

## Semi-parametric estimation of incubation and generation times by means of Laguerre polynomials

Alexander Kreiss & Ingrid Van Keilegom

To cite this article: Alexander Kreiss & Ingrid Van Keilegom (2022): Semi-parametric estimation of incubation and generation times by means of Laguerre polynomials, Journal of Nonparametric Statistics, DOI: [10.1080/10485252.2022.2028281](https://doi.org/10.1080/10485252.2022.2028281)

To link to this article: <https://doi.org/10.1080/10485252.2022.2028281>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 02 Feb 2022.



Submit your article to this journal [↗](#)



Article views: 83



View related articles [↗](#)



View Crossmark data [↗](#)

# Semi-parametric estimation of incubation and generation times by means of Laguerre polynomials

Alexander Kreiss<sup>a,b</sup> and Ingrid Van Keilegom<sup>b</sup>

<sup>a</sup>Department of Statistics, London School of Economics, London, UK; <sup>b</sup>Faculty of Economics and Business (FEB), KU Leuven, Leuven, Belgium

## ABSTRACT

In epidemics many interesting quantities, like the reproduction number, depend on the incubation period (time from infection to symptom onset) and/or the generation time (time until a new person is infected from another infected person). Therefore, estimation of the distribution of these two quantities is of distinct interest. However, this is a challenging problem since it is normally not possible to obtain precise observations of these two variables. Instead, in the beginning of a pandemic, it is possible to observe for transmission pairs the time of symptom onset for both people as well as a window for infection of the first person (e.g. because of travel to a risk area). In this paper we suggest a simple semi-parametric sieve-estimation method based on Laguerre-Polynomials for estimation of these distributions. We provide detailed theory for consistency and illustrate the finite sample performance for small datasets via a simulation study.

## ARTICLE HISTORY

Received 31 December 2020  
Accepted 30 December 2021

## KEYWORDS

Laguerre-polynomials;  
semi-parametric estimation;  
sieve-estimation; epidemics

## 2020 MATHEMATICS

**SUBJECT CLASSIFICATIONS**  
62G07; 62G20

## 1. Introduction

A dominating question in the public evaluation of the COVID-19 situation during the 2020 pandemic is the estimation of the basic reproduction number  $R_0$ , the number of new infections which are caused (on average) from a single infected individual. While this number is prominently discussed in the news about COVID-19, it is an important variable for disease transmission in general (Wallinga and Lipsitch 2007; Leavitt et al. 2020). In order to estimate the reproduction number of a disease the so-called generation time  $G$  plays an important role (see Euler–Lotka-Equation e.g. in Britton and Scalia Tomba 2019). The generation time  $G$  is defined as the time difference between the infection time of a randomly chosen infected individual and his or her infector. Let  $\varphi_G$  denote the density of the generation time  $G$  and let  $i(t)$  be the expected incidence at time  $t$  assuming a certain model of transmission (*‘the average community rate of new infections’*, Britton and Scalia Tomba 2019, p. 2). By formulating a renewal equation for  $i(t)$  and assuming an exponential growth  $i(t) = Ce^{rt}$ , one obtains  $R_0^{-1} = F_{R_0}(G)$  (the Euler–Lotka-Equation, see e.g.

**CONTACT** Alexander Kreiss  a.kreiss@lse.ac.uk  The London School of Economics and Political Science, Department of Statistics, Houghton Street, London, WC2A 2AE, UK

Wallinga and Lipsitch 2007; Britton and Scalia Tomba 2019), where

$$F_{R_0}(G) := \mathbb{E} \left( e^{-rG} \right) = \int_0^\infty e^{-rt} \varphi_G(t) dt. \quad (1)$$

Hence, the basic reproduction number  $R_0$  is a function of  $\varphi_G$ .

We suppose to observe transmission pairs, i.e. two infected people  $A$  and  $B$  where it is known that  $A$ , the *index case*, infected  $B$ , the *secondary case*. We observe for  $A$  and  $B$  their times of symptom onset and we observe, in addition, that the infection of the index case happened in a certain interval. We do not assume information about the infection of the secondary case through the index case. Such data can arise e.g. if the index case got infected during a travel to a region where the virus is circulating and infected the secondary case back home where the virus is not spreading yet (see Bi et al. 2020; Lauer et al. 2020; Xia et al. 2020 for examples related to travel to and from Wuhan in the early days of the pandemic). This leads to observations of the infection times with measurement error. Hence, the generation time and the incubation times  $I_1, I_2$  (the time from infection to symptom onset) are not directly observed. The serial interval  $S$  (the time between the two symptom onsets), in contrast, can be observed. Coming back to the example of estimation of  $F_{R_0}(G)$  (and hence  $R_0$ ), the most natural estimator, i.e. replacing the expectation by the sample average, is not feasible from our observations because the exact infection times are not observed. In reality it is rarely possible to make observations of  $G$  directly for the same reason (cf. Nishiura, Kakehashi, and Inaba 2009). However, the urgency of the situation requires estimation with imperfect data (Ferretti et al. 2020b).

We can write down a likelihood for observing the symptom onset times and the exposure interval. This likelihood can be written in terms of  $\varphi_G$  and other quantities. If it was possible to formulate a family of generation time densities which is indexed by  $R_0$  we could simply maximise the likelihood over  $R_0$ . In view of (1) this seems difficult. We follow therefore a different idea: Find a nonparametric estimate  $\hat{\varphi}_{G,n}$  of  $\varphi_G$  and study  $F_{R_0}(\hat{\varphi}_{G,n})^{-1}$  as an estimator of  $R_0$ . Such *plug-in estimators* have been studied e.g. in Shen (1997) and it was argued there that smoothness of  $F_{R_0}$  can compensate for typical drawbacks of nonparametric estimators like a slow convergence rate.

This approach is not limited to  $R_0$  but can be applied to other interesting functions  $F$  of  $\varphi_G$ , we call them *features of  $G$* . One example are tests: Some methods for estimation of  $R_0$  make implicit assumptions about the generation time (cf. Wallinga and Lipsitch 2007). Testing for such assumptions can be realised within this framework if the test statistic  $T(\hat{\varphi}_{G,n})$  can be written as a function of an estimate  $\hat{\varphi}_{G,n}$  of  $\varphi_G$ , e.g. similarly to Härdle and Mammen (1993) who use an  $L^2$  type test-statistic. Further interesting features of  $\varphi_G$  are for instance variances, mean values, quantiles or the probability of pre-symptomatic infection  $\mathbb{P}(G \leq I)$  where  $I$  denotes the incubation time (the time from infection to symptom onset). This is a feature of the joint variable  $(G, I)$ . It is therefore interesting to study efficient estimation of general features of  $G$  or of  $(G, I)$ . We will focus in the following on features of  $G$ . More precisely, we will estimate  $F(G)$  by firstly finding a semi-parametric sieve-estimator  $\hat{\varphi}_{G,n}$  of  $\varphi_G$  based on Laguerre polynomials and consider then the *plug-in* estimator which estimates  $F(G)$  through replacing  $\varphi_G$  by  $\hat{\varphi}_{G,n}$  in the definition of  $F$ .

In the beginning of a pandemic it is important to estimate its transmission characteristics (Bi et al. 2020) and it is common to replace the unobserved generation time  $G$  by

the serial interval  $S$  (cf. Wallinga and Lipsitch 2007; Britton and Scalia Tomba 2019; Ferretti et al. 2020b), because this can be observed in clinical studies. In certain situations this can be a promising approach, cf. Remark 2.2. However, this practice can also yield biased estimates (cf. Britton and Scalia Tomba 2019). In this paper, we avoid this issue by estimating  $\varphi_G$  directly. Moreover, since the interest lies in a function of  $\varphi_G$  rather than  $\varphi_G$  itself, it might be desirable for researchers to not impose parametric assumptions on  $\varphi_G$  but to consider a nonparametric approach instead. In addition, a nonparametric approach avoids issues like selecting a parametric family or important aspects being hidden through this choice (cf. Groeneboom 2021).

The main contribution of this paper is to provide the first step towards efficient, semi-parametric estimation of features of the generation time: Consistent density estimation. Based on such an estimator the theory for efficient, semi-parametric inference methods about features of  $\varphi_G$  can be developed. To this end we extend the methodology of Ferretti et al. (2020b) and Ganyani et al. (2020) in two ways: Firstly, we estimate the densities of the incubation period and the generation time jointly in one step with the same data and, secondly, we provide a semi-parametric framework which makes no a-priori assumptions about the incubation and generation times. We pursue this by constructing a sieve estimator based on Laguerre polynomials highlighting the flexibility of Laguerre polynomials as approximating functions. In order to identify the model we need to specify a model for the relation between the exposure window and the infection time but we will not make parametric assumptions about the distributions of the incubation period and the generation time. Thus, we will do semi-parametric estimation.

General background information about data analysis in disease transmission can e.g. be found in Held, Hens, O'Neill, and Wallinga (2019) and Chowell, Hayman, Bettencourt, and Castolli-Chavez (2009). The estimation of  $R_0$  from observations of the serial interval in particular is discussed in Lipsitch et al. (2003). As an alternative (Ferretti et al. 2020b) suggest the following two step procedure: Firstly, fit a parametric model for the incubation time (cf. Bi et al. 2020; Lauer et al. 2020) and then, secondly, use this to fit a parametric model to the generation time. Groeneboom (2021) describes how to do nonparametric estimation of the incubation time. Problems related to under-reporting or delays (cf. Azmon, Faes, and Hens 2014) are not considered here because we have data in mind which was collected by researchers rather than observational data from self-reporting. In addition to this type of experimental data, it is also possible to collect larger sets with covariates and use prediction techniques to identify transmission pairs (cf. Leavitt et al. 2020). Specific parametric results about the above mentioned quantities for Covid-19 can e.g. be found in Bi et al. (2020), Ferretti et al. (2020b), Ganyani et al. (2020), Lauer et al. (2020), Tindale et al. (2020) and in many other places. The exact mathematical setting described below is close to de-convolution and measurement error settings, classical results about which can be found for example in Fan (1991), Devroye (1989), and Carroll, Ruppert, Stefanski, and Crainiceanu (2006). General background about sieve estimators and series estimation can be found in Newey (1997) and Chen (2007) and the references therein. Ai and Chen (2003) and Newey and Powell (2003) use sieve estimation based on moment conditions while we start from a likelihood. Further reading about plug-in sieve estimates can be found in Shen (1997) and Chen and Shen (1998).

The structure of this paper is as follows: In Section 2 we introduce the exact modelling framework. Afterwards, in Section 3, we provide a sieve estimator for the incubation

period and generation time based on Laguerre polynomials. Its properties will be discussed in Section 4 and we show that our estimator is asymptotically consistent. The asymptotic distribution of the methodology will be studied in Section 5 by means of simulations and it will be applied to a real-world dataset consisting of 191 SARS-COV-2 transmission pairs that has also been used by Hart, Maini, and Thompson (2021) and Ferretti et al. (2020a). The R-code which is used for these computations is available on github (<https://github.com/akreiss/SemiParametric-Laguerre.git>). Finally, we finish the paper with some concluding remarks in Section 6. Additional simulation results and some proofs are collected in the Appendix.

## 2. Model

We study observations of transmission pairs as shown in Figure 1. The first person, the index case, gets infected at time  $T_1$  which is unobserved. However, it is known to lie in the interval  $[0, W]$ , the *exposure period*. The exact conditional distribution of  $T_1$  in  $[0, W]$  is allowed to depend on a random variable  $C$  which determines for example the location of the infection. In certain situations, it will turn out that the likelihood will depend only on  $\min(W, S_1)$ , where  $S_1 \geq T_1$  denotes the time at which Person 1 shows symptoms. In that sense, observation of  $W$  is only then required if  $W \leq S_1$ . Otherwise we do not have to observe  $W$  but we have to observe that  $W \geq S_1$ . The *incubation period* is defined as  $I_1 := S_1 - T_1$ . Moreover, Person 1 is known to have infected Person 2, the secondary case. Similarly as above, we define  $S_2, T_2, I_2$  for the second person: Person 2 shows symptoms at the observed time  $S_2$ . But, of course, the infection time  $T_2$  and hence the incubation time  $I_2 := S_2 - T_2$  are unobserved. The generation time is thus  $G := T_2 - T_1$  and it is also unobserved, while the serial interval  $S := S_2 - S_1$  can be observed. Such data was for example collected by Ferretti et al. (2020b), Bi et al. (2020), and Lauer et al. (2020).

The following relations can be directly read off from the definitions and Figure 1.

$$S_1 = T_1 + I_1, \quad (2)$$

$$S_2 = T_1 + I_2 + G, \quad (3)$$

$$S = G + I_2 - I_1. \quad (4)$$

Let  $(W_i, T_{1,i}, C_i, S_{1,i}, S_{2,i}, G_i)_{i=1, \dots, n}$  be iid copies of  $(W, T_1, C, S_1, S_2, G)$ . However, we suppose that our observations are only given by  $(S_{1,i}, S_{2,i}, \tilde{W}_i, C_i)_{i=1, \dots, n}$ , where  $\tilde{W}_i :=$



**Figure 1.** Schematic depiction of the important quantities and their relation.

$\min(S_{1,i}, W_i)$ . Based on our observations, we can only make direct inference about the joint distribution of  $(S_1, S_2)$  and the serial interval  $S := S_2 - S_1$ . But we have only indirect information about  $I, G$  and  $T_1$  by means of (2)–(3) (note that (4) is the difference of (2) and (3)). Informally speaking, we have three unknown densities and two equations. In other words, identification of the joint-distribution of the three random variables  $(I, G, T_1)$  from a joint distribution of the two variables  $(S_1, S_2)$  is impossible without additional assumptions. Throughout we will assume the following to be true:

**Assumption (M) Model:**  $I_1, I_2$  and  $G$  are non-negative and  $C$  takes values in  $\{1, \dots, K_C\}$ . Their distributions have the following densities (with respect to the Lebesgue measure) or count measures and relate as follows

$$\begin{aligned} W &\sim \varphi_W \\ C &\sim p_C, \text{ independent of } W \\ T_1 &\sim \varphi_{T_1}(\cdot \mid W, C) \text{ cond. on } (W, C) \\ I_1, I_2 &\stackrel{iid}{\sim} \varphi_I, \text{ independent of } (T_1, W, C) \\ G &\sim \varphi_G, \text{ independent of } (I_1, I_2, T_1, W, C). \end{aligned} \tag{5}$$

It holds that  $\varphi_{T_1}(t \mid w, c) = \mathbb{1}(t \in [0, w])n(w, c)h(w - t \mid c)$  where  $h : [0, \infty) \rightarrow [0, \infty)$  is integrable. ■

When inspecting the likelihood of Ferretti et al. (2020b) and the assumptions of Ganyani et al. (2020), we see the same independence assumptions for  $I_1, I_2, G$  and  $T_1$ . Independence between  $I_1$  and  $T_1$  can also be found in Groeneboom (2021). The variable  $C$  can be understood as covariate and will be interpreted as the location of the infection. It can be used to capture heterogeneity in the infection dynamics: In a location with an increasing number of infections, it is more likely that the infection time  $T_1$  lies towards the end of the infection window  $[0, W]$  conditionally on  $(C, W)$ . In contrast, in locations with a low number of infections,  $T_1$  is possibly uniformly distributed in  $[0, W]$ . See Remark 2.1 for further details and a motivation for certain choices of  $h$ . It has been reported (see e.g. Tindale et al. 2020) that the transmission dynamics, i.e.  $\varphi_I$  and  $\varphi_G$ , can be different in different scenarios. Thus, we need to restrict to data coming from the same environment. Moreover, the above assumption means that we observe true transmission pairs (see e.g. Tindale et al. 2020 for a parametric model which allows for an unknown number of intermediate infectors).

**Remark 2.1:** All discussion below is conditional on  $W$ . The dataset we imagine is recorded in the beginning of a pandemic, e.g. a person comes home after travelling to a region where the virus is circulating. Suppose that the time points, at which the person of interest has an infectious contact (i.e. a contact which definitely leads to an infection), are the jump points of a counting process  $N$  with intensity function  $\lambda : [0, W] \rightarrow [0, \infty)$ . It is plausible that the intensity of infection is proportional to the number of infections in the population. Since (in most cases) the infection numbers behave exponentially, we also consider  $\lambda(r) = \lambda_0 \exp(-\alpha(W - r)) \sim \exp(\alpha r)$ , where  $\alpha \in \mathbb{R}$  describes the exponential growth (or decay) of the number of infections. In this notation,  $T_1$ , the time of infection, is the first jump of

$N$ . In an early stage of the pandemic it is plausible that there is only one infectious contact. Then, using that counting processes have independent increments and that the number of jumps in a certain interval is Poisson distributed with parameter equal to the integral of the intensity over the respective interval, we obtain

$$\begin{aligned} \mathbb{P}(T_1 \leq s \mid N(t) = 1) &= \frac{\mathbb{P}(N([0, s]) = 1, N((s, t]) = 0)}{\mathbb{P}(N([0, t]) = 1)} \\ &= \frac{\int_0^s \lambda(r) \, dr}{\int_0^t \lambda(r) \, dr} = \frac{e^{-\alpha(W-s)} - e^{-\alpha W}}{e^{-\alpha(W-t)} - e^{-\alpha W}}. \end{aligned}$$

Deriving the distribution function above yields that the density of  $T_1$  is proportional to  $\exp(-\alpha(W-s))$ . Thus, we will later consider primarily the situation in which  $h(u \mid c) = \exp(-r(C)u)$ , where  $r(C)$  is the (known) exponential growth of the infection numbers in location  $C$ . Note that taking  $\alpha \rightarrow 0$ , i.e. the case of a constant intensity function, yields  $\mathbb{P}(T_1 \leq s \mid N(t) = 1) \rightarrow s/t$  (the distribution function of a uniform distribution on  $[0, t]$ ).

The discussion in Remark 2.1 is then plausible, if it is plausible to assume that the person of interest has infectious contacts with a rate which is so low/the time  $W$  is so short that there is only one infectious contact in  $[0, W]$ . In particular in the beginning of a pandemic this seems reasonable. Note that for the subsequent infection (within the observed transmission pair) this assumption is no longer valid: It is, intuitively speaking, more likely that infections, e.g. in a household happen directly after the infected person has returned, rather than days after.

Note finally that in the special case  $r(c_0) = 0$  for some location  $c_0$ , i.e. if the infection level remains constant, we get  $h(\cdot \mid c_0) = 1$  (then  $n(w, c_0) = 1/w$ ), i.e.  $T_1 \sim \mathcal{U}([0, W])$  (uniform distribution). From a mathematical standpoint, this case leads to another very intuitive result: The conditional distribution of  $S_1$  given  $W$  and  $C = c_0$  is given by the convolution through (2)

$$\begin{aligned} \mathbb{P}(S_1 \leq x \mid W, C = c_0) &= \mathbb{P}(T_1 + I_1 \leq x \mid W, C = c_0) \\ &= \frac{1}{W} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}(t \in [0, W]) \mathbb{1}(t + s \leq x) \varphi_I(s) \, ds \, dt \\ &= \frac{1}{W} \int_0^W \int_{-\infty}^{x-t} \varphi_I(s) \, ds \, dt. \end{aligned}$$

The conditional density of  $S_1$  can then be obtained through differentiation of the above with respect to  $x$  and is given by

$$\frac{1}{W} \int_0^W \varphi_I(x-t) \, dt = \frac{1}{W} \int_{x-W}^x \varphi_I(u) \, du.$$

Hence, the likelihood of observing  $S_1 = x$  is given by the probability of observing an incubation period between  $x-W$  and  $x$ , the natural lower and upper bounds. This way we can identify  $\varphi_I$  from observations of  $S_1$ . Then, in a second step, we can use (4) to identify  $\varphi_G$  from observing  $S$ . Hence, under the assumption that  $T_1 \sim \mathcal{U}([0, W])$  conditionally on  $W$ , we can identify the distributions of  $I$  and  $G$  from the distribution of  $(S_1, S_2)$ . We make this mathematically precise also in the general case in Corollary 4.4.



We mentioned in the introduction that using the serial interval in place of the generation time can lead to biases. We discuss in the next remark in which situations this practice is safe to use and we argue that it does not work for  $R_0$ .

**Remark 2.2 (Serial Interval vs. Generation Time):** Consider the set-up from Assumption (M) and let  $\varphi_S$  denote the density of  $S$ . Our interest lies in estimation of  $F(G)$ , where  $F$  is an arbitrary feature of  $G$ . As a first example let  $F$  be linear, e.g.  $F(G) := \mathbb{E}(G) = \int_{\mathbb{R}} t\varphi_G(t) dt$  be the expectation. In that case by (4)

$$F(G) = F(S) + F(I_1) - F(I_2) = F(S) = \int_{-\infty}^{\infty} t\varphi_S(t) dt$$

and  $F(G)$  can probably be well (maybe even optimally) estimated by estimating the integral on the right based on observations of the serial interval alone. For the case of the basic reproduction number, however,  $F_{R_0}(G) = \mathbb{E}(\exp(-rG))$  is a moment generating function, i.e. it is non-linear. Other examples of interesting non-linear features of interest are quantiles, a quadratic test statistic or the probability of pre-symptomatic infection  $\mathbb{P}(G \leq I_1)$ . For these features, we cannot simply estimate the serial interval in place of the generation time, but it is necessary to have an estimate of the density  $\varphi_G$  available.

Remark 2.2 together with the unavailability of observations from  $G$  shows that estimation of features like  $R_0$  is non-trivial and requires more complicated methods. Before turning to the estimators we make a remark about asymptomatic patients.

**Remark 2.3 (Asymptomatic Patients):** It is unclear how asymptomatic patients can be handled in this framework. However, it should be emphasised that asymptomatic means here infected people who infect others but who never show symptoms. Pre-symptomatic infections are well allowed, i.e. Person 1 is allowed to infect Person 2 before showing symptoms. However, we assume here that Person 1 and Person 2 both will eventually show symptoms.

### 3. Methodology

We introduce now our estimator by specifying approximations to  $\varphi_I$  and  $\varphi_G$ . Consider the following class of densities on the non-negative real line for  $m \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$

$$\mathcal{G}_m := \{ \varphi_{\theta} : \theta \in \mathbb{R}^{m+1}, \|\theta\|_2 = 1 \},$$

where  $\|\cdot\|_2$  is the Euclidean norm and

$$\varphi_{\theta}(x) = \mathbb{1}(x \geq 0) e^{-x} \left( \sum_{k=0}^m \theta^{(k)} L_k(x) \right)^2 \tag{6}$$

with  $\theta = (\theta^{(0)}, \dots, \theta^{(m)})$  and

$$L_k(x) := \sum_{i=0}^k \binom{k}{i} \frac{(-x)^i}{i!}$$

being the  $k$ th Laguerre polynomial. Since the Laguerre polynomials form an orthonormal system of functions on  $[0, \infty)$  with respect to the weight function  $e^{-x}$  it is simple to



show that the above construction yields a density under the simple condition  $\|\theta\|_2 = 1$ , cf. Lemma 3.1. Moreover, it is well known that a large class of densities can be approximated by Laguerre polynomials. Define to this end the Hellinger distance of two distributions on  $\mathcal{X}$  which have densities  $\varphi_1, \varphi_2$  with respect to a measure  $\mu$  via

$$\rho_H(\varphi_1, \varphi_2) := \left( \int_{\mathcal{X}} \left( \varphi_1(x)^{\frac{1}{2}} - \varphi_2(x)^{\frac{1}{2}} \right)^2 d\mu(x) \right)^{\frac{1}{2}}.$$

The following lemma is essentially a rephrasing of Theorem 1 in Chapter II.8 of Nikiforov and Uvarov (1988) which we state here for the convenience of the reader.

**Lemma 3.1:** *For any  $m \in \mathbb{N}_0$  and any  $\theta \in \mathbb{R}^{m+1}$  with  $\|\theta\|_2 = 1$  we have that  $\varphi_\theta$  as defined in (6) is a density function on  $[0, \infty)$ . Suppose moreover that  $\varphi$  is an arbitrary density on  $[0, \infty)$  such that  $p(x) := \sqrt{e^x \varphi(x)}$  is continuous on  $[0, \infty)$  and has a piecewise continuous derivative  $p'(x)$ . Suppose that*

$$\int_0^\infty p'(x)^2 x e^{-x} dx < +\infty. \quad (7)$$

*Then, for any sequence  $m_n \rightarrow \infty$  there are  $\theta_n \in \mathbb{R}^{m_n+1}$  with  $\|\theta_n\| = 1$  such that for  $n \rightarrow \infty$*

$$\rho_H(\varphi_{\theta_n}, \varphi)^2 = \int_0^\infty (\varphi_{\theta_n}(x)^{1/2} - \varphi(x)^{1/2})^2 dx \rightarrow 0.$$

*Moreover,  $\varphi_{\theta_n} \rightarrow \varphi$  locally uniformly, i.e.  $\sup_{x \in K} |\varphi_{\theta_n}(x) - \varphi(x)| \rightarrow 0$  for any compact set  $K \subseteq (0, \infty)$ .*

**Proof:** Let  $p_{m_n}(x; \theta_n) := \sum_{k=0}^{m_n} \theta_n^{(k)} L_k(x)$ , where  $\theta_n^{(k)}$  denotes the  $k$ th entry of  $\theta_n$ . Then  $\varphi_{\theta_n}(x) = \mathbb{1}(x \geq 0) e^{-x} p_{m_n}(x; \theta_n)^2$ . The first part of the lemma is a simple calculation using the orthonormality of the Laguerre polynomials, i.e. use that

$$\int_0^\infty e^{-x} L_k(x) L_l(x) dx = \mathbb{1}(k = l).$$

For the second statement we note

$$\int_0^\infty e^{-x} p(x)^2 dx = \int_0^\infty \varphi(x) dx = 1 < \infty.$$

This and (7) are exactly the conditions of Theorem 1 in Chapter II.8 of Nikiforov and Uvarov (1988) which we stated in the Appendix as Theorem A.2 for the convenience of the reader. The theorem states that the sequence  $\tilde{p}_n(x) := \sum_{k=0}^{m_n} c_k L_k(x)$  with  $c_k =$

$\int_0^\infty p(x)L_k(x)e^{-x} dx$  converges locally uniformly, i.e. uniformly on compact sets, to  $p$ . Let

$$C_n := \int_0^\infty e^{-x}\tilde{p}_n(x)^2 dx = \sum_{k=0}^{m_n} c_k^2$$

and  $\theta_n^{(k)} := c_k/C_n^{1/2}$  for  $k = 0, \dots, m_n$ . Clearly,  $\|\theta_n\|_2 = 1$ . By Parseval's Identity  $C_n \rightarrow \sum_{k=0}^\infty c_k^2 = \int_0^\infty e^{-x}p(x)^2 dx = 1$ . Hence, we can conclude that also

$$p_{m_n}(x; \theta_n) = \sum_{k=0}^{m_n} \theta_n^{(k)} L_k(x) = C_n^{-\frac{1}{2}} \sum_{k=0}^{m_n} c_k L_k(x) = C_n^{-\frac{1}{2}} \tilde{p}_n(x)$$

converges locally uniformly to  $p$  because  $p$  is bounded on compact sets  $K \subseteq (0, \infty)$ . From this we conclude also the locally uniform convergence of  $\varphi_{\theta_n}$  to  $\varphi$ . Moreover, Theorem A.2 states that

$$\int_0^\infty e^{-x}(\tilde{p}_n(x) - p(x))^2 dx \rightarrow 0.$$

The proof of the Lemma is complete since the above implies

$$\begin{aligned} & \int_0^\infty (\varphi_{\theta_n}^{1/2} - \varphi^{1/2})^2 dx \\ &= \int_0^\infty e^{-x} (p(x) - p_{m_n}(x; \theta_n))^2 dx \\ &= \int_0^\infty e^{-x} \left( p(x) - \tilde{p}_n(x) + \tilde{p}_n(x) \left( 1 - C_n^{-\frac{1}{2}} \right) \right)^2 dx \\ &\leq 2 \int_0^\infty e^{-x} (p(x) - \tilde{p}_n(x))^2 dx + 2 \int_0^\infty e^{-x} \tilde{p}_n(x)^2 dx \left( 1 - C_n^{-\frac{1}{2}} \right)^2 \rightarrow 0. \quad \blacksquare \end{aligned}$$

The conditions of Lemma 3.1 cover a wide class of piecewise continuously differentiable densities like any sub-Gaussian density, densities with compact support and sub-exponential densities. We consider condition (7) therefore as not-restrictive. Before we can write down the likelihood, we have to find the conditional density of  $(S_1, S_2, W)$  given  $C$ .

**Lemma 3.2:** *Suppose that the random variables  $W, C, T_1, I_1, I_2, G$  are related as in Assumption (M). The conditional joint density of  $(S_1, S_2, W)$  given  $C$ , where  $S_1 := I_1 + T_1$  and  $S_2 := T_1 + I_2 + G$ , is given by*

$$\begin{aligned} & f(x_1, x_2, \omega | c) \\ &= n(\omega, c) \varphi_W(\omega) \int_0^{x_2} \varphi_G(y) \int_0^{\min(x_1, \omega)} h(\omega - t | c) \varphi_I(x_1 - t) \varphi_I(x_2 - t - y) dt dy. \end{aligned}$$

for  $x_1, x_2, \omega \geq 0$  and 0 otherwise.

**Proof:** Let  $x_1, x_2, \omega \geq 0$  be arbitrary. Note firstly that, by Assumption (M), the conditional density of  $(W, T_1)$  given  $C$  is given by

$$\varphi_{T_1}(t | w, c)\varphi_W(w) = \mathbb{1}(t \in [0, \omega])n(w, c)h(w - t | c)\varphi_W(w).$$

Below integrals of the type  $\int_0^a$  are to be understood over the set  $(\min(0, a), \max(0, a))$ . Since all integrands are supported in  $[0, \infty)$ , integrals are zero when  $a < 0$ . We have by the independence assumptions in (M)

$$\begin{aligned} & \mathbb{P}(S_1 \leq x_1, S_2 \leq x_2, W \leq \omega | C) \\ &= \mathbb{P}(T_1 + I_1 \leq x_1, T_1 + G + I_2 \leq x_2, W \leq \omega | C) \\ &= \mathbb{E}(\mathbb{1}(W \leq \omega)\mathbb{P}(T_1 + I_1 \leq x_1, T_1 + G + I_2 \leq x_2 | T_1, W, C) | C) \\ &= \mathbb{E}(\mathbb{1}(W \leq \omega)\mathbb{P}(I_1 \leq x_1 - T_1 | T_1, C) \cdot \mathbb{P}(G + I_2 \leq x_2 - T_1 | T_1, C) | C) \\ &= \mathbb{E}\left(\mathbb{1}(W \leq \omega) \int_0^{x_1 - T_1} \varphi_I(a) da \cdot \int_0^{x_2 - T_1} \int_0^\infty \varphi_G(y)\varphi_I(b - y) dy db \middle| C\right) \\ &= \int_0^\omega \int_0^v n(v, c)h(v - t | c)\varphi_W(v) \\ &\quad \times \int_0^{x_1 - t} \varphi_I(a) da \cdot \int_0^{x_2 - t} \int_0^\infty \varphi_G(y)\varphi_I(b - y) dy db dt dv. \end{aligned}$$

We obtain furthermore by differentiating under the integral and using that  $\varphi_I$  and  $\varphi_G$  are supported on the non-negative real line

$$\begin{aligned} & f(x_1, x_2, \omega | c) \\ &= \frac{d^3}{dx_1 dx_2 d\omega} \mathbb{P}(S_1 \leq x_1, S_2 \leq x_2, W \leq \omega | C = c) \\ &= \frac{d^2}{dx_1 dx_2} n(\omega, c)\varphi_W(\omega) \int_0^\omega h(\omega - t | c) \\ &\quad \times \int_0^{x_1 - t} \varphi_I(a) da \cdot \int_0^{x_2 - t} \int_0^\infty \varphi_G(y)\varphi_I(b - y) dy db dt \\ &= n(\omega, c)\varphi_W(\omega) \int_0^\omega h(\omega - t | c)\varphi_I(x_1 - t) \cdot \int_0^\infty \varphi_G(y)\varphi_I(x_2 - t - y) dy dt \\ &= n(\omega, c)\varphi_W(\omega) \int_0^\infty \varphi_G(y) \int_0^\omega h(\omega - t | c)\varphi_I(x_1 - t)\varphi_I(x_2 - t - y) dt dy \\ &= n(\omega, c)\varphi_W(\omega) \int_0^{x_2} \varphi_G(y) \int_0^{\min(x_1, \omega)} h(\omega - t | c)\varphi_I(x_1 - t)\varphi_I(x_2 - t - y) dt dy. \end{aligned}$$

In the last line above we can replace the upper integration bound of the outer integral by  $x_2$  because for  $y > x_2$ , we have  $x_2 - t - y < 0$  and hence  $\varphi_I(x_2 - t - y) = 0$ .  $\blacksquare$

By using the density from above, we obtain an expression for the likelihood for estimation of  $\varphi_I$  and  $\varphi_G$ .

**Lemma 3.3:** Let the observations  $(S_{1,i}, S_{2,i}, W_i, C_i)_{i=1, \dots, n}$  be iid copies of  $(S_1, S_2, W, C)$  which are related as in Assumption (M). Let  $m_1, m_2 \in \mathbb{N}$  and  $\theta_1 \in \mathbb{R}^{m_1+1}$ ,  $\theta_2 \in \mathbb{R}^{m_2+1}$ . Consider the candidate densities  $\varphi_{I, \theta_1} \in \mathcal{G}_{m_1}$  and  $\varphi_{G, \theta_2} \in \mathcal{G}_{m_2}$ . The likelihood for this configuration is given by, denote  $\tilde{W}_i := \min(W_i, S_{1,i})$

$$\begin{aligned} & \mathcal{L}_n(\varphi_{I, \theta_1}, \varphi_{G, \theta_2}) \\ &= \sum_{i=1}^n \log \left( \int_0^{S_{2,i}} \varphi_{G, \theta_2}(y) \int_0^{\tilde{W}_i} h(W_i - t \mid C_i) \varphi_{I, \theta_1}(S_{1,i} - t) \varphi_{I, \theta_1}(S_{2,i} - t - y) dt dy \right) \\ & \quad + C_n, \end{aligned} \quad (8)$$

where  $C_n$  is a random constant which does not depend on  $\varphi_{I, \theta_1}$  and  $\varphi_{G, \theta_2}$ . In the specific case of  $h(u \mid c) = \exp(-r(c)u)$  for some exponential parameter  $r(c) \in \mathbb{R}$  we even get that

$$\begin{aligned} \mathcal{L}_n &= \sum_{i=1}^n \log \left( \int_0^{S_{2,i}} \varphi_{G, \theta_2}(y) \right. \\ & \quad \left. \times \int_0^{\tilde{W}_i} \exp(r(C_i)t) \varphi_{I, \theta_1}(S_{1,i} - t) \varphi_{I, \theta_1}(S_{2,i} - t - y) dt dy \right) + \tilde{C}_n \end{aligned} \quad (9)$$

depends on  $W_i$  only through  $\tilde{W}_i$ . Above  $\tilde{C}_n$  is a random constant which does not depend on  $\varphi_{I, \theta_1}$  and  $\varphi_{G, \theta_2}$ .

**Proof:** Let  $f_{\varphi_{I, \theta_1}, \varphi_{G, \theta_2}}(x_1, x_2, \omega \mid c)$  denote the conditional density of  $(S_1, S_2, W)$  given  $C$  as defined in Lemma 3.2 when  $I$  has density  $\varphi_{I, \theta_1}$  and  $G$  has density  $\varphi_{G, \theta_2}$ . By independence, we have

$$\begin{aligned} \mathcal{L}_n &= \sum_{i=1}^n \log \left( f_{\varphi_{I, \theta_1}, \varphi_{G, \theta_2}}(S_{1,i}, S_{2,i}, W_i \mid C_i) p_C(C_i) \right) \\ &= \sum_{i=1}^n \log \left( \int_0^{S_{2,i}} \varphi_{G, \theta_2}(y) \right. \\ & \quad \left. \times \int_0^{\min(S_{1,i}, W_i)} h(W_i - t \mid C_i) \varphi_{I, \theta_1}(S_{1,i} - t) \varphi_{I, \theta_1}(S_{2,i} - t - y) dt dy \right) \\ & \quad + \sum_{i=1}^n \log \left( n(W_i, C_i) \varphi_W(W_i) p_C(C_i) \right), \end{aligned}$$

where the latter term is independent of  $\varphi_{I, \theta_1}$  and  $\varphi_{G, \theta_2}$ . This finishes the proof of (8). For the specific choice  $h(u \mid c) = \exp(-r(c)u)$  we get

$$\begin{aligned} \mathcal{L}_n &= \sum_{i=1}^n \log \left( \int_0^{S_{2,i}} \varphi_{G, \theta_2}(y) \right. \\ & \quad \left. \times \int_0^{\min(S_{1,i}, W_i)} \exp(r(C_i)t) \varphi_{I, \theta_1}(S_{1,i} - t) \varphi_{I, \theta_1}(S_{2,i} - t - y) dt dy \right) \end{aligned}$$

$$+ \sum_{i=1}^n (\log (n(W_i, C_i) \varphi_W(W_i) p_C(C_i)) - r(C_i) W_i)$$

which finishes the proof of (9) since the second line does not depend on  $\varphi_{I,\theta_1}$  or  $\varphi_{G,\theta_2}$ .  $\blacksquare$

Note that Lemma 3.3 implies that for given  $\varphi_W$  and  $p_C$ , the likelihood to be optimised is, conveniently, independent of  $\varphi_W$  and  $p_C$ . Thus, we can treat  $\varphi_W$  and  $p_C$  as known without loosing any practicality and, therefore, we may define the following estimators for  $\varphi_I$  and  $\varphi_G$ : For given sequences  $(m_{1,n})_{n \in \mathbb{N}}, (m_{2,n})_{n \in \mathbb{N}} \subseteq \mathbb{N}_0$ , we study

$$\begin{aligned} (\hat{\varphi}_{I,n}, \hat{\varphi}_{G,n}) := & \operatorname{argmax}_{\varphi_{I,\theta_1} \in \mathcal{G}_{m_{1,n}}, \varphi_{G,\theta_2} \in \mathcal{G}_{m_{2,n}}} \sum_{i=1}^n \log \left( \int_0^{S_{2,i}} \varphi_{G,\theta_2}(y) \int_0^{\tilde{W}_i} h(W_i - t | C_i) \right. \\ & \left. \times \varphi_{I,\theta_1}(S_{1,i} - t) \varphi_{I,\theta_1}(S_{2,i} - t - y) dt dy \right). \end{aligned}$$

At this point it is not clear that we can identify the parameters  $\theta_1, \theta_2$ . Later, in Corollary 4.4, we will see that consistent estimation of the distribution functions of  $\varphi_I$  and  $\varphi_G$  is possible if a mild assumption on the characteristic functions holds. Under the same assumption, a similar proof-technique can be applied to show that two different sets of parameters lead to different likelihoods.

#### 4. Theory

In this section we will prove and discuss a consistency result. For  $\varphi_I$  and  $\varphi_G$  being arbitrary densities of  $I_1, I_2$  and  $G$ , respectively, we denote by  $f_{\varphi_I, \varphi_G}$  the conditional density of  $(S_1, S_2, W)$  given  $C$  as defined in Lemma 3.2. Then,  $f_{\varphi_I, \varphi_G} p_C$  denotes the joint density of  $(S_1, S_2, W, C)$ . When  $\varphi_I$  and  $\varphi_G$  are chosen from the approximating spaces  $\mathcal{G}_{m_1}$  and  $\mathcal{G}_{m_2}$ , the following set denotes the set of all possible approximations to the joint density of  $(S_1, S_2, W, C)$

$$\mathcal{F}_{m_1, m_2} := \{f_{\varphi_I, \varphi_G} p_C : \varphi_I \in \mathcal{G}_{m_1}, \varphi_G \in \mathcal{G}_{m_2}\}.$$

In the following we will always assume that we are in the setting presented in Section 2.

**Theorem 4.1:** *Suppose (M) holds true and let  $m_{1,n}, m_{2,n} = n^\beta$  and  $\varepsilon_n = n^{-\gamma}$  for some  $\beta, \gamma > 0$  such that  $\gamma < 1/2$  and  $\beta < 1 - 2\gamma$ . Suppose that the true densities  $\varphi_I$  and  $\varphi_G$  fulfil the assumptions of Lemma 3.1 and let  $\varphi_{I,n}$  and  $\varphi_{G,n}$  be the sequences from Lemma 3.1 such that  $\rho_H(\varphi_I, \varphi_{I,n}) \rightarrow 0$  and  $\rho_H(\varphi_G, \varphi_{G,n}) \rightarrow 0$ . Suppose that there is  $\alpha \in (0, 1/2)$  such that for some  $N_0 \in \mathbb{N}_0$*

$$\sup_{n \geq N_0} C_\alpha(f_{\varphi_I, \varphi_G}, f_{\varphi_{I,n}, \varphi_{G,n}}) < \infty, \tag{10}$$

where

$$C_\alpha(f_1, f_2) := \left( \sum_{c=1}^{K_C} \int_{[0, \infty)^3} \left( \frac{f_1(x_1, x_2, \omega | c)}{f_2(x_1, x_2, \omega | c)} \right)^{\frac{\alpha}{1-\alpha}} f_1(x_1, x_2, \omega | c) p_C(c) d(x_1, x_2, \omega) \right)^{1-\alpha}. \tag{11}$$

Then, we have that

$$\rho_H(f_{\varphi_I, \varphi_G} P_C, f_{\hat{\varphi}_{I,n}, \hat{\varphi}_{G,n}} P_C) = O_P \left( \max \left( \varepsilon_n, \max \left( \rho_H(\varphi_I, \varphi_{I,n}), \rho_H(\varphi_G, \varphi_{G,n}) \right)^\alpha \right) \right) = o_P(1).$$

The proof of this result will be given later in Section 4.2. We will begin with a discussion of the result and its assumptions in the next subsection.

#### 4.1. Discussion of Theorem 4.1

Before turning to the assumptions of Theorem 4.1 we make a remark about the convergence rate.

**Remark 4.2:** Note that the requirements on  $\beta$  and  $\gamma$  in Theorem 4.1 allow in the limit for the choice  $\beta = 1/5$  and  $\gamma = 2/5 < 1/2$ . This yields the *classical* convergence rate  $n^{-2/5}$  for series estimation with polynomials if the true density is sufficiently smooth (cf. Newey 1997 and Proposition 3.6 in Chen 2007).

We continue with a discussion of the condition (10) above. This can be understood as a tail condition: For any compact set  $K \subseteq (0, \infty)^3$  we have by Lemma 3.1 that  $f_{\varphi_{I,n}, \varphi_{G,n}}$  converges uniformly to  $f_{\varphi_I, \varphi_G}$ . Thus, if we restrict the integral in the definition of  $C_\alpha(f_{\varphi_I, \varphi_G}, f_{\varphi_{I,n}, \varphi_{G,n}})$  to  $K$ , the restricted integral remains bounded. Note also that by choosing  $\alpha > 0$  small, possible singularities of  $f_2$  are integrable. Hence, (10) is only restrictive for the integral over  $K^c$ . Thus, if  $f_{\varphi_I, \varphi_G}$  is compactly supported (10) follows. Note furthermore that  $f_{\varphi_I, \varphi_G}$  is compactly supported if  $\varphi_I, \varphi_G$  and  $\varphi_W$  are compactly supported. In our specific epidemics setting for COVID-19, this is a highly plausible assumption because people stop being infectious at some point. However, for other diseases people can stay infectious for an extended period of time, e.g. for Hepatitis C, the mean generation time is about 20 years (cf. Wallinga and Lipsitch 2007).

In those cases of non-compactly supported distributions, (10) is a restriction: It is required that the approximations do not decrease much faster to zero than the actual density. The meaning of *much faster* is to be understood relative to  $f_{\varphi_I, \varphi_G}$  and it can be adjusted by choosing  $\alpha > 0$  small. Any polynomial difference can therefore be handled.

Theorem 4.1 shows, strictly speaking, that we can consistently estimate the joint distribution of  $(S_1, S_2, W, C)$ . However, from a practical point of view, the estimation of  $\varphi_I$  and  $\varphi_G$  is of interest. In order to ensure identifiability of  $\varphi_I$  and  $\varphi_G$ , we need the following additional assumption:

**Assumption 1 (Characteristic Functions):** Let  $\Phi_I$  denote the characteristic function of  $I$ . There are functions  $z_1 : \mathbb{R} \rightarrow \mathbb{R}$  and  $z_2 : \mathbb{R} \rightarrow \mathbb{R}$  such that for almost all  $x \in \mathbb{R}$

$$\Phi_I(x) \neq 0 \quad \text{and} \quad \mathbb{E} \left( e^{iz_1(x)W + iz_2(x)C} \mathbb{E} \left( e^{ixT_1} \mid W, C \right) \right) \neq 0.$$

In particular the second requirement above looks cryptic in full generality. In the following remark we show that in specific scenarios it simplifies to simpler conditions which are more interpretable.

**Remark 4.3:** The condition

$$\mathbb{E} \left( e^{iz_1(x)W + iz_2(x)C} \mathbb{E} \left( e^{ixT_1} \mid W, C \right) \right) \neq 0 \quad (12)$$

above can be rewritten in more specific scenarios.

- (1) If  $\Phi_{T_1}(x) := \mathbb{E}(\exp(ixT_1)) \neq 0$  almost everywhere, (12) holds because we may simply choose  $z_1 \equiv z_2 \equiv 0$ .
- (2) If  $C \equiv 1$  and  $T_1$  is uniformly distributed in  $[0, W]$ , i.e.  $h(w-t \mid c) \equiv 1$  and  $n(w, c) = 1/w$ , we obtain with  $z_2 \equiv 0$  and  $z_1(x) = -x/2$  that (note that  $\sin(x)/x \rightarrow 1$  for  $x \rightarrow 0$  such that the below inequality chain can be used also for  $x = 0$ )

$$\begin{aligned} & \mathbb{E} \left( e^{iz_1(x)W + iz_2(x)C} \mathbb{E} \left( e^{ixT_1} \mid W, C \right) \right) \\ &= \mathbb{E} \left( e^{-i\frac{x}{2}W} \frac{1}{W} \int_0^W e^{ixt} dt \right) \\ &= \mathbb{E} \left( \frac{1}{ixW} \left( e^{i\frac{x}{2}W} - e^{-i\frac{x}{2}W} \right) \right) = \mathbb{E} \left( \frac{1}{\frac{x}{2}W} \sin \left( \frac{x}{2}W \right) \right) = \mathbb{E} \left( \int_0^1 \cos \left( \frac{x}{2}Wt \right) dt \right) \\ &= \int_0^1 \operatorname{Re} \left( \mathbb{E} \left( e^{i\frac{x}{2}Wt} \right) \right) dt = \int_0^1 \operatorname{Re} \left( \Phi_W \left( \frac{x}{2}t \right) \right) dt, \end{aligned}$$

where  $\Phi_W$  is the characteristic function of  $W$ . The above is non-zero for almost all  $x \in \mathbb{R}$  (implying (12)) for many standard distributions like the exponential distribution which characteristic function has a strictly positive real part.

Our setting is very similar to the de-convolution set-up: In (2) the signal  $I_1$  is perturbed by the noise  $T_1$  (the distribution of which is determined by  $(W, C)$ ) and in (3) the signal  $G$  is perturbed by the noise  $I_2 + T_1$ . Thus it is clear that assumptions on  $W, C$  and  $I_1$  similar to those for the de-convolution set-up are necessary for identification. The assumption of almost everywhere non-vanishing characteristic functions has already been mentioned in Devroye (1989) (see the Remark below Theorem 1 therein) to be necessary to guarantee consistent estimation.

Let  $F_I$  and  $F_G$  denote the distribution functions corresponding to the densities  $\varphi_I$  and  $\varphi_G$ , respectively. Similarly, define  $\hat{F}_{I,n}$  and  $\hat{F}_{G,n}$  to be the distribution functions of  $\hat{\varphi}_{I,n}$  and  $\hat{\varphi}_{G,n}$ , respectively. The following corollary ensures under Assumption (C) that  $\hat{F}_{I,n}$  and  $\hat{F}_{G,n}$  are consistent estimators for  $F_I$  and  $F_G$ , respectively.

**Corollary 4.4:** *Suppose that, in addition to the assumptions of Theorem 4.1, Assumption (C) holds true. We then have*

$$\left\| \hat{F}_{I,n} - F_I \right\|_{\infty} = o_P(1) \quad \text{and} \quad \left\| \hat{F}_{G,n} - F_G \right\|_{\infty} = o_P(1).$$

The detailed proof is given in Section A.3 in the Appendix.

We finally come back to estimating the basic reproduction number  $R_0$ . The above framework provides us with a methodology to consistently estimate the reproduction number without making parametric statements about the incubation period or the generation time.



From (1) it is evident that we require next to an estimator for  $\varphi_G$  also an estimate  $\hat{r}_n$  of the growth rate of the expected incidence. In Ferretti et al. (2020b) the exponential growth rate of the reported numbers is estimated to be  $r = 0.14$ . But, strictly speaking, what we need here is the growth rate of the infection numbers which is (intuitively speaking) very similar but can be different due to under-reporting and delays. The estimator is specified in the following corollary, the proof of which can also be found in Section A.3 in the Appendix.

**Corollary 4.5:** *Let  $\hat{r}_n$  denote a consistent estimate of the exponential growth rate of the expected incidence  $r > 0$ , i.e. let  $\hat{r}_n - r = o_p(1)$ . Under the conditions of Theorem 4.1 and Assumption (C), we have*

$$\int_0^\infty e^{-\hat{r}_n t} \hat{\varphi}_{G,n}(t) dt \xrightarrow{\mathbb{P}} F_{R_0}(G).$$

**4.2. Proof of Theorem 4.1**

In order to prove Theorem 4.1 we need to introduce the same distance relation which was used also by Wong and Shen (1995). Define for any two densities  $f_1, f_2$  with respect to a measure  $\mu$  on a space  $\mathcal{X}$  and  $\alpha \neq 0$  the distance relation

$$\rho_\alpha(f_1, f_2) := \frac{1}{\alpha} \int_{\mathcal{X}} f_1(x) \left( \left( \frac{f_1(x)}{f_2(x)} \right)^\alpha - 1 \right) d\mu(x).$$

Note that

$$\begin{aligned} \rho_{-1/2}(f_1, f_2) &= -2 \int_{\mathcal{X}} \left( \sqrt{f_1(x)f_2(x)} - f_1(x) \right) d\mu(x) = 2 - 2 \int_{\mathcal{X}} \sqrt{f_1(x)f_2(x)} d\mu(x) \\ &= \int_{\mathcal{X}} \left( f_1(x) + f_2(x) - 2\sqrt{f_1(x)f_2(x)} \right) d\mu(x) \\ &= \int_{\mathcal{X}} \left( \sqrt{f_1(x)} - \sqrt{f_2(x)} \right)^2 d\mu(x) \\ &= \rho_H(f_1, f_2)^2. \end{aligned}$$

A second property of  $\rho_\alpha$  we mention here, is non-negativity for  $\alpha \geq -1$ : Let  $g_\alpha : (0, \infty) \rightarrow \mathbb{R}$  be given by  $g_\alpha(x) := \frac{1}{\alpha}(x^{-\alpha} - 1)$ . Then  $g''_\alpha(x) = (\alpha + 1)x^{-\alpha-2} \geq 0$  because  $\alpha \geq -1$ . Hence,  $g_\alpha$  is a convex function. We hence obtain by Jensen's Inequality (for the measure with density  $f_1$  with respect to  $\mu$ )

$$\rho_\alpha(f_1, f_2) = \int_{\mathcal{X}} g_\alpha \left( \frac{f_2(x)}{f_1(x)} \right) f_1(x) d\mu(x) \geq g_\alpha \left( \int_{\mathcal{X}} \frac{f_2(x)}{f_1(x)} f_1(x) \mu(x) \right) = 0.$$

Lemma 3.1 shows that the sieve spaces  $\mathcal{G}_m$  lie dense in a large class of densities with respect to the Hellinger distance. However, for our consistency result, we require that the sieve spaces provide also good approximations with respect to  $\rho_\alpha$  for  $\alpha > 0$ . The following lemma provides the main tool.

**Lemma 4.6:** *Let  $\varphi_{I,1}, \varphi_{I,2}$  be two densities of incubation periods and let  $\varphi_{G,1}, \varphi_{G,2}$  be two densities of generation times and consider  $f_1(x_1, x_2, \omega | c)$  and  $f_2(x_1, x_2, \omega | c)$  defined as in*

*Lemma 3.2* using  $(\varphi_{I,1}, \varphi_{G,1})$  and  $(\varphi_{I,2}, \varphi_{G,2})$ , respectively. Define  $C_\alpha(f_1, f_2)$  as in (11). Then, for any  $\alpha \in (0, 1/2)$ ,

$$\rho_\alpha(f_1 p_C, f_2 p_C) \leq \frac{1}{\alpha} C_\alpha(f_1, f_2) \left( 2\rho_H(\varphi_{G,1}, \varphi_{G,2})^2 + 8\rho_H(\varphi_{I,1}, \varphi_{I,2})^2 \right)^\alpha. \quad (13)$$

The proof of the above Lemma is presented in Section A.4 in the Appendix. Before we can prove the consistency result, we need as a last preparation a bound on the bracketing entropy of the spaces  $\mathcal{F}_{m_1, m_2}$ .

**Definition 4.7:** Let  $\varepsilon > 0$ ,  $\mathcal{F}_0 \subseteq \mathcal{F}$  classes of functions and  $\rho : \mathcal{F} \times \mathcal{F} \rightarrow [0, \infty)$  a metric on  $\mathcal{F}$  be given. The *bracketing number*  $\mathcal{N}_{[]}(\varepsilon, \mathcal{F}_0, \rho)$  is defined as the smallest number of pairs  $(l_i, u_i) \in \mathcal{F}^2$  of functions such that  $\rho(l_i, u_i) \leq \varepsilon$  for all pairs and such that for any  $f \in \mathcal{F}_0$  there is a pair  $(l_i, u_i)$  with  $l_i(x) \leq f(x) \leq u_i(x)$ . The pairs  $(l_i, u_i)$  are called *brackets*.

**Lemma 4.8:** For  $m_1, m_2 \geq 1$  and any  $\varepsilon > 0$  with  $\varepsilon \leq (15m_1 m_2^2/4)^{1/2}$  we have

$$\log \mathcal{N}_{[]}(\varepsilon, \mathcal{F}_{m_1, m_2}, \rho_H) \leq (m_1 + m_2) \log \left( \frac{\pi \sqrt{15m_1 m_2^2}}{\varepsilon} \right)$$

This Lemma is also proven in Section A.4 in the Appendix. We have now all ingredients together to prove the main result of this paper.

**Proof of Theorem 4.1:** This theorem is a consequence of Theorem 4 from Wong and Shen (1995). For the convenience of the reader, we have stated the result in Section A.1 in the Appendix as Theorem A.1. Since we are interested in an asymptotic result we may assume below that  $n \geq N_0$ .

We apply Theorem A.1 with  $Y_i = (S_{1,i}, S_{2,i}, W_i, C_i)$  and  $\mathcal{F}_n = \mathcal{F}_{m_{1,n}, m_{2,n}}$ . We show next that the entropy condition is fulfilled: We note firstly that  $(15m_{1,n} m_{2,n}^2)^{1/2} \rightarrow \infty$  and  $\varepsilon_n \rightarrow 0$ . Hence, the condition  $\varepsilon_n \leq (15m_{1,n} m_{2,n}^2)^{1/2}$  is eventually fulfilled and we may apply Lemma 4.8. Thus, we obtain for a suitable  $c_2 < \infty$

$$\begin{aligned} & \frac{1}{\sqrt{n\varepsilon_n^2}} \int_{\frac{\varepsilon_n^2}{2^8}}^{\sqrt{2}\varepsilon_n} \sqrt{\log \mathcal{N}_{[]} \left( \frac{u}{c_1}, \mathcal{F}_{m_{1,n}, m_{2,n}}, \rho_H \right)} du \\ & \leq \frac{1}{\sqrt{n\varepsilon_n^2}} \int_{\frac{\varepsilon_n^2}{2^8}}^{\sqrt{2}\varepsilon_n} \sqrt{(m_{1,n} + m_{2,n}) \log \left( \frac{c_1 \pi \sqrt{15m_{1,n} m_{2,n}^2}}{u} \right)} du \\ & \leq \frac{\sqrt{2}\varepsilon_n}{\sqrt{n\varepsilon_n^2}} \sqrt{(m_{1,n} + m_{2,n}) \log \left( \frac{2^8 c_1 \pi \sqrt{15m_{1,n} m_{2,n}^2}}{\varepsilon_n^2} \right)} \\ & \leq 2n^{-\frac{1}{2} + \gamma + \frac{1}{2}\beta} \sqrt{\log \left( 2^8 c_1 \pi \sqrt{15n^{\frac{3}{2}\beta + 2\gamma}} \right)} \leq c_2, \end{aligned}$$

because  $-\frac{1}{2} + \gamma + \frac{1}{2}\beta > 0$ . Hence, the entropy condition of Theorem A.1 is fulfilled. Moreover, by the assumptions and Lemma 4.6 we find that

$$\delta_n(\alpha) \leq \rho_\alpha(f_{\varphi_I, \varphi_G} p_C, f_{\varphi_{I,n}, \varphi_{G,n}} p_C) \rightarrow 0.$$

Hence, we eventually have  $\delta_n(\alpha) < 1/\alpha$ . Under these conditions, we have that  $\varepsilon_n^*(\alpha) \rightarrow 0$  and  $n\varepsilon_n^*(\alpha)^2 \geq n\varepsilon_n^2 = n^{1-2\gamma} \rightarrow \infty$  since  $\gamma < 1/2$ . Hence, we conclude from Theorem A.1 that  $\rho_H(f_{\varphi_I, \varphi_G} p_C, f_{\hat{\varphi}_{I,n}, \hat{\varphi}_{G,n}} p_C) = O_P(\varepsilon_n^*(\alpha))$  and the proof of the theorem is complete since

$$\varepsilon_n^*(\alpha) = O(\max(\varepsilon_n, \max(\rho_H(\varphi_I, \varphi_{I,n}), \rho_H(\varphi_G, \varphi_{G,n})))^\alpha)$$

by Lemma 4.6. ■

### 5. Empirical studies

The analytic evaluation of the likelihood presented in Lemma 3.3 is very tedious. Therefore, the following results were obtained by numerically approximating the integrals. As illustrated by the simulations below this approximation does not cause problems in the estimation. However, we believe that a speed-up of the method is possible if the integrals are analytically computed. In order to efficiently enforce the constraint that  $\|\theta_1\|_2 = \|\theta_2\|_2 = 1$ , we optimise the angles of the polar-coordinates of  $\theta_1$  and  $\theta_2$  and fix their radii to 1. The angles can then be optimised under the box constraint  $[0, \pi]$  (note that  $\theta$  and  $-\theta$  yield the same model). Finally, we note that the likelihood can have singularities because the Laguerre densities (6) can be zero. We overcome the resulting problems in the optimisation, similar to Zhang and Davidian (2008), by repeating the optimisation of the likelihood with several randomly chosen starting values. Below we always considered 10 different, random starting values.

We begin with an illustration of our methodology with synthetic data in Section 5.1 followed by a brief analysis of a real-world dataset in Section 5.2. The R-code which is used for both parts is available on github (<https://github.com/akreiss/SemiParametric-Laguerre.git>).

#### 5.1. Simulation study

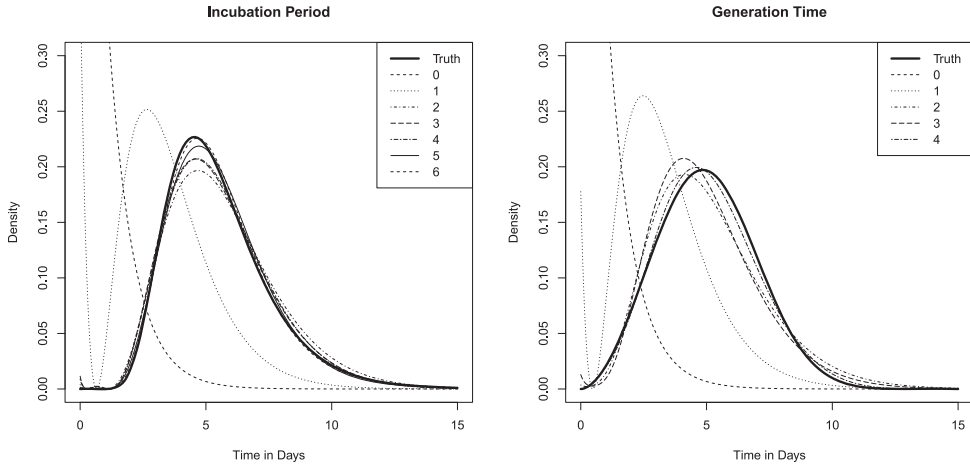
It appears to be very difficult to obtain data on transmission pairs, e.g. an early dataset of Ferretti et al. (2020b) contains only 40 transmission pairs (the data is available on the website of the journal on <https://doi.org/10.1126/science.abb6936> in the supplementary material section). To this end, we regard it useful to illustrate in the simulation experiments that our methodology works also for small datasets and we choose  $n = 40$  observations as well. In this simulation study we use as a data generating process the model as in (5) with the following choices:

$$\begin{array}{ll} \varphi_W & \text{exponential with rate } \lambda = 0.3820225 \\ p_C & \mathbb{P}(C = 0) = 0.65 \text{ and } \mathbb{P}(C = 1) = 0.35 \\ \varphi_{T_1} & \varphi_{T_1}(t \mid w, c) = \mathbb{1}(t \in [0, w]) \begin{cases} \frac{re^{-r(w-t)}}{1-e^{-rw}}, & c = 1 \\ \frac{1}{w}, & c = 0 \end{cases} \end{array}$$

$$r = \log(2)/5$$

$$\varphi_I \quad \varphi_{I,0} = \text{log-normal distribution: meanlog} = 1.644, \text{ standard dev.} = 0.363$$

$$\varphi_G \quad \varphi_{G,0} = \text{Weibull distribution: shape} = 2.826, \text{ scale} = 5.665$$



**Figure 2.** Best approximations of  $\varphi_I$  (left) and  $\varphi_G$  (right) through Laguerre densities of the form (6) for various choices of  $m_1$  and  $m_2$ .

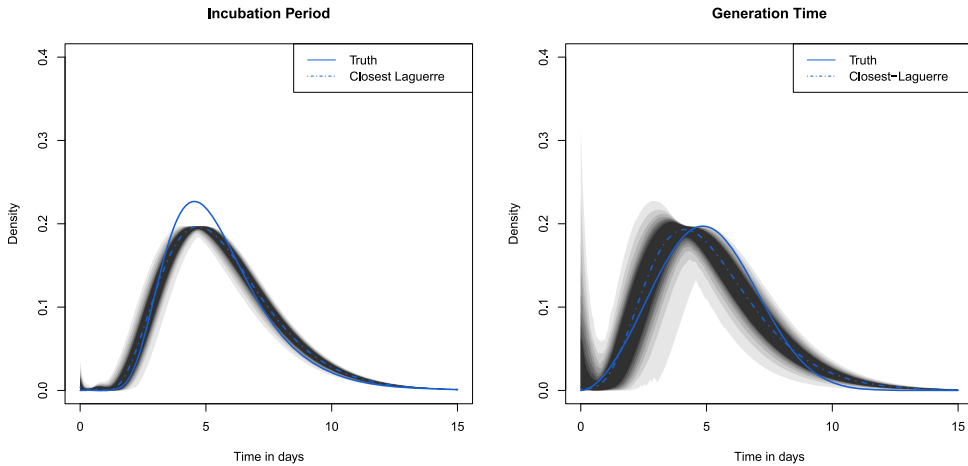
All quantities above are chosen to imitate the data used in Ferretti et al. (2020b): The average window length  $\mathbb{E}(W_i)$  equals roughly the average length of the observed exposure windows in the dataset of Ferretti et al. (2020b). Likewise, in their dataset the authors distinguish two locations, those with an exponential growth of cases with growth rate  $\log(2)/5$  (in roughly 35% of the cases) and those with no exponential growth (in roughly 65% of the cases). Moreover, the density of the incubation period  $\varphi_I = \varphi_{I,0}$  is the distribution of incubation times fitted by Ferretti et al. (2020b) and Lauer et al. (2020). Lastly, the density of the generation time is taken from Ferretti et al. (2020b). We show firstly in Figure 2 that these densities can be well approximated through Laguerre type densities as defined in (6). The graphs show those Laguerre densities which minimise the Hellinger distance to the true densities. Not-surprisingly, larger choices of the degrees  $m_1$  and  $m_2$  lead to better approximations. However, we also see that, in both cases, already degree two yields reasonable approximations. Note in particular that the flat beginning of the density of the incubation period can be well captured by the approximating densities.

In order to determine the degree of the approximating Laguerre polynomials, we use the information criterion BIC. Since the computations are quite intensive we do the model selection only for one dataset and choose the selected model for all subsequent repetitions. The resulting values of the BIC are shown in Table 1 and it can be seen that the minimal value is obtained for  $m_1 = 2$  and  $m_2 = 2$ . For the remainder of this Section, we use these degrees.

We simulate now  $N = 1000$  datasets according to the above specified data generating process and fit our model to it. Figure 3 shows the true densities as solid lines and corresponding closest Laguerre densities (6) when choosing degrees  $m_1 = m_2 = 2$ , respectively. The shaded areas show point-wise confidence bands (constructed based on simulations),

**Table 1.** BIC values for one synthetic dataset.

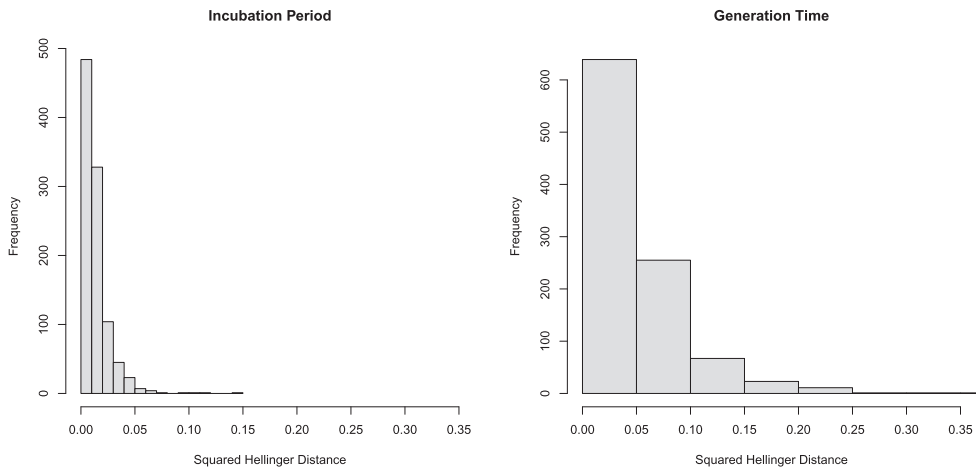
		$m_2$			
		1	2	3	4
$m_1$	1	486.77	421.91	418.82	422.37
	2	413.26	395.89	399.58	403.27
	3	416.95	397.70	401.10	404.55
	4	418.39	401.02	404.71	408.19



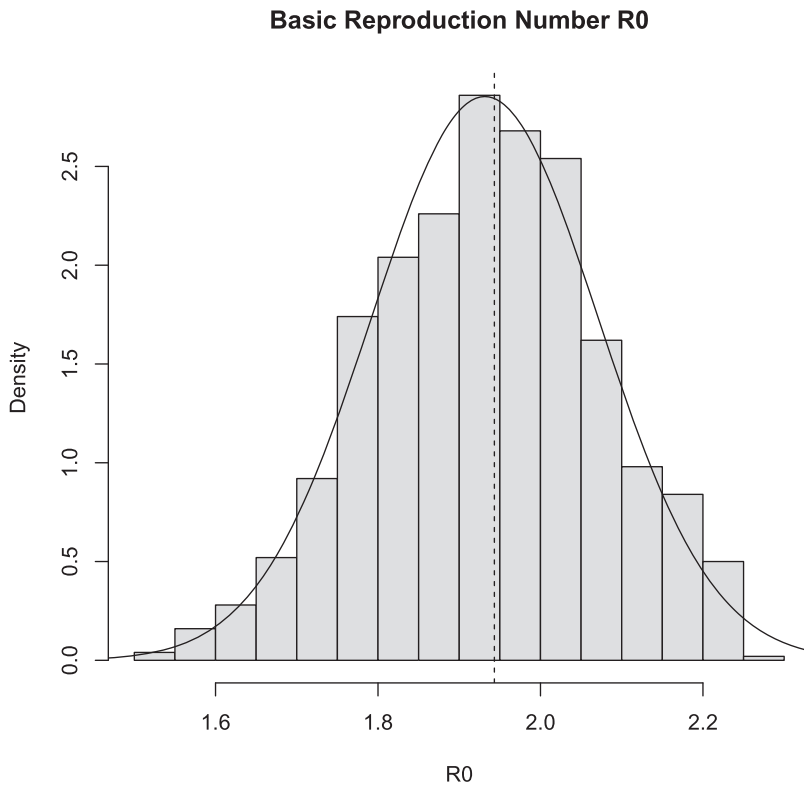
**Figure 3.** Estimation results for simulated data with  $n = 40$  observations. True densities are shown as solid lines and the dashed lines show the Laguerre densities which come closest to the true densities. The shaded areas show the point-wise simulated confidence areas from 99% (lightest gray) over 95% in 5%-steps to 60% in black.

e.g. 90% of the estimators lie point-wise in the confidence band of level 90%. Different confidence levels (99, 95, 90, 85, 80, 75, 70, 65, 60) are indicated by different intensities of gray (from light gray to black). It can be seen that the estimation works visibly quite well for the incubation period. The height of the mode is underestimated but this comes from the fact that fitting an order two Laguerre density cannot do better. The estimation for the generation time works a bit less good but still the general trend is captured well by most estimators if we keep in mind that we have here only 40 observations of a heavily convolved variable. In order to assess the fit of the nonparametric estimator a bit more formally, we compare the estimators to the true densities in terms of the squared Hellinger-distance  $\rho_H^2$ . The resulting histogram is shown in Figure 4. Note that squared Hellinger distances are bounded from above by 2. It can be seen that the estimation of the Incubation Period works better than the estimation of the Generation Time. This is not surprising because we observe transmission pairs each of which contains two independent realisations of incubation periods but only one generation time. Overall the fit of the generation time appears to be fairly good (90% of the distances are smaller than 0.103 and 50% of the distances are smaller than 0.034).

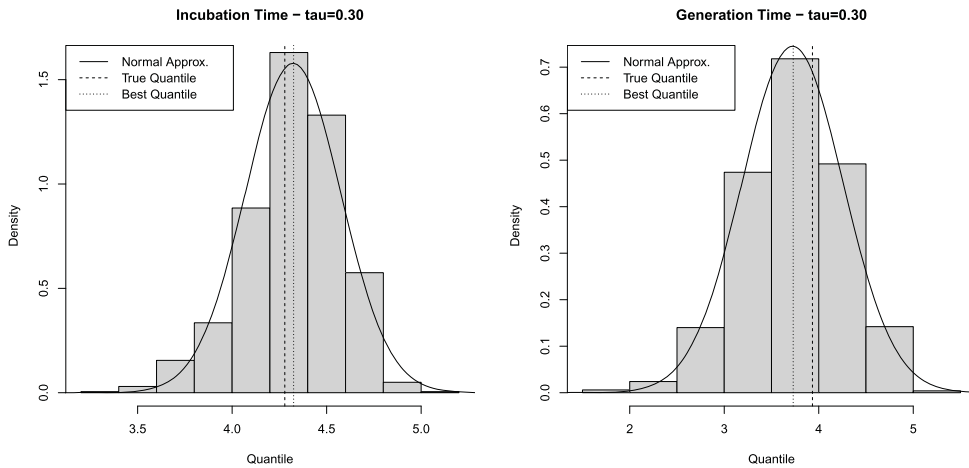
In order to illustrate that the corresponding plug-in feature estimates enjoy asymptotic normality properties, we consider estimation of the basic reproduction number of a fictional pandemic and estimation of quantiles. In order to estimate the basic reproduction



**Figure 4.** Histograms of squared Hellinger distances of estimates to true densities.



**Figure 5.** Histograms of estimated basic reproduction numbers of fictional pandemic. The dotted line indicates the true  $R_0$  and the solid line is the density of a normal distribution with the sample mean and standard deviation.



**Figure 6.** Histograms of estimated 30%-quantile for the incubation time (left) and generation time (right). The dashed lines indicate the respective true values, the dotted lines the quantiles of the best approximating densities and the solid line is the density of a normal distribution with the corresponding sample mean and standard deviation.

number, we plug the true and the estimated generation time densities into Equation (1) and take the inverse of it (in this fictional pandemic we choose  $r = \log(2)/5$  which means that the case numbers double every five days). The histogram of the estimates is shown in Figure 5. We emphasise that these are simulated results and we cannot draw any conclusions about COVID-19. We can see that the estimation appears to be almost unbiased even in finite samples. This is remarkable because the estimator is based on a nonparametric estimator which is biased (we do not control for over- or under-smoothing). Moreover, the approximation through a normal distribution seems to be reasonably accurate. This motivates further research in establishing a formal asymptotic normality result for this type of semi-parametric inference. Figure 6 shows the histograms of the estimated 30%-quantiles. We can see that the estimates appear to have a bias in finite samples (the dashed lines indicate the true quantiles). This bias seems to stem from the fact that the approximation through low dimensional Laguerre polynomials is not perfect because the estimates centralise around the quantiles of the best Laguerre approximations (dotted lines). This effect remains true for other quantiles which are reported in Appendix A.2.1.

When inspecting the BIC values in Table 1 we see that the next smallest scores are obtained for the models  $(m_1 = 2, m_2 = 3)$  and  $(m_1 = 3, m_2 = 2)$ . In order to check the robustness of the model selection, we also inspect the next smallest BIC model, i.e.  $(m_1 = 3, m_2 = 2)$ . The results are similar and are shown in Appendix A.2.1.

## 5.2. Real data application

Finally we apply our methodology to a dataset containing 191 transmission pairs which has been studied by Hart et al. (2021) and Ferretti et al. (2020a) and can be downloaded from <https://elifesciences.org/articles/65534/figures#content> (see the source data corresponding to Figure 2). The data-set is a compilation of five data-sets: Ferretti et al. (2020b), He



**Table 2.** BIC values for transmission pair data.

		$m_2$					
		1	2	3	4	5	6
$m_1$	1	1572.22	1474.16	1460.14	1455.04	1460.29	1459.01
	2	1452.06	1423.16	1408.44	1413.69	1418.70	1417.45
	3	1436.92	1409.72	1411.41	1412.70	1417.56	1421.27
	4	1424.08	1408.84	1405.98	1410.06	1415.09	1415.36
	5	1425.95	1412.56	1408.15	1413.13	1418.06	1420.15
	6	1428.97	1414.82	1413.40	1417.10	1422.33	1424.50

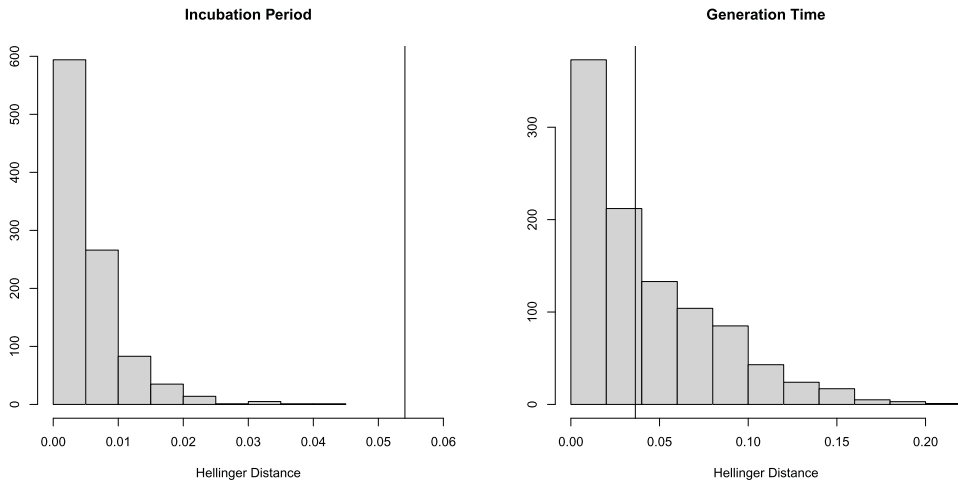
et al. (2020), Xia et al. (2020), Cheng et al. (2020), and Zhang et al. (2020). In all data-sets the authors had access to or collected data on transmission pairs. In all cases the authors ensure that the transmission pairs are indeed true transmission pairs e.g. by examining contact and travel histories or quarantines of the involved people. The data was collected with different targets concerning the transmissibility of SARS-COV-2. Thus we can reasonably illustrate our methodology on this dataset: Semi-parametric estimation of features of the generation time and the incubation time.

The dataset contains symptom onset dates for all transmission pairs, but the exposure window is not always reported. In that case we impute the dataset in the same way as Ferretti et al. (2020b): The beginning of the exposure window is at the earliest 60 days before symptom onset of the infector. The end of the infection window is at the latest the symptom onset time of any of the two people in the pair or the end of the exposure window of the second person which is reported in some cases. Since the data is discrete, i.e. we know only the days of symptom onset rather than the exact time, we suppose that the exact time is uniformly distributed throughout the day. Therefore we add a uniform random time between 0 and 24 hours to the symptom onset times and exposure window end points. We stress that our interest lies in the theoretical analysis of the methodology and we provide here an illustration for how our methodology can be used. A complete data analysis would for example also require a robustness analysis against potential issues like the question whether the transmission pairs are random samples from the pandemic.

In the following our aim is to use our semi-parametric approach to construct a test whether the parametric fit suggested by Ferretti et al. (2020b) is appropriate for the data. Let  $\varphi_{I,0}$  and  $\varphi_{G,0}$  denote the densities as defined in Section 5.1. More formally, we would like to test the hypotheses

$$H_0^{(I)} : \varphi_I = \varphi_{I,0} \quad \text{and} \quad H_0^{(G)} : \varphi_G = \varphi_{G,0}.$$

As test statistics we consider  $\rho_H(\hat{\varphi}_{I,n}, \tilde{\varphi}_{I,0})^2$  and  $\rho_H(\hat{\varphi}_{G,n}, \tilde{\varphi}_{G,0})^2$ , where  $\tilde{\varphi}_{I,0}$  and  $\tilde{\varphi}_{G,0}$  denote the best approximations with respect to the Hellinger distance of  $\varphi_{I,0}$  and  $\varphi_{G,0}$  through Laguerre polynomials, respectively. In order to choose the degrees of the approximation we use in the same way as before the BIC. The resulting values are shown in Table 2 and it can be seen that  $m_1 = 4$  and  $m_2 = 3$  yields the smallest BIC. Figure 2 shows that  $m_1 = 4$  yields already a good approximation to  $\varphi_{I,0}$ , similarly we see that  $m_2 = 3$  allows a good approximation of  $\varphi_{G,0}$ . However, in both cases, the representation is not perfect. Therefore, it is very important that we compare the estimate with the closest Laguerre-type density  $\tilde{\varphi}_{G,0}$  rather than with  $\varphi_{G,0}$  directly.



**Figure 7.** Histograms of squared Hellinger distances of  $N = 1000$  simulations from the model under  $H_0^{(I)} \cap H_0^{(G)}$ . Distances are computed between the estimated model and the closest Laguerre type model. The vertical lines indicate the observed values of the test statistic.

In order to assess the distribution of the test statistic under  $H_0^{(I)}$ ,  $H_0^{(G)}$  or  $H_0^{(I)} \cap H_0^{(G)}$ , we suppose that the densities of the incubation period and generation time are given by  $\varphi_{I,0}$  and  $\varphi_{G,0}$ , respectively, and simulate data accordingly including the rounding to complete days and adding uniform noise (note that when testing for  $H_0^{(I)}$  or  $H_0^{(G)}$  we suppose that the respective other density is correctly specified). We do this  $N = 1000$  times and show in Figure 7 the histograms of the squared Hellinger distances between the estimated densities and those Laguerre densities of type (6) (with  $m_1 = 4$  and  $m_2 = 3$ ) which lie closest to the true densities. Note, that in Figure 4 we compared the estimates with the true densities, so both figures cannot be compared. The observed squared Hellinger distances from the dataset are 0.05408553 for the incubation period and 0.03635778 for the generation time. These values are shown as vertical lines in Figure 7. In our simulations none of the simulated Hellinger distances for the incubation period are larger than the observed distance. Consequently, in none of the simulated cases both distances are simultaneously larger than the observed distances. For the generation time 44.8% of the simulated distances are larger than the observed distance. We conclude that the data-set we considered here suggests that the suggested parametric fit for the incubation time might not be correct. For the generation time, the data-set contains no evidence which suggests that the parametric fit might be wrong.

In Table 2 we see that the BIC values for all choices of  $m_1 = 3, 4, 5$  and  $m_2 = 2, 3, 4$  are all very similar. In order to check the robustness of the method to the model selection, we provide in Section A.2.2 in the Appendix a similar analysis for the second smallest BIC:  $m_1 = 5$  and  $m_2 = 3$ .

We stress that the above analysis should be understood as a recipe for a data analysis rather than an in-depth analysis of the provided dataset. Other types of semi-parametric analyses like the ones outlined in Section 5.1 can be implemented in a similar way.

## 6. Conclusion and related questions

We have introduced a semi-parametric estimator for the generation time and incubation period from observational data. We have shown that both distributions can be identified from the observations and we presented a simple, consistent semi-parametric estimator which is based on Laguerre polynomials. These results are the first steps for more results in the general realm of semi-parametric inference for epidemics: As specific examples we mention the reproduction number  $R_0$  and tests for parametric assumptions. But also the probability of pre-symptomatic infection  $\mathbb{P}(G \leq I_1)$  can be of interest. All these quantities are continuous functions of the densities  $\varphi_I$  and  $\varphi_G$  and therefore it can be expected that asymptotic normality results for estimators based on our estimators  $\hat{\varphi}_{I,n}$  and  $\hat{\varphi}_{G,n}$  can be proven. However, it should be mentioned that such results for  $R_0$  are possibly more challenging because they require another estimator  $\hat{r}_n$ .

This work can be extended in several directions. It might for example not be clear when symptoms exactly start. Therefore, it might be possible that just a window for symptom onset can be supplied (Lauer et al. 2020). Moreover, it can be of interest to include asymptomatic patients by including a certain *cure-probability*, i.e. the probability with which patients will never show symptoms. As an alternative one could also consider follow up studies in which the symptom onset time of patients is considered as a censored variable (in which case asymptomatic patients can be interpreted as patients who show symptoms at  $\infty$ ). Finally, the dataset by Ferretti et al. (2020b) includes also a window for infection of the second person. It would of course be interesting to include this information in the model. But we would like to point out that this is not entirely trivial because it is unclear how to model the time of infection within this window. As we motivated in the discussion after Remark 2.1 a simple uniformity assumption is possibly difficult to justify. Therefore, we would suggest to include this distribution in the estimation in a suitable way. In theoretic terms, the most interesting question would certainly be to establish asymptotic normality results which allow the researcher to make quantitative statements. Such statements can for example be achieved by following the semi-parametric framework as e.g. in Shen (1997). Another interesting question would be how to adequately incorporate covariates in the model. It could be the case that, e.g. younger and older people have different incubation periods or generation times. Finally, in a different branch of literature, one tries to avoid the assumption of independence between symptom onset and infection and rather models the infection time relative to the symptom onset (cf. Chun, Baek, and Kim 2020; Ferretti et al. 2020a; Hart et al. 2021). Our methodology can be applied to this setting as well with some adjustments of the theory.

## Acknowledgements

We are grateful to Niel Hens and Cathrine Aeckerle-Willems for useful discussions. Moreover we thank an unknown associate editor and two anonymous referees for reading and commenting our paper. Their remarks have lead to a great improvement of the paper.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the European Research Council under Grant 2016-2021, Horizon 2020/ERC grant agreement No. 694409.

## References

- Ai, C., and Chen, X. (2003), 'Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions', *Econometrica*, 71(6), 1795–1843.
- Azmon, A., Faes, C., and Hens, N. (2014), 'On the Estimation of the Reproduction Number Based on Misreported Epidemic Data', *Statistics in Medicine*, 33(7), 1176–1192.
- Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T., Gao, W., Cheng, C., Tang, X., Wu, X., Wu, Y., Sun, B., Huang, S., Sun, Y., Zhang, J., Ma, T., Lessler, J., and Feng, T. (2020), 'Epidemiology and Transmission of Covid-19 in 391 Cases and 1286 of Their Close Contacts in Shenzhen, China: A Retrospective Cohort Study', *The Lancet Infectious Diseases*, 20(8), 911–919.
- Billingsley, P. (1968), *Convergence of Probability Measures*, New York, NY: John Wiley and Sons.
- Britton, T., and Scalia Tomba, G. (2019), 'Estimation in Emerging Epidemics: Biases and Remedies', *Journal of the Royal Society, Interface*, 16(150). <https://doi.org/10.1098/rsif.2018.0670>
- Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, New York, NY: Chapman and Hall/CRC.
- Chen, X. (2007), 'Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models', in *Handbook of Econometrics* (Vol. 6), Amsterdam: Elsevier, pp. 5549–5632.
- Chen, X., and Shen, X. (1998), 'Sieve Extremum Estimates for Weakly Dependent Data', *Econometrica*, 66(2), 289–314.
- Cheng, H.-Y., Jian, S.-W., Liu, D.-P., Ng, T.-C., Huang, W.-T., Lin, H.-H., and for the Taiwan COVID-19 Outbreak Investigation Team (2020), 'Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk At Different Exposure Periods Before and After Symptom Onset', *JAMA Internal Medicine*, 180(9), 1156–1163.
- Chowell, G., Hayman, J.M., Bettencourt, L.M.A., and Castolli-Chavez, C. (2009), *Mathematical and Statistical Estimation Approaches in Epidemiology*, Dordrecht: Springer.
- Chun, J.Y., Baek, G., and Kim, Y. (2020), 'Transmission Onset Distribution of Covid-19', *International Journal of Infectious Diseases*, 99, 403–407. ISSN 1201-9712. <https://doi.org/10.1016/j.ijid.2020.07.075>.
- Devroye, L. (1989), 'Consistent Deconvolution in Density Estimation', *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 17(2), 235–239.
- Fan, J. (September, 1991), 'On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems', *Annals of Statistics*, 19(3), 1257–1272.
- Ferretti, L., Ledda, A., Wymant, C., Zhao, L., Ledda, V., Abeler-Dörner, L., Kendall, M., Nurtay, A., Cheng, H.-Y., Ng, T.-C., Lin, H.-H., Hinch, R., Masel, J., Kilpatrick, A.M., and Fraser, C. (2020a), 'The Timing of Covid-19 Transmission', medRxiv. <https://doi.org/10.1101/2020.09.04.20188516>.
- Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., and Fraser, C. (2020b), 'Quantifying Sars-Cov-2 Transmission Suggests Epidemic Control with Digital Contact Tracing', *Science*, 368(6491), eabb6936.
- Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., and Hens, N. (2020), 'Estimating the Generation Interval for Coronavirus Disease (Covid-19) Based on Symptom Onset Data, March 2020', *Eurosurveillance*, 25(17), Article ID 2000257. <https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257>.
- Groeneboom, P. (2021), 'Estimation of the Incubation Time Distribution for COVID-19', *Statistica Neerlandica*, 75(2), 161–179.
- Härdle, W., and Mammen, E. (December, 1993), 'Comparing Nonparametric Versus Parametric Regression Fits', *Annals of Statistics*, 21(4), 1926–1947. <https://doi.org/10.1214/aos/1176349403>.

- Hart, W.S., Maini, P.K., and Thompson, R.N. (April, 2021), ‘High Infectiousness Immediately Before Covid-19 Symptom Onset Highlights the Importance of Continued Contact Tracing’, *eLife*, 10, Article ID e65534. <https://doi.org/10.7554/eLife.65534>.
- He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B.J., Li, F., and Leung, G.M. (2020), ‘Temporal Dynamics in Viral Shedding and Transmissibility of Covid-19’, *Nature Medicine*, 26(5), 672–675.
- Held, L., Hens, N., O’Neill, P., and Wallinga, J. (2019), *Handbook of Infectious Disease Data Analysis*, New York, NY: Chapman and Hall/CRC.
- 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. <https://doi.org/10.7326/M20-0504>.
- Leavitt, S.V., Lee, R.S., Sebastiani, P., Horsburgh Jr., C.R., Jenkins, H.E., and White, L.F. (March, 2020), ‘Estimating the Relative Probability of Direct Transmission Between Infectious Disease Patients’, *International Journal of Epidemiology*, 49(3), 764–775. <https://doi.org/10.1093/ije/dyaa031>.
- Lipsitch, M., Cohen, T., Cooper, B., Robins, J.M., Ma, S., James, L., Gopalakrishna, G., Chew, S.K., Tan, C.C., Samore, M.H., Fisman, D., and Murray, M. (2003), ‘Transmission Dynamics and Control of Severe Acute Respiratory Syndrome’, *Science*, 300(5627), 1966–1970.
- Newey, W.K. (1997), ‘Convergence Rates and Asymptotic Normality for Series Estimators’, *Journal of Econometrics*, 79(1), 147–168.
- Newey, W.K., and Powell, J.L. (2003), ‘Instrumental Variable Estimation of Nonparametric Models’, *Econometrica*, 71(5), 1565–1578.
- Nikiforov, A.F., and Uvarov, V.B. (1988), *Special Functions of Mathematical Physics*, Basel: Birkhäuser.
- Nishiura, H., Kakehashi, M., and Inaba, H. (2009), ‘Two Critical Issues in Quantitative Modeling of Communicable Diseases: Inference of Unobservables and Dependent Happening’, in *Mathematical and Statistical Estimation Approaches in Epidemiology*, Springer Netherlands, pp. 53–87. [https://doi.org/10.1007/978-90-481-2313-1\\_3](https://doi.org/10.1007/978-90-481-2313-1_3).
- Ossiander, M. (1987), ‘A Central Limit Theorem Under Metric Entropy with  $l_2$  Bracketing’, *Annals of Probability*, 15(3), 897–919.
- Shen, X. (December, 1997), ‘On Methods of Sieves and Penalization’, *Annals of Statistics*, 25(6), 2555–2591.
- Shiryayev, A.N. (2016), *Probability-1: Volume 1*, New York: Springer. ISBN 978-0-387-72206-1.
- Tindale, L.C., Stockdale, J.E., Coombe, M., Garlock, E.S., Lau, W.Y.V., Saraswat, M., Zhang, L., Chen, D., Wallinga, J., and Colijn, C. (June, 2020), ‘Evidence for Transmission of Covid-19 Prior to Symptom Onset’, *eLife*, 9, Article ID e57149. <https://doi.org/10.7554/eLife.57149>.
- van der Vaart, A.W., and Wellner, J.A. (1996), *Weak Convergence and Empirical Process Theory*, New York, NY: Springer.
- Wallinga, J., and Lipsitch, M. (2007), ‘How Generation Intervals Shape the Relationship Between Growth Rates and Reproductive Numbers’, *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 599–604. <https://doi.org/10.1098/rspb.2006.3754>.
- Wong, W.H., and Shen, X. (April, 1995), ‘Probability Inequalities for Likelihood Ratios and Convergence Rates of Sieve Mles’, *Annals of Statistics*, 23(2), 339–362.
- Xia, W., Liao, J., Li, C., Li, Y., Qian, X., Sun, X., Xu, H., Mahai, G., Zhao, X., Shi, L., Liu, J., Yu, L., Wang, M., Wang, Q., Namat, A., Li, Y., Qu, J., Liu, Q., Lin, X., Cao, S., Huan, S., Xiao, J., Ruan, F., Wang, H., Xu, Q., Ding, X., Fang, X., Qiu, F., Ma, J., Zhang, Y., Wang, A., Xing, Y., and Xu, S. (2020), ‘Transmission of Corona Virus Disease 2019 During the Incubation Period May Lead to a Quarantine Loophole’, medRxiv. <https://doi.org/10.1101/2020.03.06.20031955>.
- Zhang, M., and Davidian, M. (2008), ‘“smooth” Semiparametric Regression Analysis for Arbitrarily Censored Time-to-Event Data’, *Biometrics*, 64(2), 567–576. <https://doi.org/10.1111/j.1541-0420.2007.00928.x>.
- Zhang, J., Litvinova, M., Wang, W., Wang, Y., Deng, X., Chen, X., Li, M., Zheng, W., Yi, L., Chen, X., Wu, Q., Liang, Y., Wang, X., Yang, J., Sun, K., Longini Jr., I.M., Halloran, M.E., Wu, P., Cowling,

B.J., Merler, S., Viboud, C., Vespignani, A., Ajelli, M., and Yu, H. (2020), ‘Evolving Epidemiology and Transmission Dynamics of Coronavirus Disease 2019 Outside Hubei Province, China: a Descriptive and Modelling Study’, *The Lancet Infectious Diseases*, 20(7), 793–802.

## Appendix

### A.1 A consistency result and approximation through Laguerre polynomials

In this section we state two results from the literature which are relevant for this paper.

The following Theorem is just a re-formulation for a special case of Theorem 4 in Wong and Shen (1995) which is stated here for the convenience of the reader: In their paper the authors study approximate sieve estimation, i.e. they allow that the estimator only approximately maximises the likelihood. In their notation this means that we assume in the present paper that  $\eta_n = 0$ , this is already included in the following version of Theorem 4 of Wong and Shen (1995).

**Theorem A.1:** *Let  $Y_1, \dots, Y_n$  be iid observations which have a density  $p_0$ . Let moreover  $\mathcal{F}_n$  be an arbitrary sequence of sieve spaces for density estimation and let  $\hat{p}_n$  denote the sieve-MLE. Suppose that there are constants  $c_1, c_2 > 0$  and a sequence  $\varepsilon_n > 0$  such that*

$$\int_{\frac{\varepsilon_n}{2^8}}^{\sqrt{2}\varepsilon_n} \sqrt{\log \mathcal{N}_{[]} \left( \frac{u}{c_1}, \mathcal{F}_n, \rho_H \right)} \, du \leq c_2 \sqrt{n} \varepsilon_n^2.$$

*Let  $\delta_n(\alpha) := \inf_{q \in \mathcal{F}_n} \rho_\alpha(p_0, q)$  for  $\alpha \in (0, 1]$ . Suppose that there is  $\alpha \in (0, 1]$  such that  $\delta_n(\alpha) < 1/\alpha$ . Then, there is a constant  $c > 0$  (which depends on the model) such that for*

$$\varepsilon_n^*(\alpha) := \max \left( \varepsilon_n, \sqrt{\frac{4\delta_n(\alpha)}{c}} \right)$$

*we have for another constant  $C > 0$*

$$\mathbb{P} \left( \rho_H(\hat{p}_n, p_0) \geq \varepsilon_n^*(\alpha) \right) \leq 5 \exp \left( -Cn\varepsilon_n^*(\alpha)^2 \right) + \exp \left( -\frac{1}{4}n\alpha c\varepsilon_n^*(\alpha)^2 \right).$$

The next result is a statement about approximating functions by using Laguerre polynomials. The following is a combination of Theorem 1 and Remark 2 in Chapter II.8 of Nikiforov and Uvarov (1988). We formulate this theorem here in our setting. The original statement is more general.

**Theorem A.2:** *Let  $f : [0, \infty) \rightarrow \mathbb{R}$  be continuous and have a piecewise continuous derivative  $f'$ . Consider the series*

$$f_N(x) := \sum_{k=0}^N c_k L_k(x), \quad c_k := \int_0^\infty f(x) L_k(x) e^{-x} \, dx.$$

*If the integrals*

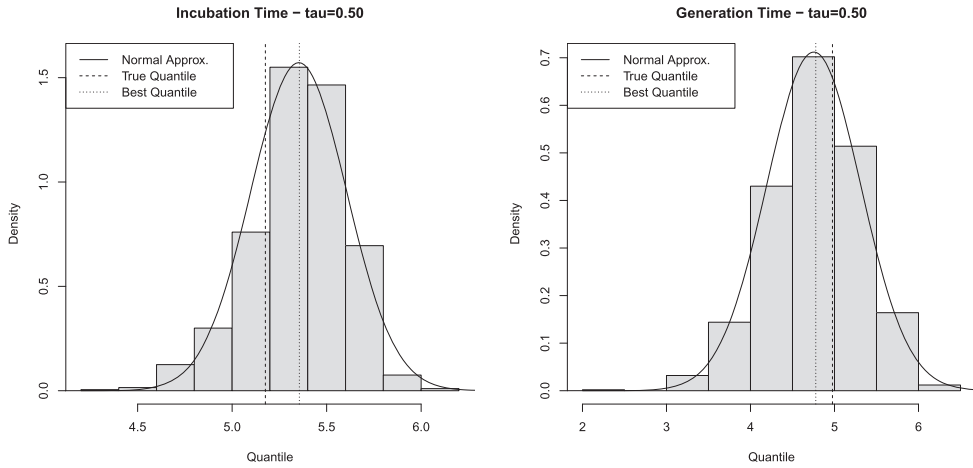
$$\int_0^\infty f(x)^2 e^{-x} \, dx \text{ and } \int_0^\infty f'(x)^2 x e^{-x} \, dx$$

*converge, we have  $f_N \rightarrow f$  for  $N \rightarrow \infty$  uniformly over compact sets  $[x_1, x_2] \subseteq (0, \infty)$  and  $\lim_{N \rightarrow \infty} \int_0^\infty [f(x) - f_N(x)]^2 e^{-x} \, dx \rightarrow 0$ .*

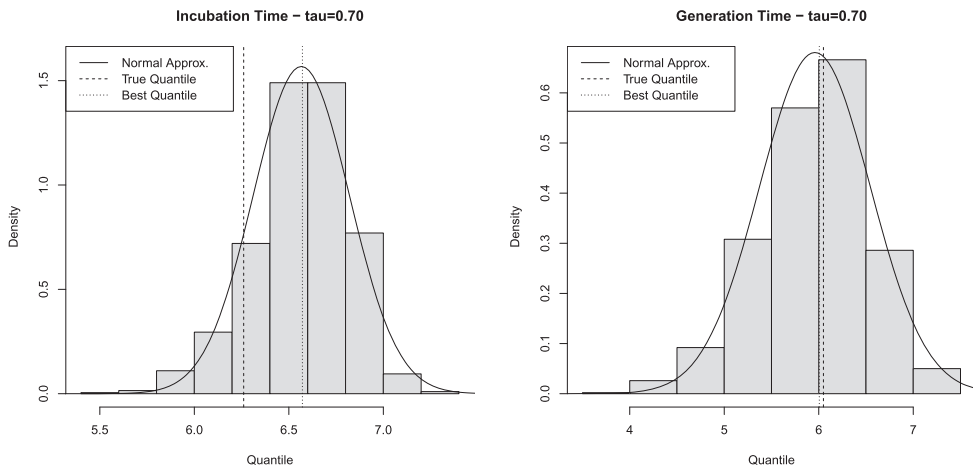
### A.2 Further empirical results

#### A.2.1 Simulation study

In this Section we present additional simulation results complementing the results from Section 5.1. Figures A1–A3 show histograms for estimation of the 50%-, 70%- and 90%-



**Figure A1.** Histograms of estimated 50%-quantile for the incubation time (left) and generation time (right). The dashed lines indicate the respective true values, the dotted lines the quantiles of the best approximating densities and the solid line is the density of a normal distribution with the corresponding sample mean and standard deviation.

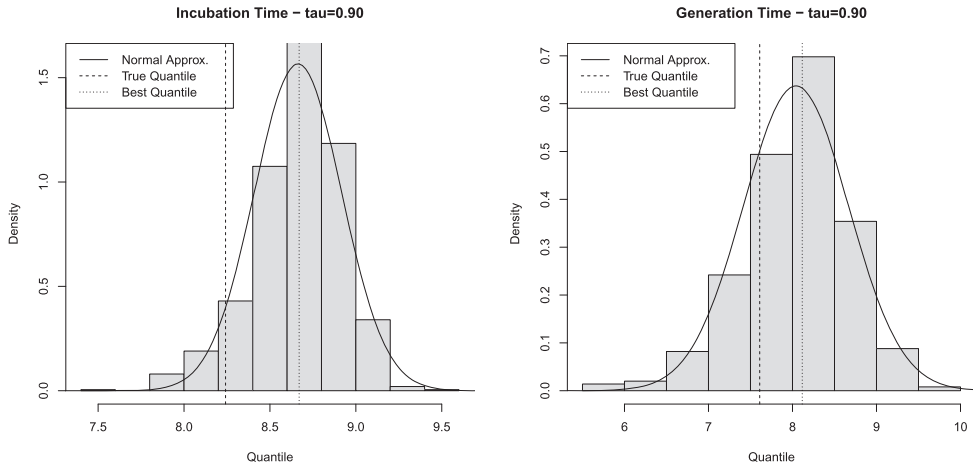


**Figure A2.** Histograms of estimated 70%-quantile for the incubation time (left) and generation time (right). The dashed lines indicate the respective true values, the dotted lines the quantiles of the best approximating densities and the solid line is the density of a normal distribution with the corresponding sample mean and standard deviation.

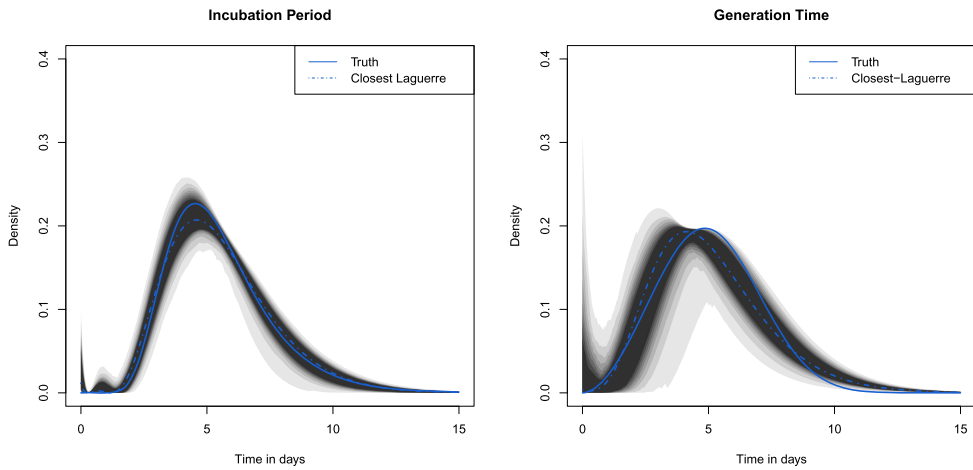
quantiles. The results are very similar to the results for the 30%-quantile which are discussed in Section 5.1.

Next, we use the same set-up as in Section 5.1 in the main text, however, here we choose as a model  $m_1 = 3$  and  $m_2 = 2$ , i.e. the model has now more flexibility for the incubation time. The estimation results are visualised in Figure A4 and the difference in the squared Hellinger distance is shown in Figure A5. In Figure A4 it appears that for the incubation time the mode moves closer to its true location, it is sometimes even overestimated (compared to the case  $m_2 = 2$  which was shown in Section 5.1). Moreover, the estimated incubation time densities seem to fluctuate more, i.e. they show a higher variance due to the higher flexibility. The estimates for the generation time appear





**Figure A3.** Histograms of estimated 90%-quantile for the incubation time (left) and generation time (right). The dashed lines indicate the respective true values, the dotted lines the quantiles of the best approximating densities and the solid line is the density of a normal distribution with the corresponding sample mean and standard deviation.

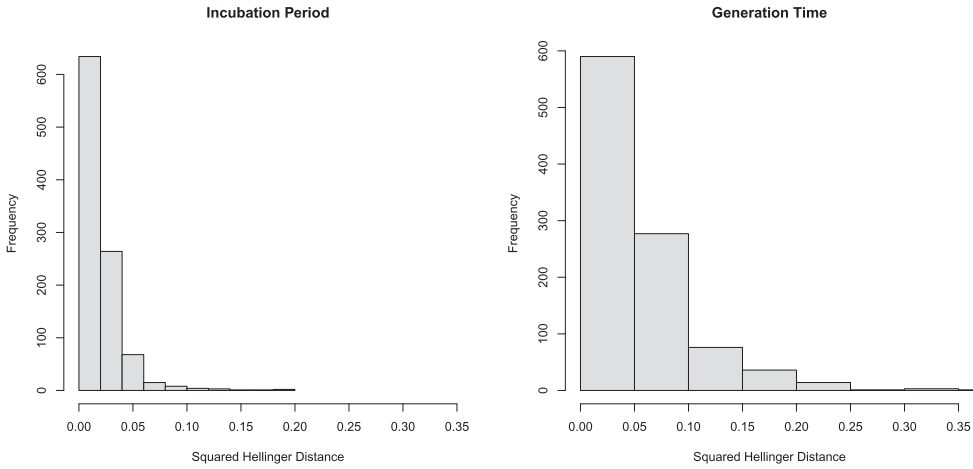


**Figure A4.** Estimation results for simulated data with  $n = 40$  observations. True densities are shown as solid lines and the dashed lines show the Laguerre densities which come closest to the true densities. The shaded areas show the point-wise simulated confidence areas from 99% (lightest gray) over 95% in 5%-steps to 60% in black. Model choices:  $m_1 = 2$  and  $m_2 = 3$ .

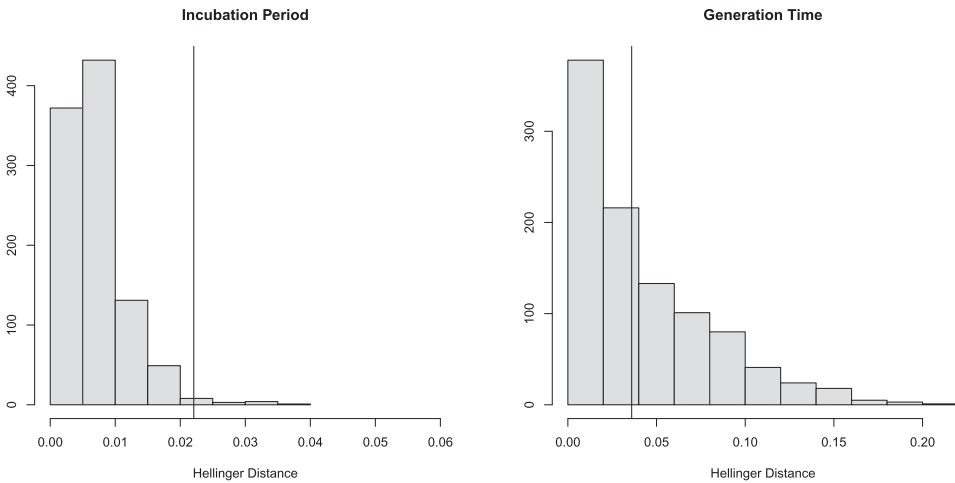
to be almost identical. In terms of the Hellinger distances, cf. histograms in Figure A5, the results appear to be very similar to the results obtained in Section 5.1. In general we see that estimation with  $n = 40$  observations yields reasonable results for both model complexities. However, we expect that the estimation can be improved if more observations are available enabling the method to choose better approximations.

### A.2.2 Real data application

In this section we show an analysis similar to that from Section 5.2 in the main text, but here we choose  $m_1 = 5$  and  $m_2 = 3$ . We begin with a simulation: We generate  $N = 1000$  datasets assuming



**Figure A5.** Histograms of squared Hellinger distances of estimates to true densities. Model Choices:  $m_1 = 3$  and  $m_2 = 2$ .



**Figure A6.** Histogram of squared Hellinger distances of  $N = 1000$  simulations from the model under  $H_0^{(I)} \cap H_0^{(G)}$ . Distances are computed between the estimated model and the closest Laguerre type model (for  $m_1 = 5$  and  $m_2 = 3$ ). The vertical lines indicate the observed values of the test statistic.

that the models specified in  $H_0^{(I)}$  and  $H_0^{(G)}$  are correct. Figure A6 shows the squared Hellinger distances of the estimates to the closest Laguerre type densities with  $m_1 = 5$  and  $m_2 = 3$ . The vertical lines show the distances which are obtained from the real data set. In case of the incubation period only 1% of the simulated distances are larger, for the generation time this percentage is higher, 44.9%. Finally, both distances are simultaneously larger for both densities in 0.3% of the simulated cases. Thus, we conclude that also in this larger model class there is no evidence in the dataset which would contradict the parametric model for the generation time fitted by Ferretti et al. (2020b). However the parametric model for the incubation time used by Ferretti et al. (2020b) and Lauer et al. (2020) might be questionable depending on the desired level of the test.

### A.3 Proofs of Section 4.1

**Proof of Corollary 4.4:** For the proof we make the following definitions: Let  $\Phi_I$ ,  $\Phi_G$  and  $\Phi_{S_1, S_2, W, C}$  be the characteristic functions of the corresponding random variables, i.e. for real numbers  $x, x_1, \dots, x_4 \in \mathbb{R}$  (note that in the definition of  $\varphi_I$  it doesn't matter whether we take  $I_1$  or  $I_2$ )

$$\begin{aligned}\Phi_I(x) &:= \mathbb{E} \left( e^{ixI_1} \right), \quad \Phi_G(x) := \mathbb{E} \left( e^{ixG} \right), \\ \Phi_{S_1, S_2, W, C}(x_1, x_2, x_3, x_4) &:= \mathbb{E} \left( e^{i(x_1 S_1 + x_2 S_2 + x_3 W + x_4 C)} \right).\end{aligned}$$

As a first step, we compute  $\Phi_{S_1, S_2, W, C}$ . We have by definition of  $(S_1, S_2, W, C)$  and independence

$$\begin{aligned}\Phi_{S_1, S_2, W, C}(x_1, x_2, x_3, x_4) &= \mathbb{E} \left( e^{i(x_1 + x_2)T_1} e^{ix_1 I_1} e^{ix_2 I_2} e^{ix_3 G} e^{ix_4 C} \right) \\ &= \mathbb{E} \left( e^{ix_3 W + ix_4 C} \mathbb{E} \left( e^{i(x_1 + x_2)T_1} \mid W, C \right) \right) \Phi_I(x_1) \Phi_I(x_2) \Phi_G(x_2) \\ &= \Psi(x_1 + x_2, x_3, x_4) \Phi_I(x_1) \Phi_I(x_2) \Phi_G(x_2),\end{aligned}$$

where

$$\Psi(x, z_1, z_2) := \mathbb{E} \left( e^{iz_1 W + iz_2 C} \mathbb{E} \left( e^{ixT_1} \mid W, C \right) \right).$$

Let now  $z_1(x), z_2(x)$  be as in Assumption (C). We obtain the following two equations

$$\begin{aligned}\Phi_{S_1, S_2, W, C}(x, 0, z_1(x), z_2(x)) &= \Psi(x, z_1(x), z_2(x)) \Phi_I(x), \\ \Phi_{S_1, S_2, W, C}(0, x, z_1(x), z_2(x)) &= \Psi(x, z_1(x), z_2(x)) \Phi_I(x) \Phi_G(x).\end{aligned}$$

Note that Assumption (C) implies that  $\Phi_{S_1, S_2, W, C}(x, 0, z_1(x), z_2(x)) \neq 0$  almost everywhere. All of this, in turn, implies the relations (almost everywhere by Assumption (C))

$$\Phi_I(x) = \Psi(x, z_1(x), z_2(x))^{-1} \Phi_{S_1, S_2, W, C}(x, 0, z_1(x), z_2(x)), \quad (\text{A1})$$

$$\Phi_G(x) = \Phi_{S_1, S_2, W, C}(0, x, z_1(x), z_2(x))^{-1} \Phi_{S_1, S_2, W, C}(x, 0, z_1(x), z_2(x)). \quad (\text{A2})$$

Let  $\hat{C}_n, \hat{W}_n, \hat{T}_{1,n}, \hat{I}_{1,n}, \hat{I}_{2,n}$  and  $\hat{G}_n$  be random variables which have the relation as specified in (5) but where  $\varphi_I$  is replaced by  $\hat{\varphi}_{I,n}$  and  $\varphi_G$  is replaced by  $\hat{\varphi}_{G,n}$ . Let moreover,  $\hat{S}_{1,n}$  and  $\hat{S}_{2,n}$  be defined analogously to their corresponding quantities  $S_1$  and  $S_2$  with corresponding characteristic functions  $\hat{\Phi}_{\hat{S}_{1,n}, \hat{S}_{2,n}, \hat{W}_n, \hat{C}_n}, \hat{\Phi}_{\hat{I}_{1,n}}$  and  $\hat{\Phi}_{\hat{G}_n}$ .

Since convergence in probability implies the existence of an almost surely convergent subsequence (cf. Lemma 1.9.2 in van der Vaart and Wellner 1996), we find by Theorem 4.1 that there is an event  $\Omega$  of probability 1 and a subsequence which we indicate again by  $f_{\hat{\varphi}_{I,n}, \hat{\varphi}_{G,n}}$  such that

$$\rho_H(f_{\varphi_I, \varphi_G} P_C, f_{\hat{\varphi}_{I,n}, \hat{\varphi}_{G,n}} P_C) \rightarrow 0$$

for all results in  $\Omega$ . Theorem 1 in Chapter 3.9 in Shiryaev (2016) implies that then

$$\left| \sum_{c=1}^{K_C} \int r(s_1, s_2, w, c) (f_{\varphi_I, \varphi_G}(s_1, s_2, w \mid c) - f_{\hat{\varphi}_{I,n}, \hat{\varphi}_{G,n}}(s_1, s_2, w \mid c)) P_C(c) d(s_1, s_2, w) \right| \rightarrow 0,$$

uniformly over all functions  $r: \mathbb{R}^3 \times \{1, \dots, K_C\} \rightarrow \mathbb{C}$  with  $\|r\|_\infty \leq 1$ . Since the functions  $r(s_1, s_2, w, c) = \exp(i(x_1 s_1 + x_2 s_2 + x_3 w + x_4 c))$  are bounded for all  $x_1, x_2, x_3, x_4 \in \mathbb{R}$ , we conclude that  $\hat{\Phi}_{\hat{S}_{1,n}, \hat{S}_{2,n}, \hat{W}_n, \hat{C}_n}$  converges uniformly to  $\Phi_{S_1, S_2, W, C}$ . Hence, for almost any fixed  $x$  we eventually have  $\hat{\Phi}_{\hat{S}_{1,n}, \hat{S}_{2,n}, \hat{W}_n, \hat{C}_n}(x, 0, z_1(x), z_2(x)) \neq 0$  and thus eventually analogue versions of the relations (A1) and (A2) hold for  $\hat{\Phi}_{\hat{I}_{1,n}}$  and  $\hat{\Phi}_{\hat{G}_n}$ . As a consequence, we obtain the following point-wise almost everywhere convergences

$$\begin{aligned}\Phi_{\hat{I}_{1,n}}(x) &= \Psi(x, z_1(x), z_2(x))^{-1} \hat{\Phi}_{\hat{S}_{1,n}, \hat{S}_{2,n}, \hat{W}_n, \hat{C}_n}(x, 0, z_1(x), z_2(x)) \\ &\rightarrow \Psi(x, z_1(x), z_2(x))^{-1} \Phi_{S_1, S_2, W, C}(x, 0, z_1(x), z_2(x)) = \Phi_I(x),\end{aligned}$$

$$\begin{aligned}\Phi_{\hat{G}_n}(x) &= \Phi_{\hat{\delta}_{1,n}, \hat{\delta}_{2,n}, \hat{W}_n, \hat{C}_n}(x, 0, z_1(x), z_2(x))^{-1} \Phi_{\hat{\delta}_{1,n}, \hat{\delta}_{2,n}, \hat{W}_n, \hat{C}_n}(0, x, z_1(x), z_2(x)) \\ &\rightarrow \Phi_{S_1, S_2, W, C}(x, 0, z_1(x), z_2(x))^{-1} \Phi_{S_1, S_2, W, C}(0, x, z_1(x), z_2(x)) = \Phi_G(x).\end{aligned}$$

Now by Levy's Theorem (cf. Theorem 1.7.6 Billingsley 1968, an inspection of the proof reveals that in the univariate case convergence almost everywhere is a sufficient condition), we conclude that  $\hat{F}_{I,n} \rightarrow F_I$  and  $\hat{F}_{G,n} \rightarrow F_G$  point-wise for all realisations in  $\Omega$ . Since the distribution functions are continuous, these convergences are also uniform. Recall that the above argument holds for a subsequence and all realisations in  $\Omega$ . But we can repeat the same argument starting from a subsequence and showing in that way that every subsequence contains a sub-subsequence for which the corresponding distribution functions converge almost surely. This implies convergence in probability by Lemma 1.9.2 in van der Vaart and Wellner (1996) and the proof is complete.  $\blacksquare$

**Proof of Corollary 4.5:** We make a similar subsequence of a sub-sequence argument as in Corollary 4.4: Consider sub-sequences of  $\hat{r}_n$  and  $\hat{F}_{G,n}$  (which we again do not indicate in the notation) such that  $\hat{r}_n \rightarrow r$  almost surely and  $\|\hat{F}_{G,n} - F_G\|_\infty \rightarrow 0$  almost surely (the former is possible by assumption and the latter by Corollary 4.4). The following considerations are made for each realisation in a set of probability one on which these two convergences hold. Choose  $n$  so large such that  $\hat{r}_n \geq r/2$ . Choose moreover,  $c > 0$  such that  $t \exp(-tr/2) \leq c$  for all  $t \geq 0$ . We have

$$\begin{aligned}\int_0^\infty e^{-\hat{r}_n t} \hat{\varphi}_{G,n} dt - \int_0^\infty e^{-rt} \varphi_G(t) dt \\ = \int_0^\infty (e^{-\hat{r}_n t} - e^{-rt}) \hat{\varphi}_{G,n} dt + \int_0^\infty e^{-rt} (\hat{\varphi}_{G,n} - \varphi_G(t)) dt.\end{aligned}\tag{A3}$$

By a Taylor expansion we obtain for the first part of the right-hand side above and for some  $r^* \in [r, \hat{r}_n]$  (note that by assumption  $r^* \geq r/2$  and hence  $t \cdot \exp(-r^*t) \leq c$ )

$$\left| \int_0^\infty (e^{-\hat{r}_n t} - e^{-rt}) \hat{\varphi}_{G,n}(t) dt \right| \leq |\hat{r}_n - r| \int_0^\infty t e^{-r^* t} \hat{\varphi}_{G,n}(t) dt \leq c |\hat{r}_n - r| \rightarrow 0.$$

For the second term on the right-hand side of (A3) we obtain via integration by parts

$$\begin{aligned}\int_0^\infty e^{-rt} (\hat{\varphi}_{G,n} - \varphi_G(t)) dt \\ = e^{-rt} (\hat{F}_{G,n}(t) - F_G(t)) \Big|_{t=0}^\infty + \int_0^\infty r e^{-rt} (\hat{F}_{G,n}(t) - F_G(t)) dt\end{aligned}$$

which converges to zero since  $\hat{F}_{G,n} \rightarrow F$  uniformly. Now, a subsequence of a subsequence argument completes the proof.  $\blacksquare$

#### A.4 Proofs of Section 4.2

**Proof of Lemma 4.6:** We firstly apply Hoelder's Inequality with  $p = 1/(1 - \alpha)$  and  $q = 1/\alpha$  (use that  $\alpha \in (0, 1)$ ) to get

$$\begin{aligned}\rho_\alpha(f_1 p_C, f_2 p_C) \\ &= \frac{1}{\alpha} \sum_{c=1}^{K_C} \int_{[0, \infty)^3} f_1(x_1, x_2, \omega | c) p_C(c) \left( \left( \frac{f_1(x_1, x_2, \omega | c) p_C(c)}{f_2(x_1, x_2, \omega | c) p_C(c)} \right)^\alpha - 1 \right) d(x_1, x_2, \omega) \\ &= \frac{1}{\alpha} \sum_{c=1}^{K_C} \int_{[0, \infty)^3} \left( \frac{f_1(x_1, x_2, \omega | c)}{f_2(x_1, x_2, \omega | c)} \right)^\alpha f_1(x_1, x_2, \omega | c)^{1-\alpha} \\ &\quad \times (f_1(x_1, x_2, \omega | c)^\alpha - f_2(x_1, x_2, \omega | c)^\alpha) p_C(c) d(x_1, x_2, \omega)\end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{\alpha} \left( \sum_{c=1}^{K_C} \int_{[0,\infty)^3} \left( \frac{f_1(x_1, x_2, \omega | c)}{f_2(x_1, x_2, \omega | c)} \right)^{\frac{\alpha}{1-\alpha}} f_1(x_1, x_2, \omega | c) p_C(c) \, d(x_1, x_2, \omega) \right)^{1-\alpha} \\ &\quad \times \left( \sum_{c=1}^{K_C} \int_{[0,\infty)^3} |f_1(x_1, x_2, \omega | c)^\alpha - f_2(x_1, x_2, \omega | c)^\alpha|^{\frac{1}{\alpha}} p_C(c) \, d(x_1, x_2, \omega) \right)^\alpha. \end{aligned}$$

By using the definition of  $C_\alpha(f_1, f_2)$ , we see that (13) follows from the above if we can prove that

$$\begin{aligned} &\sum_{c=1}^{K_C} \int_{[0,\infty)^3} |f_1(x_1, x_2, \omega | c)^\alpha - f_2(x_1, x_2, \omega | c)^\alpha|^{\frac{1}{\alpha}} p_C(c) \, d(x_1, x_2, \omega) \\ &\quad \leq 2\rho_H(\varphi_{G,1}, \varphi_{G,2})^2 + 8\rho_H(\varphi_{I,1}, \varphi_{I,2})^2. \end{aligned} \quad (\text{A4})$$

We begin by applying the reverse triangle inequality for the  $L^{1/\alpha}$  norm repeatedly. Define to this end

$$\begin{aligned} B_1(t; x_1, x_2, c, \omega, y) &:= \varphi_{G,1}(y)h(\omega - t | c)\varphi_{I,1}(x_1 - t)\varphi_{I,1}(x_2 - t - y), \\ B_2(t; x_1, x_2, c, \omega, y) &:= \varphi_{G,2}(y)h(\omega - t | c)\varphi_{I,2}(x_1 - t)\varphi_{I,2}(x_2 - t - y), \\ A_1(y; x_1, x_2, c, \omega) &:= \int_0^{\min(x_1, \omega)} B_1(t; x_1, x_2, c, y) \, dt, \\ A_2(y; x_1, x_2, c, \omega) &:= \int_0^{\min(x_1, \omega)} B_2(t; x_1, x_2, c, y) \, dt. \end{aligned}$$

We have now by the reverse triangle inequality for  $L^{1/\alpha}$  that for any  $x_1, x_2, \omega \geq 0$

$$\begin{aligned} &\left| \left( \int_0^{x_2} A_1(y; x_1, x_2, c, \omega) \, dy \right)^\alpha - \left( \int_0^{x_2} A_2(y; x_1, x_2, c, \omega) \, dy \right)^\alpha \right|^{\frac{1}{\alpha}} \\ &= \left| \|A_1(\cdot; x_1, x_2, c, \omega)^\alpha\|_{L^{1/\alpha}([0, x_2])} - \|A_2(\cdot; x_1, x_2, c, \omega)^\alpha\|_{L^{1/\alpha}([0, x_2])} \right|^{\frac{1}{\alpha}} \\ &\leq \|A_1(\cdot; x_1, x_2, c, \omega)^\alpha - A_2(\cdot; x_1, x_2, c, \omega)^\alpha\|_{L^{1/\alpha}([0, x_2])}^{\frac{1}{\alpha}} \\ &= \int_0^{x_2} |A_1(y; x_1, x_2, c, \omega)^\alpha - A_2(y; x_1, x_2, c, \omega)^\alpha|^{\frac{1}{\alpha}} \, dy \\ &= \int_0^{x_2} \left| \left( \int_0^{\min(x_1, \omega)} B_1(t; x_1, x_2, c, \omega, y) \, dt \right)^\alpha - \left( \int_0^{\min(x_1, \omega)} B_2(t; x_1, x_2, c, \omega, y) \, dt \right)^\alpha \right|^{\frac{1}{\alpha}} \, dy \\ &= \int_0^{x_2} \left\| \|B_1(\cdot; x_1, x_2, c, \omega, y)^\alpha\|_{L^{1/\alpha}([0, \min(x_1, \omega)])} \right. \\ &\quad \left. - \|B_2(\cdot; x_1, x_2, c, \omega, y)^\alpha\|_{L^{1/\alpha}([0, \min(x_1, \omega)])} \right|^{\frac{1}{\alpha}} \, dy \\ &\leq \int_0^{x_2} \|B_1(\cdot; x_1, x_2, c, \omega, y)^\alpha - B_2(\cdot; x_1, x_2, c, \omega, y)^\alpha\|_{L^{1/\alpha}([0, \min(x_1, \omega)])}^{\frac{1}{\alpha}} \, dy \\ &= \int_0^{x_2} \int_0^{\min(x_1, \omega)} |(\varphi_{G,1}(y)h(\omega - t | c)\varphi_{I,1}(x_1 - t)\varphi_{I,1}(x_2 - t - y))^\alpha \\ &\quad - (\varphi_{G,2}(y)h(\omega - t | c)\varphi_{I,2}(x_1 - t)\varphi_{I,2}(x_2 - t - y))^\alpha|^{\frac{1}{\alpha}} \, dt \, dy. \end{aligned}$$

By using the above inequality chain, we obtain

$$\begin{aligned}
& \sum_{c=1}^{K_C} \int_{[0,\infty)^3} |f_1(x_1, x_2, \omega | c)^\alpha - f_2(x_1, x_2, \omega | c)^\alpha|^{\frac{1}{\alpha}} p_C(c) \, d(x_1, x_2, \omega) \\
&= \sum_{c=1}^{K_C} \int_{[0,\infty)^3} \varphi_W(\omega) n(\omega, c) p_C(c) \\
&\quad \times \left| \left( \int_0^{x_2} \varphi_{G,1}(y) \int_0^{\min(x_1, \omega)} h(\omega - t | c) \varphi_{I,1}(x_1 - t) \varphi_{I,1}(x_2 - t - y) \, dt \, dy \right)^\alpha \right. \\
&\quad \left. - \left( \int_0^{x_2} \varphi_{G,2}(y) \int_0^{\min(x_1, \omega)} h(\omega - t | c) \varphi_{I,2}(x_1 - t) \varphi_{I,2}(x_2 - t - y) \, dt \, dy \right)^\alpha \right|^{\frac{1}{\alpha}} d(x_1, x_2, \omega) \\
&= \sum_{c=1}^{K_C} \int_{[0,\infty)^3} \varphi_W(\omega) n(\omega, c) p_C(c) \\
&\quad \times \left| \left( \int_0^{x_2} A_1(y; x_1, x_2, c, \omega) \, dy \right)^\alpha - \left( \int_0^{x_2} A_2(y; x_1, x_2, c, \omega) \, dy \right)^\alpha \right|^{\frac{1}{\alpha}} d(x_1, x_2, \omega) \\
&\leq \sum_{c=1}^{K_C} \int_{[0,\infty)^3} \varphi_W(\omega) n(\omega, c) p_C(c) \\
&\quad \times \int_0^{x_2} \int_0^{\min(x_1, \omega)} |(\varphi_{G,1}(y) h(\omega - t | c) \varphi_{I,1}(x_1 - t) \varphi_{I,1}(x_2 - t - y))^\alpha \\
&\quad - (\varphi_{G,2}(y) h(\omega - t | c) \varphi_{I,2}(x_1 - t) \varphi_{I,2}(x_2 - t - y))^\alpha|^{\frac{1}{\alpha}} \, dt \, dy \, d(x_1, x_2, \omega). \tag{A5}
\end{aligned}$$

Above we have an iterated integral over a non-negative function, we may thus re-arrange the order of integration and use substitution. We substitute below  $a = x_1 - t$  for  $x_1$  and  $b = x_2 - t - y$  for  $x_2$ . Note that we implicitly take care of the integration bounds by using the indicator function and the fact that all densities are zero on the negative real line. Hence, we can continue the above inequality chain

$$\begin{aligned}
\text{(A5)} &= \sum_{c=1}^{K_C} \int_{[0,\infty)^5} \mathbb{1}(y \leq x_2, t \leq \min(x_1, \omega)) \varphi_W(\omega) n(\omega, c) p_C(c) h(\omega - t | c) \\
&\quad \times |(\varphi_{G,1}(y) \varphi_{I,1}(x_1 - t) \varphi_{I,1}(x_2 - t - y))^\alpha \\
&\quad - (\varphi_{G,2}(y) \varphi_{I,2}(x_1 - t) \varphi_{I,2}(x_2 - t - y))^\alpha|^{\frac{1}{\alpha}} \, dx_1 \, dx_2 \, dt \, dy \, d\omega \\
&= \sum_{c=1}^{K_C} \int_{[0,\infty)^5} \mathbb{1}(y \leq x_2, t \leq \min(a + t, \omega)) \varphi_W(\omega) n(\omega, c) p_C(c) h(\omega - t | c) \\
&\quad \times |(\varphi_{G,1}(y) \varphi_{I,1}(a) \varphi_{I,1}(x_2 - t - y))^\alpha \\
&\quad - (\varphi_{G,2}(y) \varphi_{I,2}(a) \varphi_{I,2}(x_2 - t - y))^\alpha|^{\frac{1}{\alpha}} \, da \, dx_2 \, dt \, dy \, d\omega \\
&= \sum_{c=1}^{K_C} \int_{[0,\infty)^5} \mathbb{1}(0 \leq b + t, t \leq \min(a + t, \omega)) \varphi_W(\omega) n(\omega, c) p_C(c) h(\omega - t | c) \\
&\quad \times |(\varphi_{G,1}(y) \varphi_{I,1}(a) \varphi_{I,1}(b))^\alpha - (\varphi_{G,2}(y) \varphi_{I,2}(a) \varphi_{I,2}(b))^\alpha|^{\frac{1}{\alpha}} \, da \, db \, dt \, dy \, d\omega. \tag{A6}
\end{aligned}$$

Note next that the indicator equals actually  $\mathbb{1}(t \in [0, \omega])$ . We again interchange the order of integration, to integrate with respect to  $t$  first and then with respect to  $\omega$ . By doing this, recalling the form of  $\varphi_{T_1}$  in Assumption (M), keeping in mind that  $|x^{2\alpha} - y^{2\alpha}| \leq |x - y|^{2\alpha}$  for all  $x, y \geq 0$  (since  $2\alpha \in (0, 1)$ ) and that  $(x - y)^2 \leq 2x^2 + 2y^2$  for all  $x, y$ , we continue

$$\begin{aligned}
 \text{(A6)} &= \sum_{c=1}^{K_C} \int_{[0, \infty)^5} \varphi_W(\omega) p_C(c) \varphi_{T_1}(t \mid \omega, c) dt d\omega \\
 &\quad \times \left| (\varphi_{G,1}(y) \varphi_{I,1}(a) \varphi_{I,1}(b))^\alpha - (\varphi_{G,2}(y) \varphi_{I,2}(a) \varphi_{I,2}(b))^\alpha \right|^{\frac{1}{\alpha}} da db dy \\
 &= \int_{[0, \infty)^3} \left| (\varphi_{G,1}(y) \varphi_{I,1}(a) \varphi_{I,1}(b))^\alpha - (\varphi_{G,2}(y) \varphi_{I,2}(a) \varphi_{I,2}(b))^\alpha \right|^{\frac{1}{\alpha}} da db dy \\
 &\leq \int_{[0, \infty)^3} \left| (\varphi_{G,1}(y) \varphi_{I,1}(a) \varphi_{I,1}(b))^{\frac{1}{2}} - (\varphi_{G,2}(y) \varphi_{I,2}(a) \varphi_{I,2}(b))^{\frac{1}{2}} \right|^2 da db dy \\
 &\leq \int_{[0, \infty)^3} 2 \left( \varphi_{G,1}(y)^{\frac{1}{2}} - \varphi_{G,2}(y)^{\frac{1}{2}} \right)^2 \varphi_{I,1}(a) \varphi_{I,1}(b) da db dy \\
 &\quad + 2 \int_{[0, \infty)^3} \varphi_{G,2}(y) \left( \varphi_{I,2}(a)^{\frac{1}{2}} \varphi_{I,2}(b)^{\frac{1}{2}} - \varphi_{I,1}(a)^{\frac{1}{2}} \varphi_{I,1}(b)^{\frac{1}{2}} \right)^2 da db dy \\
 &= 2\rho_H(\varphi_{G,1}, \varphi_{G,2})^2 + 2 \int_{[0, \infty)^2} \left( \varphi_{I,2}(a)^{\frac{1}{2}} \varphi_{I,2}(b)^{\frac{1}{2}} - \varphi_{I,1}(a)^{\frac{1}{2}} \varphi_{I,1}(b)^{\frac{1}{2}} \right)^2 da db \\
 &\leq 2\rho_H(\varphi_{G,1}, \varphi_{G,2})^2 + 4 \int_{[0, \infty)^2} \left( \varphi_{I,2}(a)^{\frac{1}{2}} - \varphi_{I,1}(a)^{\frac{1}{2}} \right)^2 \varphi_{I,2}(b) db \\
 &\quad + 4 \int_{[0, \infty)^2} \varphi_{I,1}(a) \left( \varphi_{I,1}(b)^{\frac{1}{2}} - \varphi_{I,2}(b)^{\frac{1}{2}} \right)^2 da db \\
 &= 2\rho_H(\varphi_{G,1}, \varphi_{G,2})^2 + 8\rho_H(\varphi_{I,1}, \varphi_{I,2})^2. \tag{A7}
 \end{aligned}$$

This is (A4) and the proof is complete. ■

**Proof of Lemma 4.8:** Denote for any  $m \in \mathbb{N}_0$ ,  $\Theta_m := \{\theta \in \mathbb{R}^{m+1} : \|\theta\|_2 = 1\}$ . The proof of this Lemma uses the following strategy which is similar to Lemma 2.1 in Ossiander (1987). In the interest of completeness we give the detailed proof: Let  $\delta \in (0, \sqrt{3/2}/2]$  be given. We define for any  $\theta_1 \in \Theta_{m_1}$  and  $\theta_2 \in \Theta_{m_2}$  the  $\delta$ -ball

$$B_\delta(\theta_1, \theta_2) := \{(\tilde{\theta}_1, \tilde{\theta}_2) \in \Theta_{m_1} \times \Theta_{m_2} : \|\theta_1 - \tilde{\theta}_1\|_2 \leq \delta, \|\theta_2 - \tilde{\theta}_2\|_2 \leq \delta\}.$$

Find now a set  $((\theta_{1,i}, \theta_{2,i}))_{i=1, \dots, N(\delta)} \subseteq \Theta_{m_1} \times \Theta_{m_2}$  such that

$$\bigcup_{i=1}^{N(\delta)} B_\delta(\theta_{1,i}, \theta_{2,i}) \supseteq \Theta_{m_1} \times \Theta_{m_2}.$$

In order to bound  $N(\delta)$  we construct a specific collection of pairs: Consider a grid of  $[0, \pi]^{m_1}$  with side length  $\alpha = 2\delta\sqrt{2/3}$  and construct  $\theta_{1,i}$  by taking the grid points as polar coordinates (with radius 1). Then, it is clear that for any  $\theta \in \Theta_{m_1}$ , there is a grid point  $\theta_{1,i}$  such that the difference between any two angles of the polar representations of  $\theta$  and  $\theta_{1,i}$  is smaller than  $\alpha/2 = \delta\sqrt{2/3}$ . By Lemma A.3 below and symmetry of the polar coordinates, we find that  $\|\theta - \theta_{1,i}\|_2 \leq \sqrt{3/2}\alpha/2 = \delta$ . The size of this grid can be bounded by  $(\pi/\alpha + 1)^{m_1}$ . We repeat this construction for  $\Theta_{m_2}$  and obtain

$$N(\delta) \leq \left( \frac{\sqrt{3}\pi}{2\sqrt{2}\delta} + 1 \right)^{m_1+m_2}. \tag{A8}$$

The brackets are now defined as

$$l_i(x_1, x_2, \omega, c) := \inf_{(\theta_1, \theta_2) \in B_\delta(\theta_{1,i}, \theta_{2,i})} f_{\varphi_{I,\theta_1}, \varphi_{G,\theta_2}}(x_1, x_2, \omega \mid c) p_C(c),$$

$$u_i(x_1, x_2, \omega, c) := \sup_{(\theta_1, \theta_2) \in B_\delta(\theta_{1,i}, \theta_{2,i})} f_{\varphi_{I,\theta_1}, \varphi_{G,\theta_2}}(x_1, x_2, \omega \mid c) p_C(c).$$

For any  $f_{\varphi_{I,\theta_1}, \varphi_{G,\theta_2}} p_C \in \mathcal{F}_{m_1, m_2}$  we find thus first a pair  $(\theta_{1,i}, \theta_{2,i})$  such that  $(\theta_1, \theta_2) \in B_\delta(\theta_{1,i}, \theta_{2,i})$  and thus also  $l_i \leq f_{\varphi_{I,\theta_1}, \varphi_{G,\theta_2}} p_C \leq u_i$ . It remains to compute  $\rho_H(l_i, u_i)$ . To this end, we firstly see that the same arguments which lead to (A7) (for  $\alpha = 1/2$ ) give us here the following (the sup refers always to the supremum over all pairs  $(\theta_1, \theta_2), (\tilde{\theta}_1, \tilde{\theta}_2) \in \Theta_{m_1} \times \Theta_{m_2}$  such that  $\|\theta_1 - \tilde{\theta}_1\|_2 \leq \delta$  and  $\|\theta_2 - \tilde{\theta}_2\|_2 \leq \delta$ )

$$\begin{aligned} \rho_H(l_i, u_i)^2 &\leq \sum_{c=1}^{K_C} \int_{[0, \infty)^3} \sup \left( f_{\varphi_{I,\theta_1}, \varphi_{G,\theta_2}}(x_1, x_2, \omega)^{\frac{1}{2}} - f_{\varphi_{I,\tilde{\theta}_1}, \varphi_{G,\tilde{\theta}_2}}(x_1, x_2, \omega)^{\frac{1}{2}} \right)^2 p_C(c) d(x_1, x_2, \omega) \\ &\leq \int_{[0, \infty)^3} \sup \left( (\varphi_{G,\theta_1}(y) \varphi_{I,\theta_2}(a) \varphi_{I,\theta_2}(b))^{\frac{1}{2}} - (\varphi_{G,\tilde{\theta}_1}(y) \varphi_{I,\tilde{\theta}_2}(a) \varphi_{I,\tilde{\theta}_2}(b))^{\frac{1}{2}} \right)^2 dy da db \\ &\leq \int_{[0, \infty)^3} e^{-y-a-b} \sup \left( \sum_{k=0}^{m_1} (\theta_{1,k} - \tilde{\theta}_{1,k}) L_k(y) \sum_{k=0}^{m_2} \theta_{2,k} L_k(a) \sum_{k=0}^{m_2} \theta_{2,k} L_k(b) \right. \\ &\quad \left. - \sum_{k=0}^{m_1} \tilde{\theta}_{1,k} L_k(y) \left( \sum_{k=0}^{m_2} (\tilde{\theta}_{2,k} - \theta_{2,k}) L_k(a) \sum_{k=0}^{m_2} \tilde{\theta}_{2,k} L_k(b) \right. \right. \\ &\quad \left. \left. - \sum_{k=0}^{m_2} \theta_{2,k} L_k(a) \sum_{k=0}^{m_2} (\theta_{2,k} - \tilde{\theta}_{2,k}) L_k(b) \right) \right)^2 dy da db \\ &\leq \int_{[0, \infty)^3} e^{-y-a-b} \left( 2\delta^2 \sum_{k=0}^{m_1} L_k(y)^2 \sum_{k=0}^{m_2} L_k(a)^2 \sum_{k=0}^{m_2} L_k(b)^2 \right. \\ &\quad \left. + 2 \sum_{k=0}^{m_1} L_k(y)^2 \left( 2\delta^2 \sum_{k=0}^{m_2} L_k(a)^2 \sum_{k=0}^{m_2} L_k(b)^2 + 2\delta^2 \sum_{k=0}^{m_2} L_k(a)^2 \sum_{k=0}^{m_2} L_k(b)^2 \right) \right) dy da db \\ &= 10\delta^2 m_1 m_2^2, \end{aligned}$$

where we used the Cauchy-Schwarz-Inequality in between and the integral properties of the Laguerre polynomials at the end. Thus, when putting  $\delta = \varepsilon(10m_1 m_2^2)^{-1/2}$  (the condition on  $\delta$  is fulfilled by the assumption on  $\varepsilon$ ) we find together with (A8)

$$\mathcal{N}_{[\cdot]}(\varepsilon, \mathcal{F}_{m_1, m_2}, \rho_H) \leq \left( \frac{\pi \sqrt{15m_1 m_2^2}}{2\varepsilon} + 1 \right)^{m_1 + m_2} \leq \left( \frac{\pi \sqrt{15m_1 m_2^2}}{\varepsilon} \right)^{m_1 + m_2}$$

and the proof is complete. ■

**Lemma A.3:** Let  $n \geq 2$  and  $\delta \in [0, 1/2]$ . Denote by  $e_1 := (1, 0, \dots, 0)'$  the first unit vector of  $\mathbb{R}^n$ . Let  $x \in \mathbb{R}^n$  have  $\|x\|_2 = 1$  and angles of polar coordinates  $\psi_1, \dots, \psi_{n-1} \in [0, \delta]$ . Then,  $\|e_1 - x\|_2 \leq \delta \sqrt{3/2}$ .

**Proof:** By a Taylor expansion, we have for any  $\psi \in [0, 2\pi]$  (below  $\psi^*$  denotes different intermediate values between 0 and  $\psi$ )

$$|\sin(\psi)| = |\cos(\psi^*)| \cdot |\psi| \leq |\psi|$$

$$|\cos(\psi) - 1| = |\sin(\psi^*)| \cdot |\psi| \leq |\psi|^2.$$



By the definition of polar coordinates, we compute (for  $n = 2$  the sum disappears)

$$\begin{aligned} \|e_1 - x\|_2^2 &= (1 - \cos(\psi_1))^2 + \prod_{k=1}^{n-1} \sin(\psi_k)^2 + \sum_{i=2}^{n-1} \prod_{k=1}^{i-1} \sin(\psi_k)^2 \cos(\psi_i)^2 \\ &\leq \psi_1^4 + \prod_{k=1}^{n-1} \psi_k^2 + \sum_{i=2}^{n-1} \prod_{k=1}^{i-1} \psi_k^2 \leq \delta^4 + \delta^{2(n-1)} + \sum_{i=2}^{n-1} \delta^{2(i-1)} \leq \delta^2 \frac{\delta^2 - \delta^4 + 1}{1 - \delta^2}. \end{aligned}$$

The statement follows since  $(\delta^2 - \delta^4 + 1)/(1 - \delta^2) \leq 3/2$  for  $\delta \in [0, 1/2]$ . ■