BIOMETRICS 000, 000–000 DOI: 000
000 0000

Web Based Supplementary Materials for 'An Empirically Adjusted Approach to Reproductive Number Estimation for Stochastic Compartmental Models'

Grant D. Brown[1,*], Jacob J. Oleson[1], and Aaron T. Porter[2]

[1]Department of Biostatistics, University of Iowa, Iowa City, Iowa, U.S.A.

[2]Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, U.S.A.

[*]*email:* grant-brown@uiowa.edu

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

This document provides additional information on statistical derivations and analysis output which may be of interest to readers of the main manuscript, “An Empirically Adjusted Approach to Reproductive Number Estimation for Stochastic Compartmental Models”. Section 1 contains details on two important derivations. First, we explore the parametric from of the exposure probability which arises due to the motivating assumptions discussed in Section 2.4 of the manuscript. Next, we further explore the nature of the relationship between our $\mathcal{R}^{(EA)}$ method and traditional measures of $\mathcal{R}_0$ discussed in Section 3.3. Section 2 contains addional statistical output for the three degree of freedom Ebola analysis discussed in Section 5.2, including a table of MCMC parameter quantiles, and an illustration of the employed temporal spline basis.

2. Important Derivations

2.1 *Exposure Probability: Parametric Form*

Define δ_{ij} to be the proportion of persons who are infectious in spatial unit s_j at time t_i. Then, letting $Inf(s_j, t_i)$ denote the event that a person becomes infected from contact within spatial unit s_j at time t_i and $!Inf(s_j, t_i)$ denote it's compliment. We may then partition the infection probability as in 1.

$$P(Inf(., t_i)) = 1 - P(!Inf(s_j, t_i)) \cdot P(!Inf(s_{-j}, t_i)) \tag{1}$$

where

$$P(!Inf(s_j,t_i)) = E(!Inf(s_j,t_i)) = E(E(!Inf(s_j,t_i)|K=k))$$
$$= E((1-\delta_{ij}p)^k)$$
$$= \sum_{k=0}^{\infty}(1-\delta_{ij}p)^k(\frac{\lambda_{ij}^k e^{-\lambda_{ij}}}{k!})$$
$$= \sum_{k=0}^{\infty} q_{ij}^k(\frac{\lambda_{ij}^k e^{-\lambda_{ij}}}{k!})$$
$$= \frac{e^{-\lambda_{ij}}}{e^{-q_{ij}\lambda_{ij}}}(1) = e^{-\lambda_i\cdot(1-q_{ij})} = e^{-\lambda_i\cdot p_{ij}} = e^{-\lambda_i\cdot(\delta_{ij}p)}$$

Therefore,

$$P(Inf(s_j,t_i)) = 1 - e^{-\lambda_{ij}\cdot(\delta_{ij}p)}$$

Similarly,

$$P(!Inf(s_{-j},t_i)) = \prod_{\{l\neq j\}}[P(!Inf(s_l,t_i))]$$
$$= \prod_{\{l\neq j\}}[E(!Inf(s_{-j},t_i))] = \prod_{\{l\neq j\}}[E(E(!Inf(s_{-j},t_i)|K=k))]$$
$$= \prod_{\{l\neq j\}}\left[E(1-\delta_{il}p)^k)\right]$$
$$= \prod_{\{l\neq j\}}\left[\sum_{k=0}^{\infty}(q_{il}(j))^k\frac{(\lambda_{il}\cdot f(d_{jl}))^k e^{-\lambda_{il}\cdot f(d_{jl})}}{k!}\right] = \prod_{\{l\neq j\}}\left[\frac{e^{-\lambda_{il}\cdot f(d_{jl})}}{e^{-q_{il}\lambda_{il}f(d_{jl})}}(1)\right]$$
$$= \prod_{\{l\neq j\}}\left[e^{-\lambda_{il}\cdot f(d_{jl})p_{il}}\right] = \prod_{\{l\neq j\}}\left[e^{-\lambda_{il}\cdot f(d_{jl})\cdot(\delta_{il}p)}\right]$$
$$= exp\left\{\sum_{\{l\neq j\}}[p\lambda_l\delta_{il}f(d_{jl})]\right\}$$

Thus, for the probability of infection for a person living in s_j at time t_i we have:

$$1 - \left(e^{-\lambda_{ij}\cdot(\delta_{ij}p)}\right)\left(e^{\left\{\sum_{\{l\neq j\}}\left[p\lambda_{il}\delta_{il}f(d_{jl})\right]\right\}}\right)$$
$$= 1 - exp\left\{-\delta_{ij}e^{\theta_{ij}} - \sum_{\{l\neq j\}}(f(d_{jl})\delta_{il}e^{\theta_{il}})\right\}$$

where $\theta_{ij} = log(\lambda_{ij}p)$

2.2 *$\mathcal{R}_0$ as a special case of $\mathcal{R}^{(\mathcal{EA})}$*

$$\mathcal{R}^{(EA)}(t_i) = \sum_{t=t_i}^{t_\infty}\left(\frac{N}{I_t}\right)(1 - exp\{-\frac{I_t}{N}e^{\theta}\})(1 - \pi^{(IR)})^{(t-t_i)}.$$

Then, assuming that I_t remains equal to one long enough that the remaining terms in this infinite summation are negligible, we have

$$\mathcal{R}^{(EA)}(t_i) \approx \sum_{t=t_i}^{t_\infty}\left(\frac{N}{1}\right)(1 - exp\{-\frac{1}{N}e^{\theta}\})(1 - \pi^{(IR)})^{(t-t_i)}.$$
$$= \left(\frac{N}{1}\right)(1 - exp\{-\frac{1}{N}e^{\theta}\})\sum_{t=t_i}^{t_\infty}(1 - \pi^{(IR)})^{(t-t_i)}.$$
$$= \left[\frac{(1 - exp\{-\frac{1}{N}e^{\theta}\})}{\left(\frac{1}{N}\right)}\right]\left[\sum_{t=t_i}^{t_\infty}(1 - \pi^{(IR)})^{(t-t_i)}\right].$$

Note that, taking the limit as $(N \to \infty)$, this first term is indeterminate: $\frac{0}{0}$. Thus, with an application of L'Hospital's rule we have the limit expression

$$lim_{N\to\infty}\frac{-\frac{1}{N^2}e^{\theta}exp\{-\frac{1}{N}e^{\theta}\}}{-\frac{1}{N^2}}\left[\sum_{t=t_i}^{t_\infty}(1 - \pi^{(IR)})^{(t-t_i)}\right].$$
$$= e^{\theta}\left[\sum_{t=t_i}^{t_\infty}(1 - \pi^{(IR)})^{(t-t_i)}\right].$$

Finally, by the convergence property of the geometric series we have

$$\frac{e^{\theta}}{1 - (1 - \pi^{(IR)})} = \frac{e^{\theta}}{\pi^{(IR)}}$$

3. Additional statistical output

[Table 1 about here.]

[Figure 1 about here.]

Figure 1. Three Degree of Freedom Spline Basis for Estimation and Prediction

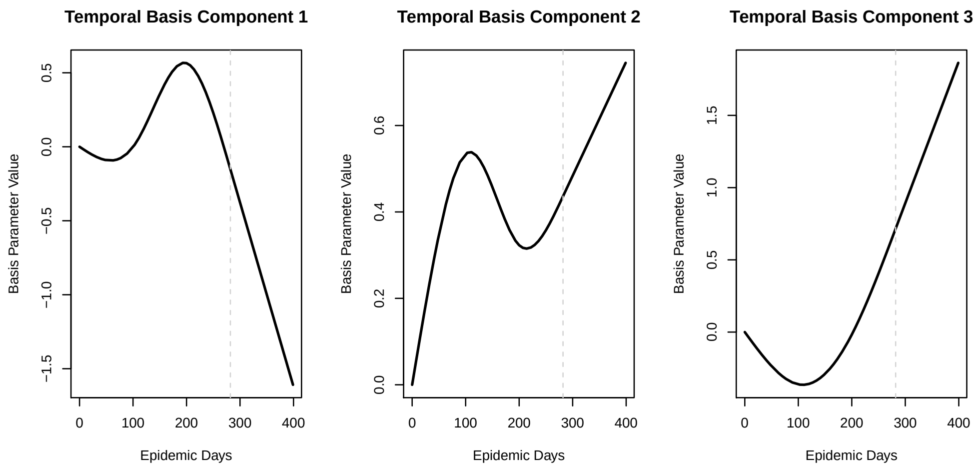

Table 1
MCMC Quantiles for Three Degree of Freedom Model Parameters

	2.5%	25%	50%	75%	97.5%
Guinea Intercept	-3.20	-3.07	-3.00	-2.93	-2.81
Liberia Intercept	-2.60	-2.44	-2.36	-2.28	-2.11
Sierra Leone Intercept	-2.32	-2.16	-2.08	-2.00	-1.84
Time component 1	-0.47	-0.36	-0.31	-0.25	-0.14
Time component 2	-0.52	-0.18	-0.02	0.15	0.46
Time component 3	-1.24	-1.17	-1.13	-1.09	-1.03
Guinea-Liberia Spatial Component	0.01	0.02	0.02	0.02	0.03
Guinea-Sierra Leone Spatial Component	0.06	0.07	0.07	0.08	0.08
Liberia-Sierra Leone Spatial Component	0.02	0.03	0.04	0.05	0.06
Overdispersion Precision	6.54	6.62	6.67	6.71	6.80
E to I Transition Parameter	0.19	0.19	0.19	0.19	0.20
I to R Transition Parameter	0.13	0.14	0.14	0.14	0.14
E to I Transition Probability	0.17	0.17	0.17	0.18	0.18
I to R Transition Probability	0.13	0.13	0.13	0.13	0.13
Days in Exposed Category	0.00	1.00	3.00	7.00	19.00
Days in Infectious Category	0.00	2.00	4.00	9.00	26.00