection rate of the virus after the these intervention measures come into place will be reduced significantly, and therefore lead to a quicker elimination of the COVID-19 pandemic. Note that the prediction for expected elimination time for regions and countries is based on the assumption that the government intervention measures are still taking into place in some form and propagation of the disease continues as in Phase 2. Relaxation of the government intervention measures will inevitably delay the disease elimination for most regions and countries.

Beside the conditional probability for the elimination time $\widetilde{T}$, the epidemic size N_t given $\{\tilde{T} \leq t\}$ can also be predicted according to the join expectation of N_t and $\{\widetilde{T} \leq t\}$ derived in Corollary 3.1. In Table 5, we report the 95% confidence interval for elimination time $\widetilde{T}$ the condition expectation of the elimination time $\widetilde{T}$, $\mathbb{E}[\widetilde{T} | \mathcal{F}_{\hat{\ell}}]$, the expected elimination date $Date_E$, and the conditional expectation of the epidemic size N_t, $\mathbb{E}[N_t | \mathcal{F}_{\hat{\ell}} \cap \{\widetilde{T} \leq t\}]$ with $t = \mathbb{E}[\widetilde{T} | \mathcal{F}_{\hat{\ell}}]$,
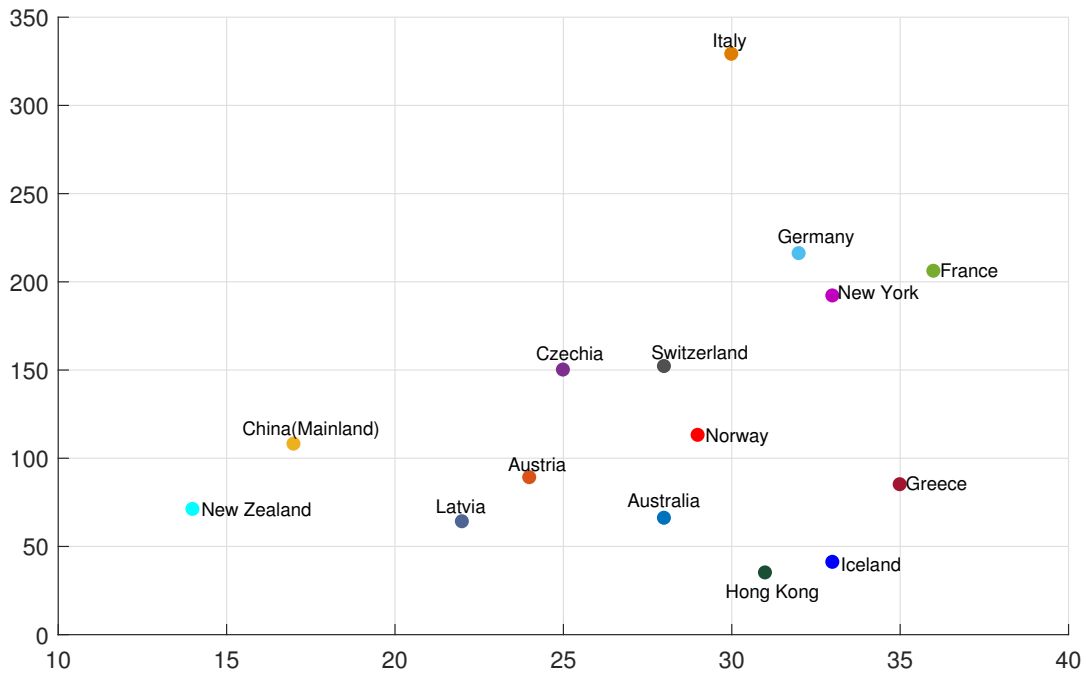
Figure 11: Comparison of the conditional expectation of the elimination time $\widetilde{T}$ and the estimated government interventions came into effect time for different regions and countries around the world. The horizontal axis represents $\hat{\ell}$, and the vertical axis represents $\mathbb{E}[\widetilde{T}|\mathcal{F}_{\hat{\ell}}]$.
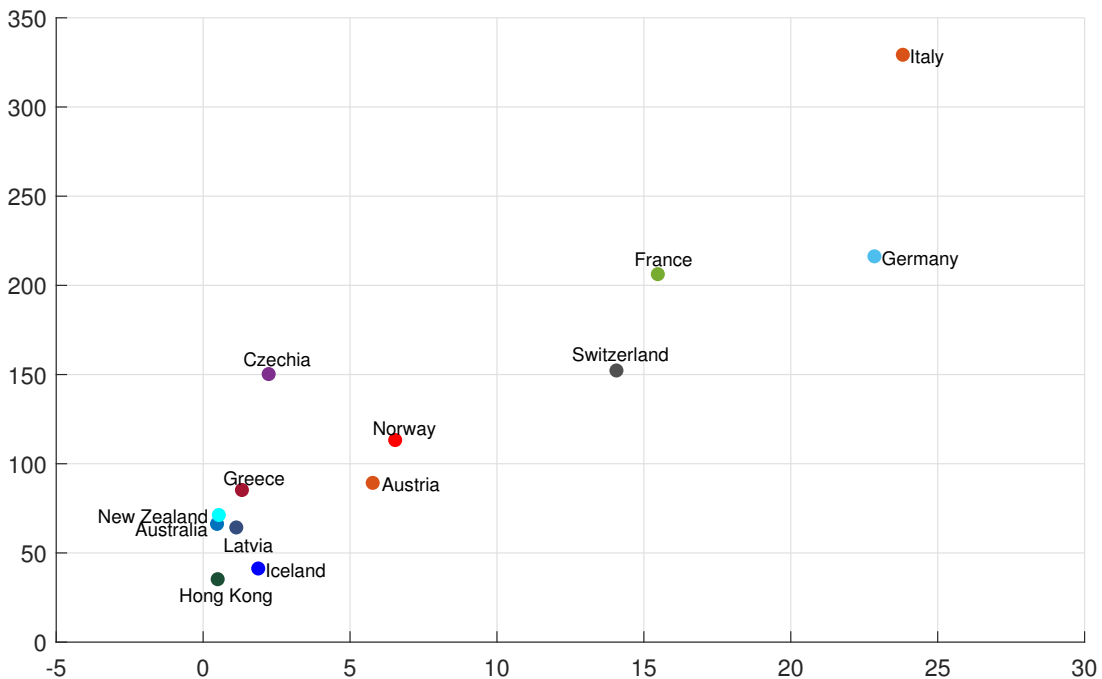
Figure 12: Comparison of the conditional expectation of the elimination time $\widetilde{T}$ and the estimated intensity for externally-exciting jumps for different regions and countries around the world. The horizontal axis represents $\hat{\varrho}$, and the vertical axis represents $\mathbb{E}[\widetilde{T}|\mathcal{F}_{\hat{\ell}}]$.
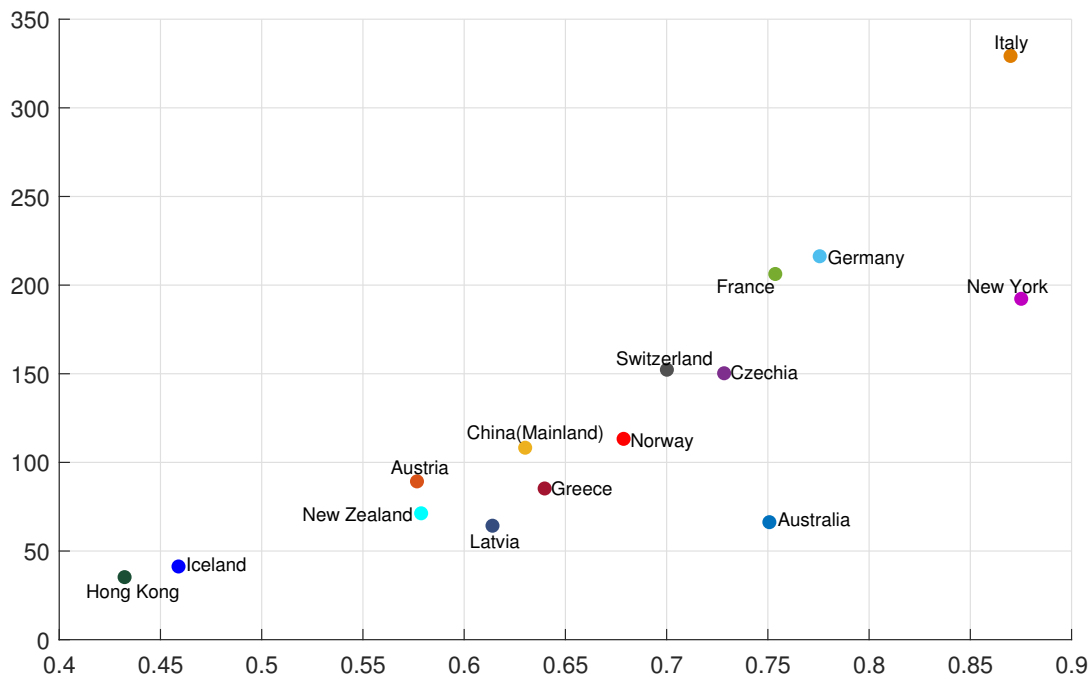
Figure 13: Comparison of the conditional expectation of the elimination time $\widetilde{T}$ and the branching ratio *after* the government interventions came into effect for different regions and countries around the world. The horizontal axis represents R_a, and the vertical axis represents $\mathbb{E}[\widetilde{T}|\mathcal{F}_{\hat{\ell}}]$.

for regions and countries with calibration parameters in Table 3. Note that, the regions and countries with more confirmed COVID-19 cases before government interventions came into effect will experience longer time to reach elimination state, like France, Germany, Italy, and New York. And the corresponding expected epidemic size for these areas are also much larger. Note that since we have smoothed the biggest jump of daily confirmed cases, adding up with the cases which have been smoothed, the actual conditional expectation of the epidemic size is about 83113, which is very close to the current total confirmed cases $82,993$ on 2020-05-27. In general, not only the expected epidemic size is very close to the actual total confirmed cases for the listed regions and countries, but also the estimated elimination date is very close to the actual eradicate date. New Zealand is one typical example that can be used to demonstrate the effectiveness of our model in predicting the key epidemiological quantities. New Zealand authority has officially declared that the country has completely eradicated COVID-19 for now with total of 1154 confirmed COVID-19 cases on 2020-06-08. This elimination date and the final epidemic size are very close to what we predicted for New Zealand under the two-phase dynamic contagion model, the predicted elimination date is around 2020-06-04 and the predicted epidemic size is about 1250. More remarkably, the historical data we used for model calibration for New Zealand is from 2020-03-12 to 2020-04-13, which is already one and half months ago. This clearly shows that the two-phase dynamic contagion model

is pretty powerful in forecasting cumulative confirmed COVID-19 cases, predicting possible elimination duration for the pandemic, and evaluating effectiveness of relevant government intervention measures.

Table 5: The 95% confidence interval for elimination time $\widetilde{T}$, the conditional expectation of the elimination time $\widetilde{T}$, the expected elimination date and the conditional expectation of the epidemic size N_t given $\widetilde{T} \leq t$ with $t = \mathbb{E}[\widetilde{T}|\mathcal{F}_{\hat{\ell}}]$ under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ for these regions and countries suggested in Table 3.

Regions	$\mathbb{P}(\widetilde{T} \in (t_1, t_2)|\mathcal{F}_{\hat{\ell}}) = 95\%$	$\mathbb{E}[\widetilde{T}|\mathcal{F}_{\hat{\ell}}]$	$Date_E$	$\mathbb{E}[N_t|\mathcal{F}_{\hat{\ell}} \cap \{\widetilde{T} \leq t\}]$
Australia	$(47, 97)$	66	2020-05-30	7059.46
Austria	$(69, 122)$	89	2020-06-21	15610.22
China (Mainland)	$(87, 142)$	108	2020-05-22	69675.10
Czechia	$(111, 216)$	150	2020-08-25	8932.448
France	$(166, 272)$	206	2020-10-23	154112.00
Germany	$(175, 285)$	216	2020-10-28	189116.50
Greece	$(59, 127)$	85	2020-06-24	2650.07
Hong Kong	$(23, 54)$	35	2020-05-05	945.81
Iceland	$(28, 61)$	41	2020-05-11	1787.79
Italy	$(264, 445)$	329	2021-02-13	276480.1
Latvia	$(40, 102)$	64	2020-05-31	764.80
New York	$(149, 261)$	192	2020-10-10	209311.80
New Zealand	$(49, 107)$	71	2020-06-04	1249.77
Norway	$(83, 162)$	113	2020-07-15	8024.73
Switzerland	$(121, 203)$	152	2020-08-22	33414.19

The conditional distribution of the final epidemic size N_∞ can be obtained by numerically inverse the probability generating function provided in Proposition 3.4. Since we assume the self-exciting jumps follows an exponential distribution after government interventions came into effect, the final epidemic size N_∞ conditional on $\mathcal{F}_\ell$ follows a mixed-Poisson distribution with probability mass function specified in (3.16). Figure 14 demonstrates the conditional probability mass function of the difference between the finial epidemic size N_∞ and the total number of confirm cases N_ℓ when government interventions came into effect, i.e., $\mathbb{P}\left(N_\infty - N_\ell = k \mid \mathcal{F}_\ell\right)$, for some regions and countries under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ for these regions and countries suggested in Table 3.

5   Concluding Remarks

In this paper, we have introduced a two-phase dynamic contagion process for modelling the current spread of COVID-19. This model allows randomness to the infectivity of individuals rather than a constant reproduction number as commonly assumed by standard models. Key episdemiological quantities, such as the distribution of final epidemic size and expected epidemic duration, are derived and estimated based on real data for various regions and countries. In addition, the associated
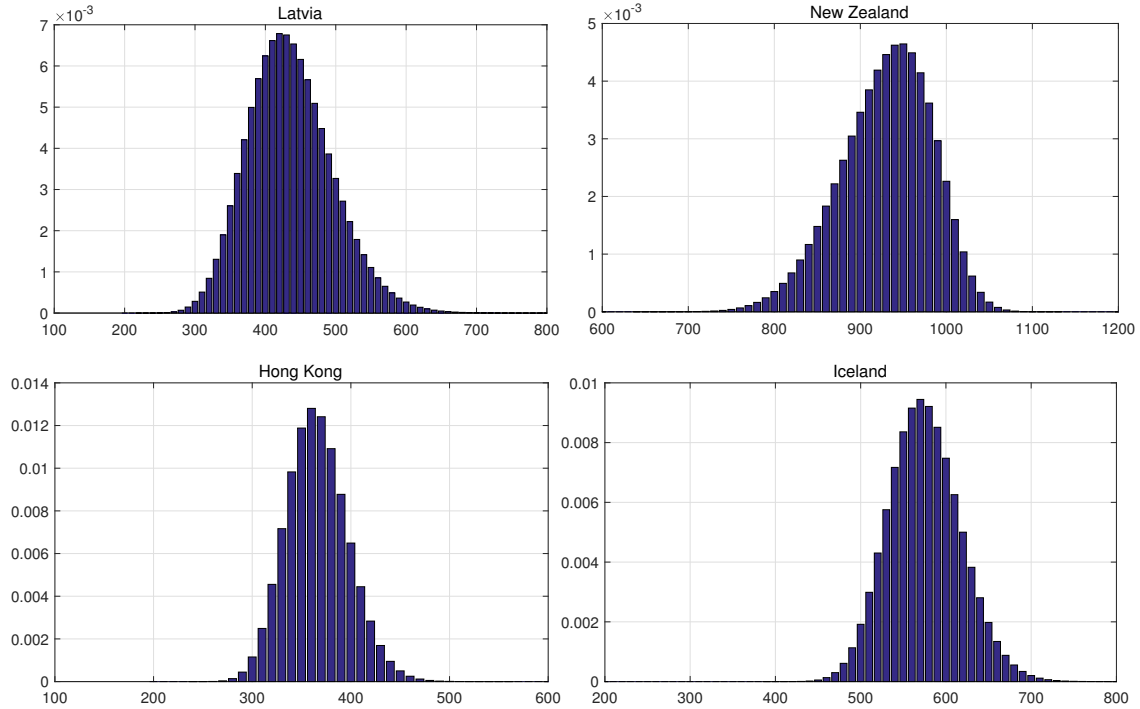
Figure 14: Probability mass function $\mathbb{P}\left(N_\infty - N_{\hat{\ell}} = k \mid \mathcal{F}_{\hat{\ell}}\right)$ for Latvia, New Zealand, Hong Kong and Iceland under the calibrated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$ for these regions suggested in Table 3.

time lag of the effect of intervention in each country or region has been estimated, and our empirical results are consistent to the incubation time of COVID-19 for most people found by existing medical study such as Lauer et al. (2020). The aim of this paper is to demonstrate that our model, as a representative of Hawkes-based processes, could be a valuable tool for epidemic modelling. In fact, the vast literature of Hawkes-based processes would also be relevant and potentially be applicable. For example, multivariate extensions of Hawkes-based processes, such as Embrechts et al. (2011) for analysing financial high-frequency data, could be adopted for modelling the cross-region epidemic contagion. Lévy-driven extensions, such as Qu et al. (2019) for portfolio credit risk, may perform better in capturing the heavy tail property of epidemic distribution. In addition, easing of the government interventions will lead to change of parameters and delay extinction times. The model can be adjusted by introducing an additional phase with change of parameters. When countries cycle between periods of restrictions and relaxations to manage COVID-19, we can adjust the two-phase dynamic contagion model by replacing the constant parameters to piecewise time dependent parameters. These are all proposed as future research.

References

Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. (2020). A multi-risk SIR model with optimally targeted lockdown. National Bureau of Economic Research Working Paper.

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606.

Allen, L. J. (2008). An introduction to stochastic epidemic models. In Allen, L. J., Brauer, F., Van den Driessche, P., and Wu, J., editors, *Mathematical Epidemiology*, chapter 3, pages 81–130. Springer, Berlin.

Alvarez, F. E., Argente, D., and Lippi, F. (2020). A simple planning problem for COVID-19 lockdown. National Bureau of Economic Research Working Paper.

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*. Springer, New York.

Atkeson, A. (2020). What will be the economic impact of COVID-19 in the US? rough estimates of disease scenarios. National Bureau of Economic Research Working Paper.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013a). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013b). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499.

Bailey, N. T. J. (1950). A simple stochastic epidemic. *Biometrika*, 37(3/4):193–202.

Bailey, N. T. J. (1953). The total size of a general stochastic epidemic. *Biometrika*, 40(1/2):177–185.

Bailey, N. T. J. (1957). *The Mathematical Theory of Epidemics*. Griffin, London.

Ball, F. (1983). The threshold behaviour of epidemic models. *Journal of Applied Probability*, 20(2):227–241.

Ball, F., Britton, T., and Neal, P. (2016). On expected durations of birth–death processes, with applications to branching processes and SIS epidemics. *Journal of Applied Probability*, 53(1):203–215.

Ball, F. and Donnelly, P. (1995). Strong approximations for epidemic models. *Stochastic Processes and their Applications*, 55(1):1–21.

Bartlett, M. (1949). Some evolutionary stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):211–229.

Bartlett, M. S. (1956). Deterministic and stochastic models for recurrent epidemics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pages 81–109.

Berger, D. W., Herkenhoff, K. F., and Mongey, S. (2020). An SEIR infectious disease model with testing and conditional quarantine. National Bureau of Economic Research Working Paper.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912.

Brauer, F., Van den Driessche, P., and Wu, J. (2008). *Mathematical Epidemiology*. Springer, Berlin.

Britton, T. (2010). Stochastic epidemic models: a survey. *Mathematical Biosciences*, 225(1):24–35.

Daley, D. J. and Gani, J. (1999). *Epidemic Modelling: An Introduction*. Cambridge University Press, Cambridge.

Dassios, A. and Zhao, H. (2011). A dynamic contagion process. *Advances in Applied Probability*, 43(3):814–846.

Dassios, A. and Zhao, H. (2017a). Efficient simulation of clustering jumps with CIR intensity. *Operations Research*, 65(6):1494–1515.

Dassios, A. and Zhao, H. (2017b). A generalised contagion process with an application to credit risk. *International Journal of Theoretical and Applied Finance*, 20(1):1–33.

Diekmann, O., Heesterbeek, H., and Britton, T. (2013). *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, New Jersey.

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2020). The macroeconomics of epidemics. National Bureau of Economic Research Working Paper.

Embrechts, P., Liniger, T., and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48A:367–378.

Fuchs, C. (2013). *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer-Verlag, Berlin.

Guerrieri, V., Lorenzoni, G., Straub, L., and Werning, I. (2020). Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages? National Bureau of Economic Research Working Paper.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(3):438–443.

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Keeling, M. J. and Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, New Jersey.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society*, 115(772):700–721.

Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., and Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 172(9):577–582.

Liu, Y., Gayle, A. A., Wilder-Smith, A., and Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2):1–4.

Martcheva, M. (2015). *An Introduction to Mathematical Epidemiology*. Springer, New York.

McKendrick, A. (1925). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130.

McNeill, W. H. (1976). *Plagues and Peoples*. Anchor, New York.

Qu, Y., Dassios, A., and Zhao, H. (2019). Efficient simulation of Lévy-driven point processes. *Advances in Applied Probability*, 51(4):927–966.

Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M. U. G., Li, B., Cai, J., Xu, B., Yang, Q., Wang, B., Yang, P., Cui, Y., Song, Y., Zheng, P., Wang, Q., Bjornstad, O. N., Yang, R., Grenfell, B. T., Pybus, O. G., and Dye, C. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, 368(6491):638–642.