

## A two-phase dynamic contagion model for COVID-19

Zezhun Chen<sup>a</sup>, Angelos Dassios<sup>a</sup>, Valerie Kuan<sup>b</sup>, Jia Wei Lim<sup>c</sup>, Yan Qu<sup>d,\*</sup>, Budhi Surya<sup>e</sup>, Hongbiao Zhao<sup>f</sup>

<sup>a</sup> London School of Economics, United Kingdom

<sup>b</sup> University College London, United Kingdom

<sup>c</sup> Brunel University London, United Kingdom

<sup>d</sup> Peking University, China

<sup>e</sup> Victoria University of Wellington, New Zealand

<sup>f</sup> Shanghai University of Finance and Economics, China

### ARTICLE INFO

JEL:

I1

H0

E6

Keywords:

Stochastic intensity model

Stochastic epidemic model

Two-phase dynamic contagion process

COVID-19

Lockdown

MSC:

Primary: 60G55

Secondary: 60J75

### 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 [24]. 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.

### 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 [35]. 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. [21], Keeling and Rohani [34], Diekmann et al. [27] and Martcheva [40]. 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. [19], Liu et al. [39] and Tian et al. [48]. More recently, Acemoglu et al. [1] develop a multi-risk SIR model, which takes into account that different subpopulations have different

risks and is applied to analysing optimal lockdown.

However, there are several limitations to use deterministic epidemic models to describe the transmission dynamics of the COVID-19 pandemic. First, the SIR and SEIR models require the assumptions of the incubation period, the duration of infection, the initial reproduction number, the number of isolated cases once identified, the subclinical infection percentage, etc, which are mostly unclear at the early stage of the pandemic. Second, for most countries and regions around the world, the existing COVID-19 recovery and mortality data are generally underreported, which would lead to underestimations of key epidemiological quantities such as reproduction number. In addition, 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 [41], and then a variety of stochastic models were studied in the literature, e.g. Bartlett [17,18] and Bailey [11–13]. For more recent developments on stochastic epidemic models in general, see Daley and Gani [23], Andersson and Britton [5], Allen [3] and Fuchs [30]. In particular, many researchers adopted

\* Corresponding author.

E-mail address: [y.qu3@lse.ac.uk](mailto:y.qu3@lse.ac.uk) (Y. Qu).

<https://doi.org/10.1016/j.rinp.2021.104264>

Received 12 January 2021; Received in revised form 12 April 2021; Accepted 13 April 2021

Available online 13 May 2021

2211-3797/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

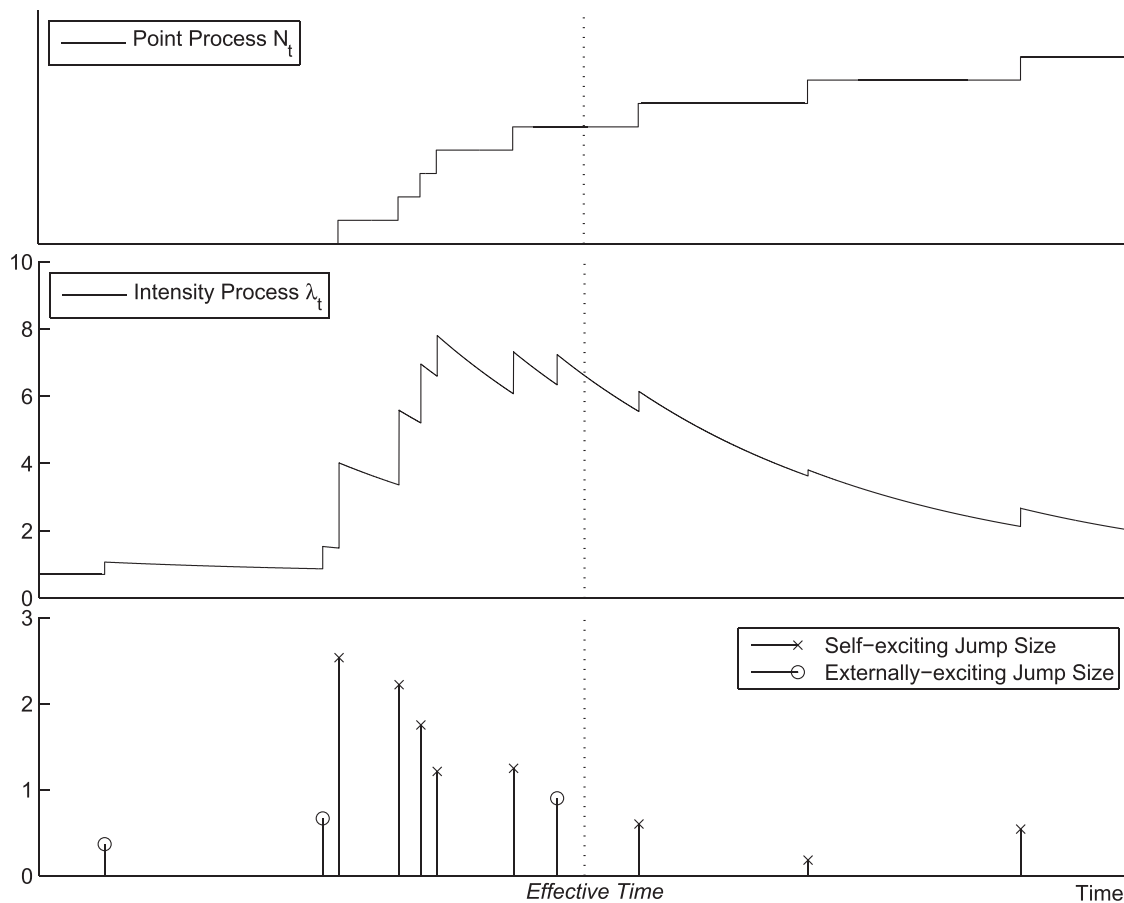


Fig. 1. Two-phase dynamic contagion process.

branching processes. Ball [14] used the birth-and-death process for constructing a sequence of general stochastic epidemics, and Ball and Donnelly [16] used branching processes to approximate the early stages of epidemic dynamics, see also Britton [22] and Ball et al. [15].

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 [24] by introducing an extra phase involving self contagion only, which is an extension of the classical Hawkes process [32,33]. 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 [20], Large [37], Embrechts et al. [29], Bacry et al. [9,10], Ait-Sahalia et al. [2], Dassios and Zhao [25,26] and Qu et al. (2019) [45]. 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. It is very simplistic to assume constant reproduction number as standard epidemic models do. However, we often observe large variations in the estimated reproduction rates within countries as well as over short periods of time. The Hawkes-based model, on the other hand, accommodates such variations as well as the occurrence of super-spreaders of the virus caused by event attending, social gathering, etc. 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 government intervention effect, under the assumption that each different governmental intervention measure imposed on a local region or country was considered as a whole package for each population we analysed. Individual policies such as school closures, physical distancing, shielding of older individuals, and isolation were not considered separately. 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 [42], and have made huge impacts to our society 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. [1], Alvarez et al. [4], Atangana and Atangana [7], Atangana [6], Atkeson [8], Eichenbaum et al. [28], Guerrieri et al. [31], Khan et al. [36], Memon et al. [43], Naik et al. [44], Sene [46,47]. 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

**Table 1**

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

Regions	Parameters					MSE( $\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell}$ ) $\frac{\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell}}{N}$
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{q}$	$\hat{\ell}$	
Heilongjiang	2.431123	8.036146	0.252324	0.934574	16	0.036414
Sichuan	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

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  $\ell$  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 via Dassios and Zhao [24].

**Definition 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

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

where

- $\lambda_0 > 0$  is the initial intensity at time  $t = 0$ ;
- $\delta > 0$  is the constant rate of exponential decay;

**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.

Regions	BR	
	$R_b$	$R_a$
Heilongjiang	1.630176	0.493166
Sichuan	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

- $N_t^* \equiv \{T_k^*\}_{k=1, \dots}$  is a Poisson process of constant rate  $\rho > 0$  arriving in time  $t \leq \ell$ ;
- $\{Z_k\}_{k=1, \dots, N_t^*}$  are independent and identically distributed. *externally-exciting* jump sizes, realised at times  $\{T_k^*\}_{k=1, \dots, N_t^*}$ , with distribution  $H(y)$ ;
- $\{Y_i\}_{i=1, \dots, N_t}$  are independent and identically distributed. *self-exciting* jump sizes of the first phase, realised at times  $\{T_i\}_{i=1, \dots, N_t}$ , 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

$$\lambda_t = \lambda_\ell e^{-\delta(t-\ell)} + \sum_{i=N_\ell+1}^{N_t} Y_i e^{-\delta(t-T_i)}, \quad t \in (\ell, \infty), \quad (2)$$

where

- $\lambda_\ell$  is the initial intensity of the second phase starting at the cutoff time point  $\ell$ , which is the terminal intensity of the first phase;
- $\{Y_i\}_{i=N_\ell+1, \dots}$  are independent and identically distributed. *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  $\ell$ . The point process  $N_t$  and its intensity process  $\lambda_t$  are illustrated in Fig. 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

$$\lambda_t = \lambda_0 e^{-\delta t} + \sum_{k=1}^{N_t^*} Z_k \mathbf{1}_{\{t \leq \ell\}} e^{-\delta(t-T_k^*)} + \sum_{i=1}^{N_t} Y_i e^{-\delta(t-T_i)}, \quad t \geq 0, \quad (3)$$

where

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

$$G(y; t) = G_1(y) \mathbf{1}_{\{t \leq \ell\}} + G_2(y) \mathbf{1}_{\{t > \ell\}}, \quad (4)$$

where smaller jumps represent usual community infection and large jumps can be caused by superspreaders of the virus and other exceptional circumstances.

- $\{Z_k\}_{k=1, \dots, N_t^*}$  are independent and identically distributed. *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  $\{T_k^*\}_{k=1, \dots}$  with sizes (marks)  $\{Z_k\}_{k=1, \dots}$ , and they disappear after time point  $\ell$  when the interventions took effect, and there will be no any increase of imported cases in a long run. The exponential decay assumption reflects that fact that carriers of the virus will eventually recover.

- 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 (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 (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  $\lambda_t$  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.

#### 3.1. Joint distribution of $(\lambda_t, N_t)$

Let  $\{\mathcal{F}_t\}_{t \geq 0}$  be the natural filtration of the point process  $N_t$ , i.e.  $\mathcal{F}_t = \sigma(N_s, s \leq t)$  and assume that the intensity process  $\{\lambda_t\}_{t \geq 0}$  being  $\mathcal{F}_t$ -adapted. The joint Laplace transform and probability generating function for  $(\lambda_t, N_t)$  is provided in Theorem 1 as below with the proof being outlined in Appendix A.

**Theorem 1.** For time  $s \leq t$ , the conditional joint Laplace transform and probability generating function for  $\lambda_t$  and the point process  $N_t$  is of the form

$$\mathbb{E}[\theta^{N_t} e^{-\nu \lambda_t} | \mathcal{F}_s] = \begin{cases} \theta^{N_s} e^{c(s)} e^{-A(s)\lambda_s}, & 0 \leq s \leq t \leq \ell, \\ \theta^{N_s} e^{-A(s)\lambda_s}, & \ell < s \leq t, \end{cases} \quad (5)$$

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

$$A'(s) - \delta A(s) + 1 - \theta \widehat{g}(A(s); s) = 0, \quad (6)$$

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

$$\widehat{g}(u; t) = \int_0^\infty e^{-uy} dG(y; t),$$

and  $c(t)$  is determined by

$$c(t) = \varrho \mathbf{1}_{\{t \leq \ell\}} \int_0^t [1 - \widehat{h}(A(u))] du, \quad (7)$$

with

$$\widehat{h}(u) = \int_0^\infty e^{-uy} dH(y).$$

The moments of  $\lambda_t$  and  $N_t$  can be obtained by differentiating the joint Laplace transform and probability generating function of  $\lambda_t$  and  $N_t$ . Based on Theorem 3.6, Lemma 3.1, Lemma 3.2 and Theorem 3.9 in Dassios and Zhao [24], we obtain the following results for the moments of  $\lambda_t$  and  $N_t$  in Proposition 1 and 2.

**Proposition 1.** The conditional expectation of the process  $\lambda_t$  given  $\mathcal{F}_s$  for  $s \leq t$  is given by

$$\mathbb{E}[\lambda_t | \mathcal{F}_s] = \begin{cases} \frac{\varrho \mu_H \mathbf{1}_{\{t \leq \ell\}}}{\kappa} + \left( \lambda_s - \frac{\varrho \mu_H \mathbf{1}_{\{t \leq \ell\}}}{\kappa} \right) e^{-\kappa(t-s)}, & \kappa \neq 0, \\ \lambda_s + \varrho \mu_H \mathbf{1}_{\{t \leq \ell\}}(t-s), & \kappa = 0, \end{cases} \quad (8)$$

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

$$\mathbb{E}[N_t | \mathcal{F}_s] = \begin{cases} N_s + \frac{\varrho \mu_H \mathbf{1}_{\{t \leq \ell\}}(t-s)}{\kappa} + \left( \lambda_s - \frac{\varrho \mu_H \mathbf{1}_{\{t \leq \ell\}}}{\kappa} \right) \frac{1 - e^{-\kappa(t-s)}}{\kappa}, & \kappa \neq 0, \\ N_s + \lambda_s(t-s) + \frac{1}{2} \varrho \mu_H \mathbf{1}_{\{t \leq \ell\}}(t-s)^2, & \kappa = 0, \end{cases} \quad (9)$$

where

$$\mu_H = \int_0^\infty y dH(y), \quad \mu_G = \int_0^\infty y dG_1(y) \mathbf{1}_{\{t \leq \ell\}} + \int_0^\infty y dG_2(y) \mathbf{1}_{\{t > \ell\}},$$

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

**Proposition 2.** The conditional second moment of the process  $\lambda_t$  given  $\mathcal{F}_s$  for  $s \leq t$  is given by

$$\mathbb{E}[\lambda_t^2 | \mathcal{F}_s] = \begin{cases} \lambda_s^2 e^{-2\kappa t} + \frac{2\varrho \mu_H \mathbf{1}_{\{t \leq \ell\}} + \mu_{2G}}{\kappa} \left( \lambda_s - \frac{\varrho \mu_H}{\kappa} \right) (e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}) \\ + \left( \frac{(2\varrho \mu_H + \mu_{2G}) \varrho \mu_H \mathbf{1}_{\{t \leq \ell\}}}{2\kappa^2} + \frac{\varrho \mu_{2H} \mathbf{1}_{\{t \leq \ell\}}}{2\kappa} \right) (1 - e^{-\kappa(t-s)}), & \kappa \neq 0, \\ \lambda_s^2 + \lambda_s \mu_{2G} t + (2\lambda_0 \mu_H + \mu_{2H}) \varrho \mathbf{1}_{\{t \leq \ell\}}(t-s) \\ + \left( \varrho^2 \mu_H^2 + \frac{1}{2} \varrho \mu_H \mu_{2G} \right) \mathbf{1}_{\{t \leq \ell\}}(t-s)^2, & \kappa = 0, \end{cases} \quad (10)$$

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

$$\mathbb{E}[\lambda_t N_t | \mathcal{F}_s] = \begin{cases} \lambda_s N_s e^{-\kappa(t-s)} + \left( \lambda_s \mu_G + \frac{\lambda_s \mu_{2G}}{\kappa} \right) (t-s) e^{-\kappa(t-s)} \\ + \left( \frac{\lambda_s^2}{\kappa} - \frac{\lambda_s \mu_{2G}}{\kappa} \right) (e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}), & \kappa \neq 0, \\ \lambda_s N_s + (\lambda_s^2 + \lambda_s \mu_G) (t-s) + \frac{1}{2} \lambda_s \mu_{2G} (t-s)^2, & \kappa = 0 \end{cases} \quad (11)$$

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

$$\mathbb{E}[N_t^2 | \mathcal{F}_s] = \begin{cases} \left( \frac{\lambda_s \mu_{2G}}{\kappa^3} - \frac{\lambda_s^2}{\kappa^2} \right) (1 - e^{-2\kappa(t-s)}) - \left( \frac{2\lambda_s \mu_G}{\kappa} + \frac{2\lambda_s \mu_{G2}}{\kappa^2} \right) (t-s) e^{-\kappa(t-s)} \\ + \left( \frac{\lambda_s + 2\lambda_s N_s}{\kappa} + \frac{2\lambda_s^2 + 2\lambda_s \mu_{2G}}{\kappa^2} \right) (1 - e^{-\kappa(t-s)}), & \kappa \neq 0, \\ \lambda_s (\mu_G + 2N_s) (t-s) + (\lambda_s^2 + \lambda_s \mu_G) (t-s)^2 + \frac{1}{3} \lambda_s \mu_{2G} (t-s)^3, & \kappa = 0, \end{cases} \quad (12)$$

where

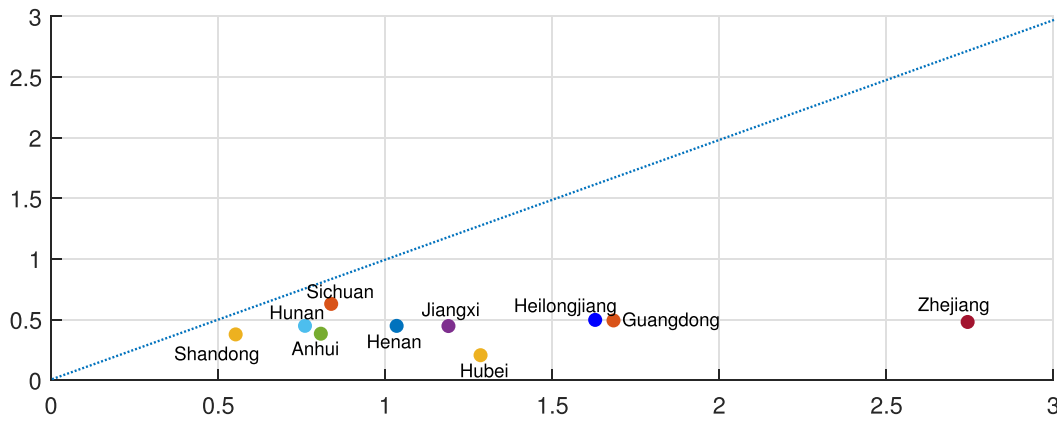


Fig. 2. Comparison of the branching ratios before and after the government interventions came into effect for regions in China. The horizontal axis represents the branching ratio before the government interventions came into effect, namely  $R_b$  and the vertical axis represents the branching ratio after the government interventions came into effect, i.e.  $R_a$ .

$$\mu_{2H} = \int_0^\infty y^2 dH(y), \quad \mu_{2G} = \int_0^\infty y^2 dG_1(y) \mathbf{1}_{\{t \leq \ell\}} + \int_0^\infty y^2 dG_2(y) \mathbf{1}_{\{t > \ell\}},$$

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

### 3.2. Probability of elimination time $\tilde{T}$

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  $\tilde{T}$  to be elimination time such that

$$\tilde{T} := \inf\{t > \ell : \forall s \geq t, N_s - N_t = 0\}. \tag{13}$$

The condition probability of the elimination time is provided in Proposition 3 with the proof being presented in Appendix B.

**Proposition 3.** For  $\ell \leq s \leq t$ , the elimination probability is given by

$$\mathbb{P}(\tilde{T} \leq t | \mathcal{F}_s) = e^{-A(s)\lambda_s}, \tag{14}$$

where  $A(s)$  is determined by the ODE in (6) with boundary condition  $A(t) = \frac{1}{\delta}$ .

### 3.3. Joint expectation of epidemic size $N_t$ and elimination time $\tilde{T}$

Given the last ever event  $\{\tilde{T} < t\}$ , based on Theorem 1 and Proposition 3, one could obtain the expected size of the epidemic at time  $t$ . The relevant details are presented in Corollary 1.

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

$$\mathbb{E}[N_t \mathbf{1}_{\{\tilde{T} \leq t\}} | \mathcal{F}_s] = \frac{d}{d\theta} \{\theta^{N_s} e^{-A(s)\lambda_s}\} |_{\theta=1^-}, \tag{15}$$

where  $A(s)$  satisfies the ODE in (6) with boundary condition  $A(t) = \frac{1}{\delta}$ .

### 3.4. Distribution of final epidemic size $N_\infty$

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  $N_\infty$  can be characterised by Proposition 4 as below [24] [Theorem 3.5].

**Proposition 4.** For  $\ell < s$ , the probability generating function of  $N_\infty$  conditional on  $\mathcal{F}_s$  is given as

$$\mathbb{E}[\theta^{N_\infty} | \mathcal{F}_s] = e^{-v^* \lambda_s}, \tag{16}$$

where

$$v^* = \frac{1}{\delta} \left( 1 - \theta \int_0^\infty e^{-v^* y} dG_2(y) \right).$$

While the government interventions come into effect, if we assume independent and identically distributed self-exciting jump sizes  $Y_i \sim \text{Exp}(\beta)$  for  $i = N_\ell + 1, \dots$ , then, we have an explicit expression for the probability generating function of  $N_\infty$  as

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

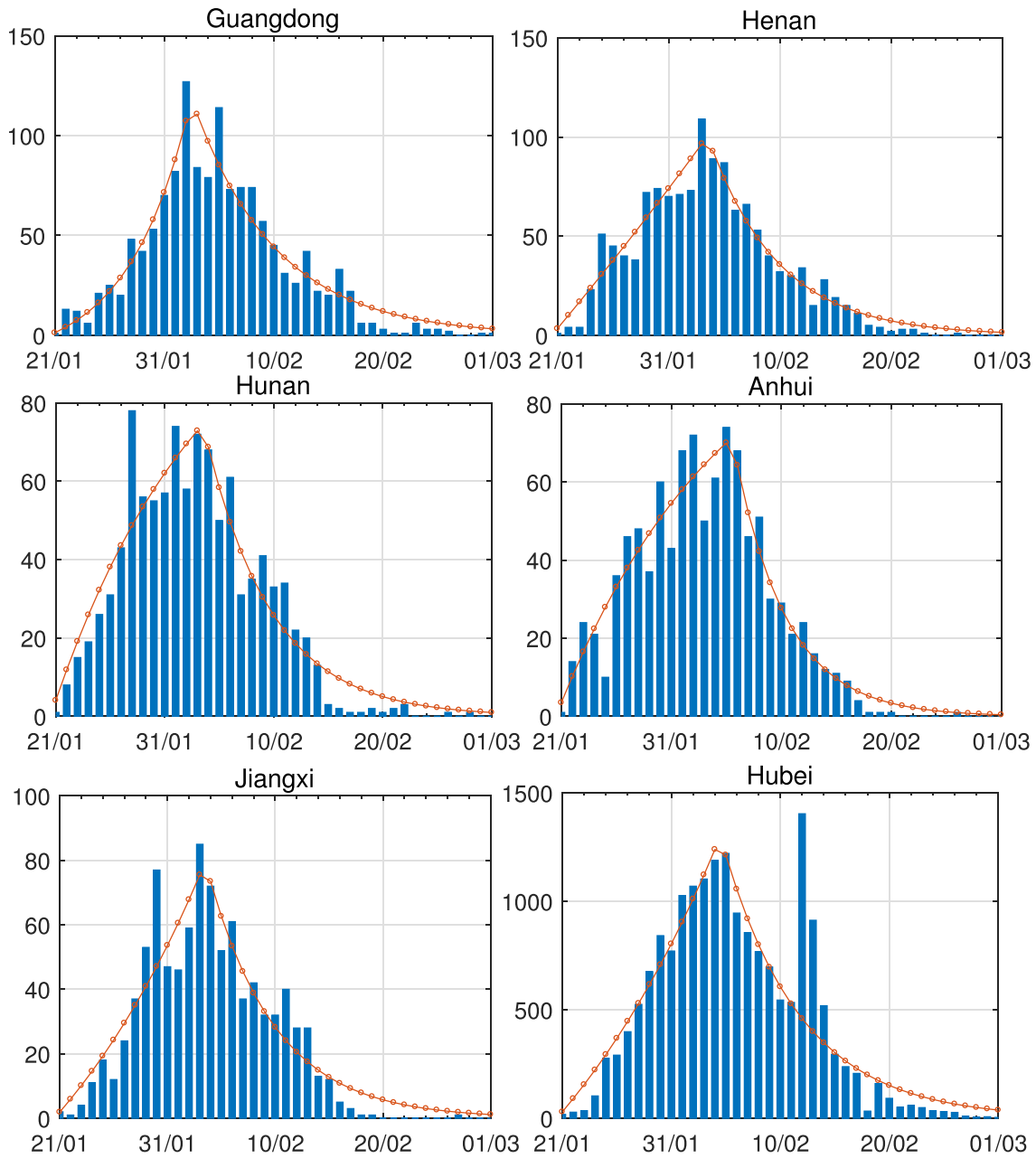
This implies that, the final epidemic size  $N_\infty$  conditional on  $\mathcal{F}_\ell$  follows a mixed-Poisson distribution with the probability mass function

$$\mathbb{P}(N_\infty = k | \mathcal{F}_s) = \int_0^\infty \frac{v^k e^{-v}}{k!} m(v) dv, \quad k = 0, 1, \dots, \tag{17}$$

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

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

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



**Fig. 3.** Model calibration comparisons between the expected daily confirmed cases under the calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  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.

**4. Empirical study**

Without loss of generality, for simplicity, we assume  $Z_k = 1$  for any  $k$ ,  $Y_i \sim \text{Exp}(\alpha)$  for  $i = 1, \dots, N_\ell$  and  $Y_i \sim \text{Exp}(\beta)$  for  $i = N_\ell + 1, \dots$ , and  $\lambda_0 = 0$  in (3) for model calibration. Other assumptions for  $Z_k$ ,  $\{Y_i\}_{i=1, \dots, N_\ell, N_\ell+1, \dots}$ ,  $\lambda_0$  can also be used if necessary. We provide a calibration scheme to estimate  $(\alpha, \beta, \delta, q, \ell)$  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  $\{C_t\}_{t=0,1,2, \dots, T}$ . 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

$$\text{MSE}(\alpha, \beta, \delta, q, \ell) = \frac{1}{T} \sum_{t=0}^{T-1} (\mathbb{E}[N_{t+1} - N_t] - C_{t+1})^2. \tag{18}$$

We consider the calibration based on minimising the MSE (18), i.e., we choose parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  such that

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

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

We provide two empirical examples that are performed in R.4.0.3.

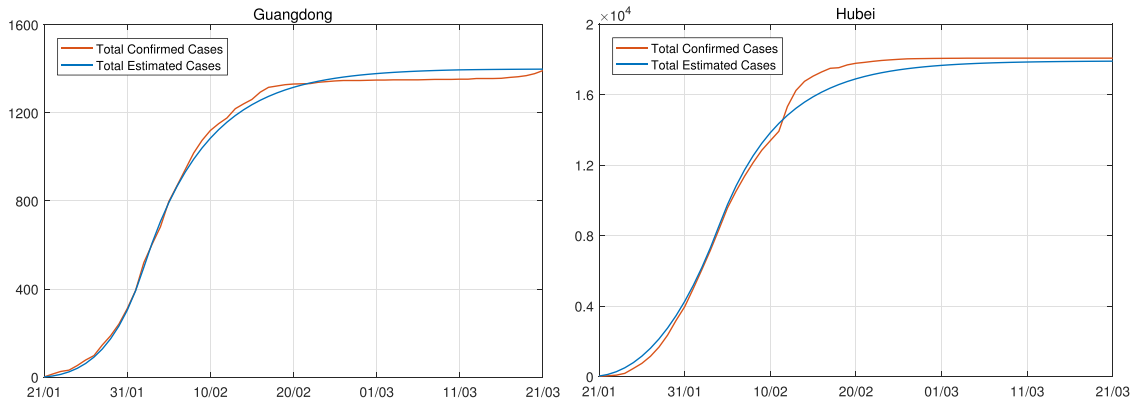


Fig. 4. Comparisons between total confirmed COVID-19 cases and total estimated cases under the calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  in Table 1 from 2020-01-19 to 2020-03-31 for Guangdong and Hubei, China.

Table 3

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

Regions	$Date_0$	$Date_G$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{q}$	$\hat{\ell}$	$MSE(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \frac{\hat{q}, \hat{\ell}}{N})$
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

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.

4.1. 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  $\hat{q}$  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  $\hat{q}$  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. [38].

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

Figs. 3 and 4 demonstrate comparisons between the expected daily/

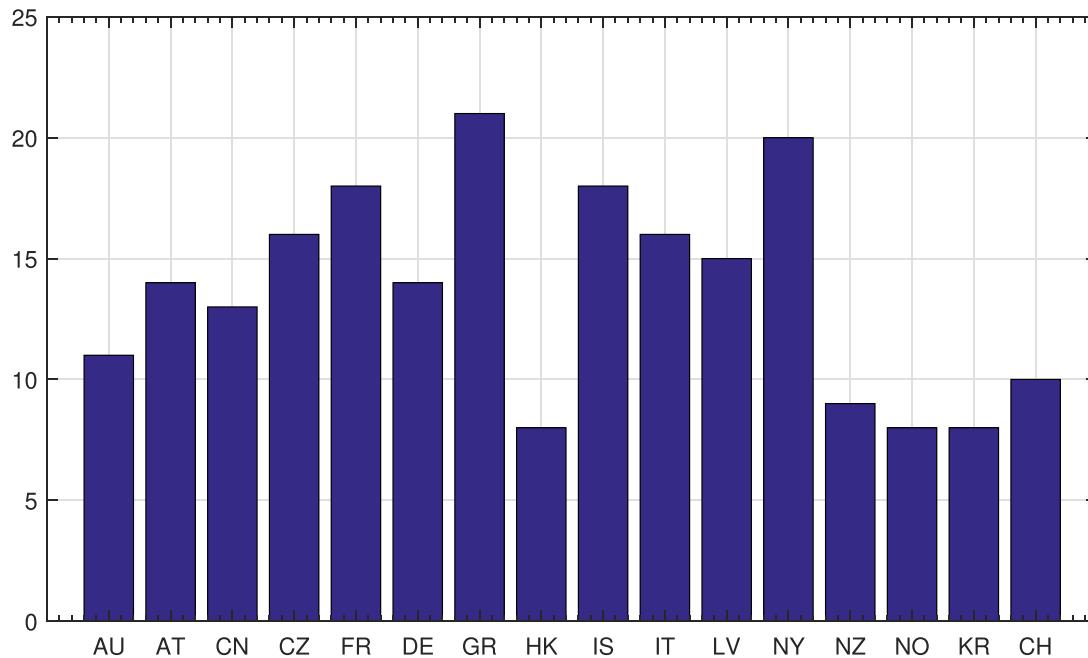


Fig. 5. Comparison of the delay period for different regions and countries around the world. The horizontal axis represents the abbreviation of regions and countries listed in Table 4 and the vertical axis represents the number of days for the government interventions came into effect after the government announcement the relevant measures.

total confirmed cases for the two-phase dynamic contagion model under the calibrated parameter  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  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.

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 world.

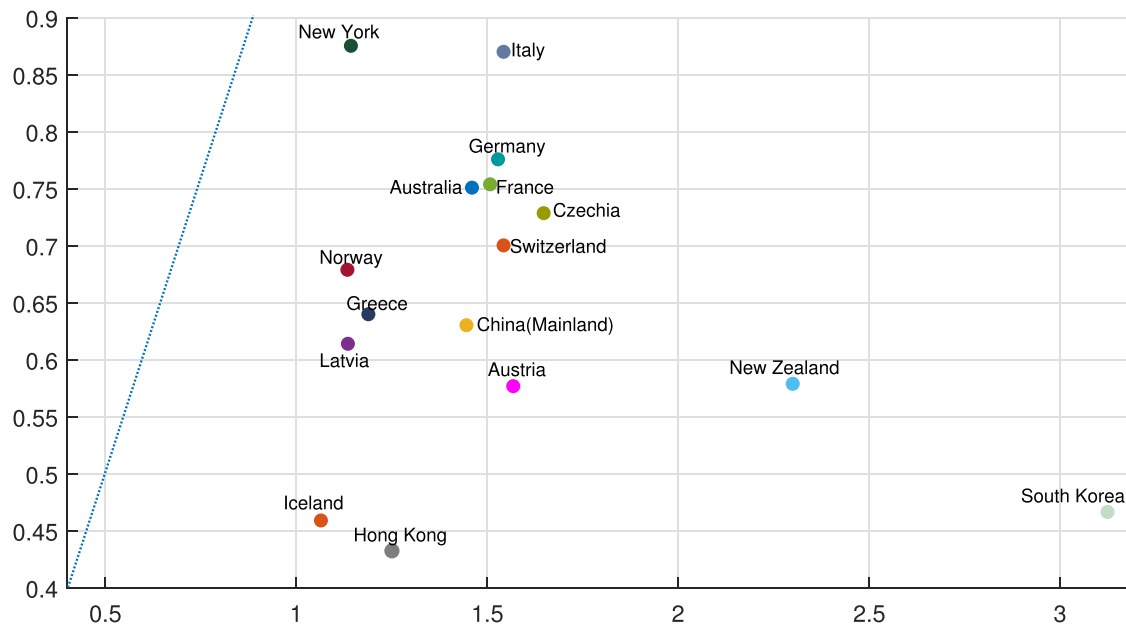
Regions	BR	
	$R_b$	$R_a$
Australia	1.461638	0.750991
Austria	1.569681	0.576945
China(Mainland)	1.447200	0.630380
Czechia	1.649399	0.728637
France	1.509016	0.753908
Germany	1.529826	0.775845
Greece	1.190377	0.639993
Hong Kong	1.252077	0.432526
Iceland	1.066200	0.459162
Italy	1.544364	0.870102
Latvia	1.136798	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

#### 4.2. 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  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  of  $(\alpha, \beta, \delta, q, \ell)$  for various countries and territories. We notice that regions and countries like Italy, China, New York have much larger  $q$  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 Fig. 5. We can see that the delay of the interventions for different regions and countries is around 8 ~ 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





**Fig. 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,  $R_a$ .

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 the 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 Fig. 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.

A comparison between the expected daily confirmed cases for the two-phase dynamic contagion model under the calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\rho})$  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 Fig. 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 Fig. 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 Fig. 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 Fig. 9, we also compare the actual total confirmed COVID-19 cases against the cumulative estimated cases with a confidence interval within two standard deviations<sup>1</sup> for some typical countries and territories. The black dash line in each graph of Fig. 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 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.

#### 4.3. Elimination probability and final epidemic size

According to Proposition 3, one could obtain the elimination probability of the epidemic by numerically solving the ODE (6). Based on the

<sup>1</sup> The standard deviations can be derived based on results in Proposition 1 and 2.

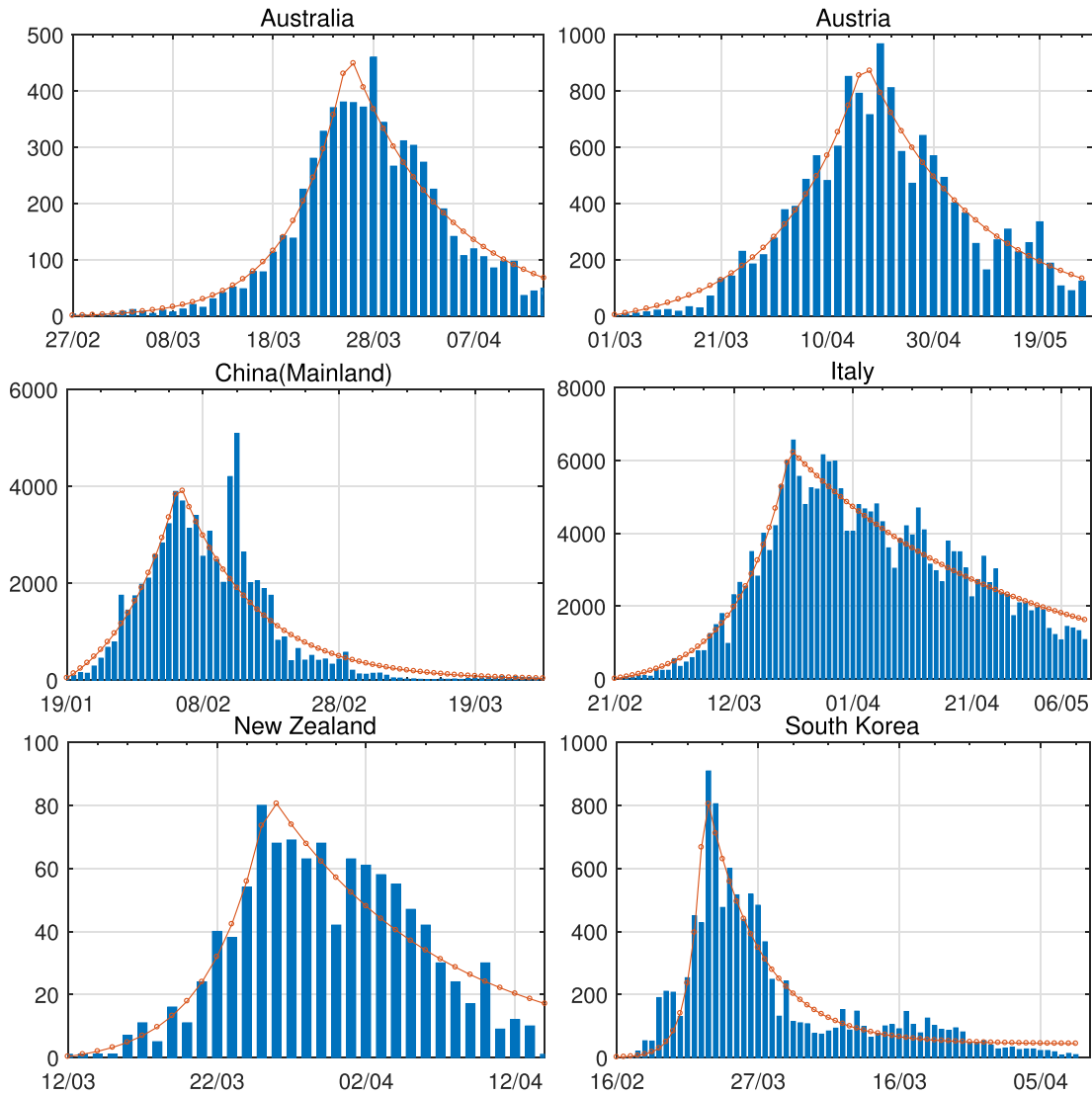
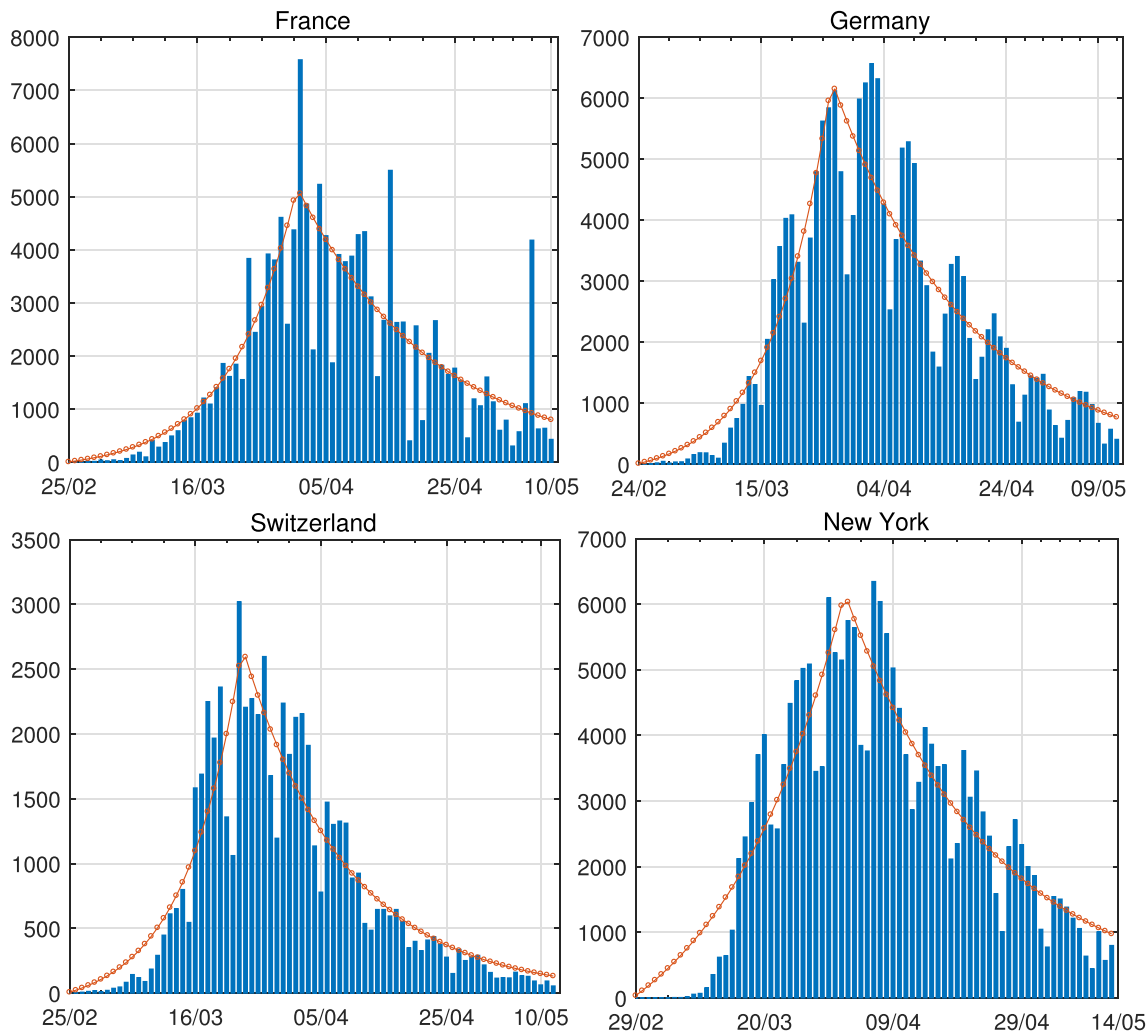


Fig. 7. Model calibration comparisons between the expected daily confirmed cases under the calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$  in Table 3 and actual daily confirmed COVID-19 cases for Australia, Austria, China(Mainland), Italy, New Zealand, and South Korea.

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. Fig. 10 illustrates how  $\mathbb{P}(\tilde{T} \leq t | \mathcal{F}_{\hat{\varrho}})$  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.

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. Fig. 11–13 illustrate comparisons between the estimated  $\hat{\ell}, \hat{\varrho}, R_a$  against  $\mathbb{E}[\tilde{T} | \mathcal{F}_{\hat{\varrho}}]$ , respectively. From Fig. 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 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. Fig. 12, and 13 clearly demonstrate that the externally-exciting jump intensity  $\varrho$  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

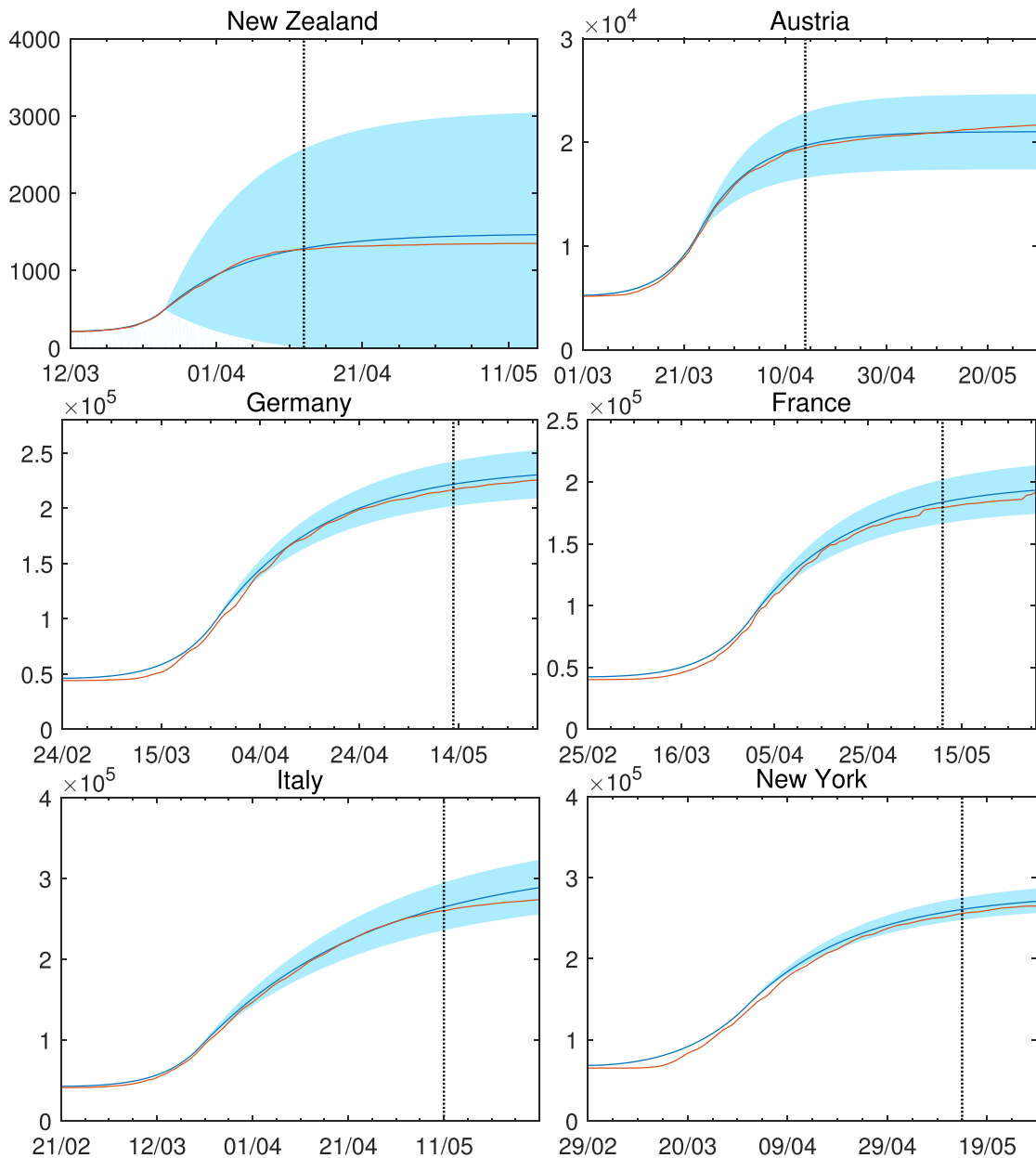


**Fig. 8.** Model calibration comparisons between the expected daily confirmed cases under the calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  in Table 3 and actual daily confirmed COVID-19 cases for France, Germany, Switzerland, New York.

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  $\tilde{T}$ , the epidemic size  $N_t$  given  $\{\tilde{T} \leq t\}$  can also be predicted according to the joint expectation of  $N_t$  and  $\{\tilde{T} \leq t\}$  derived in Corollary 1. In Table 5, we report the 95% confidence interval for elimination time  $\tilde{T}$  the condition expectation of the elimination time  $\tilde{T}, \mathbb{E}[\tilde{T} | \mathcal{F}_\rho]$ , the expected elimination date  $Date_E$ , and the conditional expectation of the epidemic size  $N_t, \mathbb{E}[N_t | \mathcal{F}_\rho \cap \{\tilde{T} \leq t\}]$  with  $t = \mathbb{E}[\tilde{T} | \mathcal{F}_\rho]$ , 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. This clearly shows that

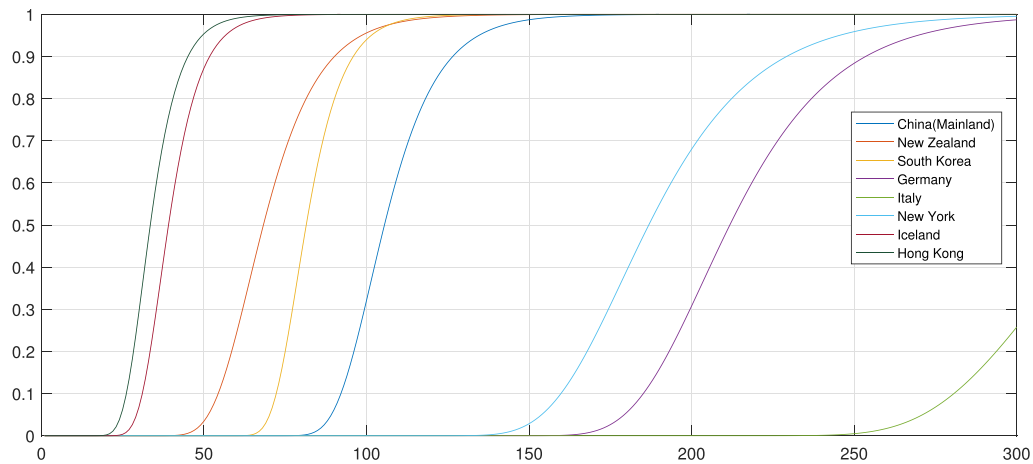


**Fig. 9.** Comparisons between total confirmed COVID-19 cases and total estimated cases under the calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  in Table 3 from mid-February onward for New Zealand, Austria, Germany, France, Italy, and New York. The red curve 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.

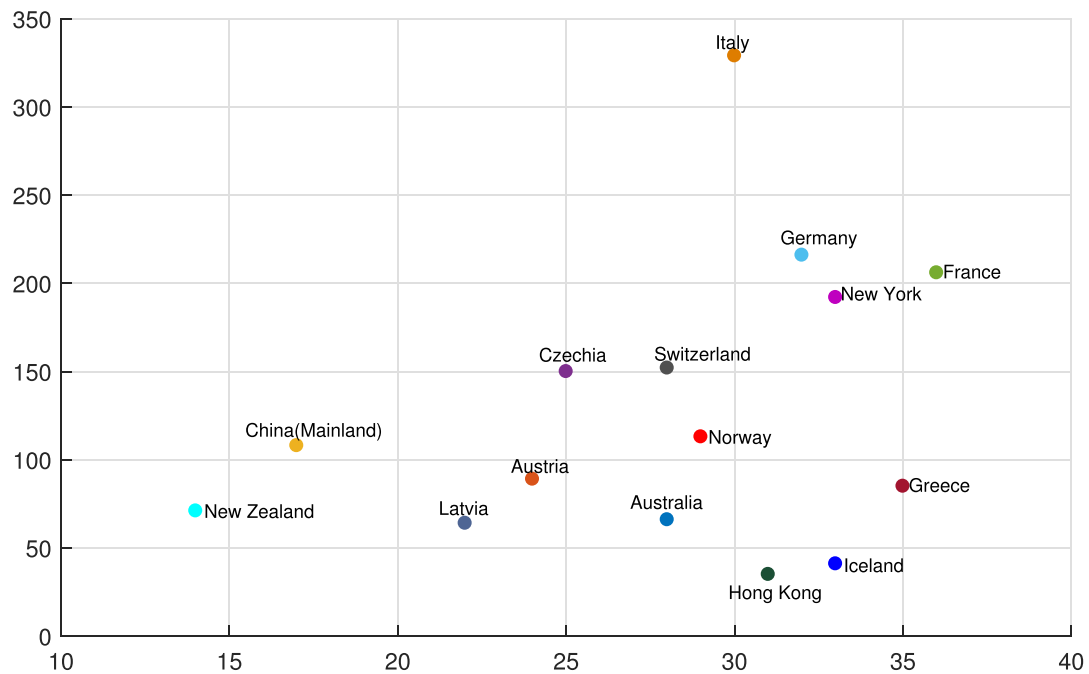
the two-phase dynamic contagion model is powerful in forecasting cumulative confirmed COVID-19 cases, predicting possible elimination duration for the pandemic, and evaluating effectiveness of relevant government intervention measures.

The conditional distribution of the final epidemic size  $N_\infty$  can be obtained by numerically inverse the probability generating function provided in Proposition 4. Since we assume the self-exciting jumps

follows an exponential distribution after government interventions came into effect, the final epidemic size  $N_\infty$  conditional on  $\mathcal{F}_\ell$  follows a mixed-Poisson distribution with probability mass function specified in (17). Fig. 14 demonstrates the conditional probability mass function of the difference between the final epidemic size  $N_\infty$  and the total number of confirm cases  $N_\ell$  when government interventions came into effect, i. e.,  $\mathbb{P}(N_\infty - N_\ell = k | \mathcal{F}_\ell)$ , for some regions and countries under the



**Fig. 10.** Conditional elimination probability  $\mathbb{P}(\tilde{T} \leq t | \mathcal{F}_{\tilde{t}})$  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.



**Fig. 11.** Comparison of the conditional expectation of the elimination time  $\tilde{T}$  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  $\mathbb{E}[\tilde{T} | \mathcal{F}_{\hat{\ell}}]$ .

calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varrho}, \hat{\ell})$  for these regions and countries suggested in Table 3.

For the current situation around the world, most countries have already relaxed some of the intervention measures, which will inevitably lead to further transmissions of the virus and cause a second wave of the COVID-19 pandemic in the near future. Therefore analysing the

key epidemiology quantities for the first wave of the COVID-19 pandemic under the current government intervention measures will also provide further guidance for authorities and public health to response for the second wave of the COVID-19 pandemic.

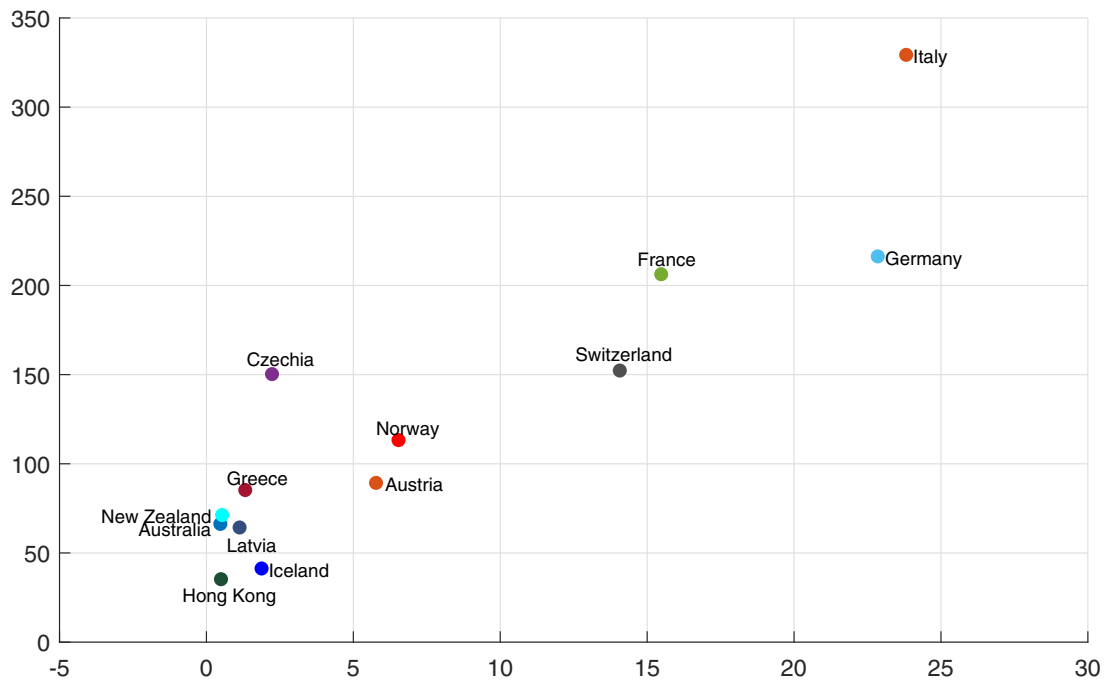


Fig. 12. Comparison of the conditional expectation of the elimination time  $\tilde{T}$  and the estimated intensity for externally-exciting jumps for different regions and countries around the world. The horizontal axis represents  $\hat{\rho}$ , and the vertical axis represents  $E[\tilde{T} | \mathcal{F}_{\hat{\rho}}]$ .

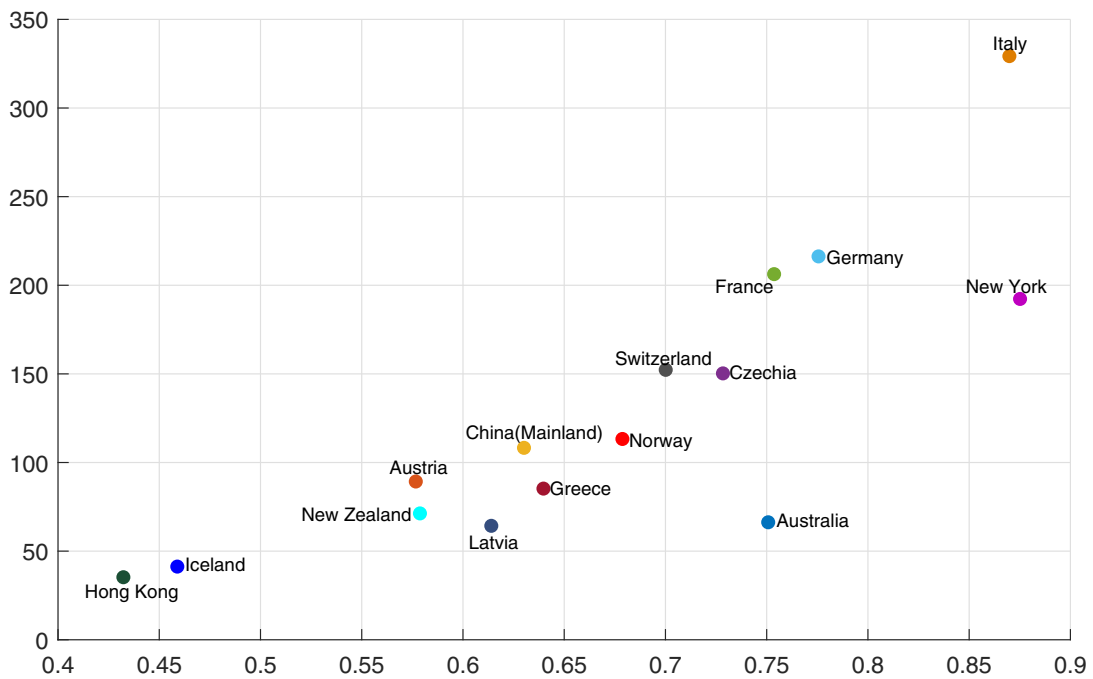


Fig. 13. Comparison of the conditional expectation of the elimination time  $\tilde{T}$  and the branching ratio after the government interventions came into effect for different regions and countries around the world. The horizontal axis represents  $R_a$ , and the vertical axis represents  $E[\tilde{T} | \mathcal{F}_{R_a}]$ .

**Table 5**

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

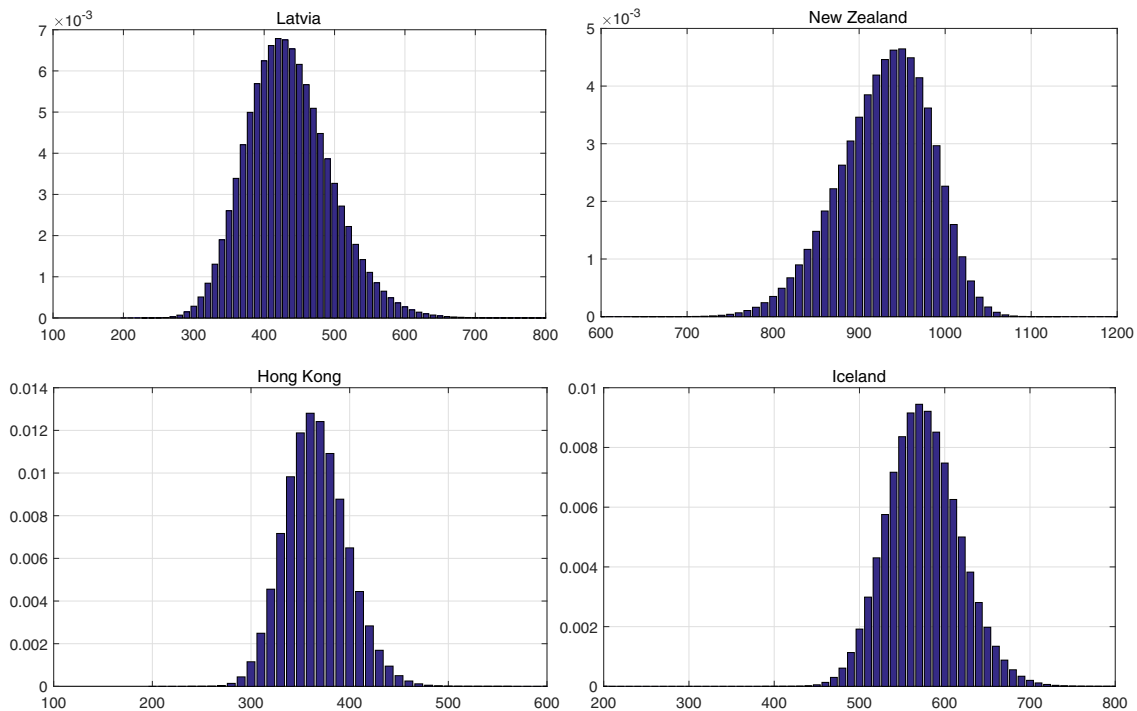
Regions	$\mathbb{P}(\tilde{T} \in (t_1, t_2)   \mathcal{F}_{\tilde{\tau}}) = 95\%$	$\mathbb{E}[\tilde{T}   \mathcal{F}_{\tilde{\tau}}]$	Date <sub>E</sub>	$\mathbb{E}[N_t   \mathcal{F}_{\tilde{\tau}} \cap \{\tilde{T} \leq t\}]$
Australia	(47, 97)	66	2020-05-30	7059.46
Austria	(69, 122)	89	2020-06-21	15610.22
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

**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 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. In addition, the associated 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. [38]. 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. [29] for analysing financial high-frequency data, could be adopted for modelling the cross-region epidemic contagion. Lévy-driven extensions, such as Qu et al. [45] 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.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



**Fig. 14.** Probability mass function  $\mathbb{P}(N_{\infty} - N_{\tilde{\tau}} = k | \mathcal{F}_{\tilde{\tau}})$  for Latvia, New Zealand, Hong Kong and Iceland under the calibrated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{q}, \hat{\ell})$  for these regions suggested in Table 3.

**Acknowledgements**

The author Hongbiao Zhao would like to acknowledge the financial support from the *National Natural Science Foundation of China*

(#71401147) and the research fund provided by the Innovative Research Team of *Shanghai University of Finance and Economics* (#2020110930) and *Shanghai Institute of International Finance and Economics*.

**Appendix A. Proof of Theorem 1**

**Proof.** For  $0 \leq s \leq t \leq \ell$ ,  $\lambda_t$  is the intensity process of the dynamic contagion process introduced in Dassios and Zhao [24]. The corresponding conditional joint Laplace transform, probability generating function for the process  $\lambda_t$  and the point process  $N_t$  is provided in Theorem 3.1 of Dassios and Zhao [24]. For  $\ell < s \leq t$ , given  $\mathcal{F}_s$ , the infinitesimal generator of the dynamic contagion process  $(\lambda_s, N_s, s)$  acting on a function  $f(\lambda, n, s)$  within its domain  $\Omega(\mathcal{A})$  is given by

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

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

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

adding the boundary condition  $A(t) = \nu$ , gives the ODE in (6).

**Appendix B. Proof of Proposition 3**

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

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

Hence, we have

$$\begin{aligned} \mathbb{P}(\widetilde{T} \leq t | \mathcal{F}_s) &= \mathbb{E}[\mathbf{1}_{\{\widetilde{T} \leq t\}} | \mathcal{F}_s] \\ &= \mathbb{E}\left[\exp\left(-\int_t^\infty \lambda_u e^{-\delta(u-t)} du\right) \middle| \mathcal{F}_s\right] \\ &= \mathbb{E}\left[\exp\left(-\frac{\lambda_t}{\delta}\right) \middle| \mathcal{F}_s\right] \end{aligned} \tag{B.1}$$

And according to Theorem 1, by setting  $\theta = 1$  in (5), the result follows immediately.

**References**

[1] Acemoglu D, Chernozhukov V, Werning I, Whinston MD. A multi-risk SIR model with optimally targeted lockdown. National Bureau of Economic Research Working Paper; 2020.

[2] Ait-Sahalia Y, Cacho-Diaz J, Laeven RJ. Modeling financial contagion using mutually exciting jump processes. *J Financial Econ* 2015;117(3):585–606.

[3] Allen LJ. An introduction to stochastic epidemic models. In: Allen LJ, Brauer F, Van den Driessche P, Wu J, editors. *Mathematical Epidemiology*, chapter 3. Berlin: Springer; 2008. p. 81–130.

[4] Alvarez FE, Argente D, Lippi F. A simple planning problem for COVID-19 lockdown. National Bureau of Economic Research Working Paper; 2020.

[5] Andersson H, Britton T. Stochastic epidemic models and their statistical analysis. New York: Springer; 2000.

[6] Atangana A. Modelling the spread of covid-19 with new fractal-fractional operators: Can the lockdown save mankind before vaccination? *Chaos Solitons Fractals* 2020;136:109860.

[7] Atangana E, Atangana A. Facemasks simple but powerful weapons to protect against covid-19 spread: can they have sides effects? *Results Phys* 2020. page 103425.

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

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

[10] Bacry E, Delattre S, Hoffmann M, Muzy J-F. Some limit theorems for Hawkes processes and application to financial statistics. *Stochast Processes Appl* 2013;123(7):2475–99.

[11] Bailey NTJ. A simple stochastic epidemic. *Biometrika* 1950;37(3/4):193–202.

[12] Bailey NTJ. The total size of a general stochastic epidemic. *Biometrika* 1953;40(1/2):177–85.

[13] Bailey NTJ. The mathematical theory of epidemics. London: Griffin; 1957.

[14] Ball F. The threshold behaviour of epidemic models. *J Appl Prob* 1983;20(2):227–41.

[15] Ball F, Britton T, Neal P. On expected durations of birth–death processes, with applications to branching processes and SIS epidemics. *J Appl Prob* 2016;53(1):203–15.

[16] Ball F, Donnelly P. Strong approximations for epidemic models. *Stochast Processes Appl* 1995;55(1):1–21.

[17] Bartlett M. Some evolutionary stochastic processes. *J R Stat Soc Ser B (Methodological)* 1949;11(2):211–29.

[18] Bartlett MS. Deterministic and stochastic models for recurrent epidemics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1956;4:81–109.

[19] Berger DW, Herkenhoff KF, Mongey S. An SEIR infectious disease model with testing and conditional quarantine. National Bureau of Economic Research Working Paper; 2020.

[20] Bowsher CG. Modelling security market events in continuous time: intensity based, multivariate point process models. *J Econometrics* 2007;141(2):876–912.

[21] Brauer F, Van den Driessche P, Wu J. *Mathematical epidemiology*. Berlin: Springer; 2008.

[22] Britton T. Stochastic epidemic models: a survey. *Math Biosci* 2010;225(1):24–35.

[23] Daley DJ, Gani J. *Epidemic modelling: an introduction*. Cambridge: Cambridge University Press; 1999.

[24] Dassios A, Zhao H. A dynamic contagion process. *Adv Appl Prob* 2011;43(3):814–46.

[25] Dassios A, Zhao H. Efficient simulation of clustering jumps with CIR intensity. *Oper Res* 2017;65(6):1494–515.



- [26] Dassios A, Zhao H. A generalised contagion process with an application to credit risk. *Int J Theor Appl Finance* 2017;20(1):1–33.
- [27] Diekmann O, Heesterbeek H, Britton T. *Mathematical tools for understanding infectious disease dynamics*. New Jersey: Princeton University Press; 2013.
- [28] Eichenbaum MS, Rebelo S, Trabandt M. *The macroeconomics of epidemics*. National Bureau of Economic Research Working Paper 2020.
- [29] Embrechts P, Liniger T, Lin L. *Multivariate Hawkes processes: an application to financial data*. *J Appl Prob* 2011;48A:367–78.
- [30] Fuchs C. *Inference for diffusion processes: with applications in life sciences*. Berlin: Springer-Verlag; 2013.
- [31] Guerrieri V, Lorenzoni G, Straub L, Werning I. *Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages?* National Bureau of Economic Research Working Paper; 2020.
- [32] Hawkes AG. Point spectra of some mutually exciting point processes. *J R Stat Soc Ser B (Methodological)* 1971;33(3):438–43.
- [33] Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 1971;58(1):83–90.
- [34] Keeling MJ, Rohani P. *Modeling infectious diseases in humans and animals*. New Jersey: Princeton University Press; 2008.
- [35] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc* 1927;115(772):700–21.
- [36] Khan MA, Atangana A, Alzahrani E, et al. The dynamics of covid-19 with quarantined and isolation. *Adv Differ Equ* 2020;2020(1):1–22.
- [37] Large J. Measuring the resiliency of an electronic limit order book. *J Financial Markets* 2007;10(1):1–25.
- [38] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Internal Med* 2020;172(9):577–82.
- [39] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med* 2020;27(2):1–4.
- [40] Martcheva M. *An introduction to mathematical epidemiology*. New York: Springer; 2015.
- [41] McKendrick A. Applications of mathematics to medical problems. *Proc Edinburgh Math Soc* 1925;44:98–130.
- [42] McNeill WH. *Plagues and peoples*. New York: Anchor; 1976.
- [43] Memon Z, Qureshi S, Memon BR. Assessing the role of quarantine and isolation as control strategies for covid-19 outbreak: a case study. *Chaos Solitons Fractals* 2021; 144:110655.
- [44] Naik PA, Yavuz M, Qureshi S, Zu J, Townley S. Modeling and analysis of covid-19 epidemics with treatment in fractional derivatives using real data from pakistan. *Eur Phys J Plus* 2020;135(10):1–42.
- [45] Qu Y, Dassios A, Zhao H. Efficient simulation of Lévy-driven point processes. *Adv Appl Prob* 2019;51(4):927–66.
- [46] Sene N. Analysis of the stochastic model for predicting the novel coronavirus disease. *Adv Differ Equ* 2020;2020(1):1–19.
- [47] Sene N. Sir epidemic model with mittag-leffler fractional derivative. *Chaos Solitons Fractals* 2020;137:109833.
- [48] Tian H, Liu Y, Li Y, Wu C-H, Chen B, Kraemer MUG, Li B, Cai J, Xu B, Yang Q, Wang B, Yang P, Cui Y, Song Y, Zheng P, Wang Q, Bjornstad ON, Yang R, Grenfell BT, Pybus OG, Dye C. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 2020;368 (6491):638–42.