Bayesian Inference for Infectious Disease Transmission Models Based on Ordinary Differential Equations

Felix Weidemann

München 2015

Bayesian Inference for Infectious Disease Transmission Models Based on Ordinary Differential Equations

Felix Weidemann

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Felix Weidemann
aus Berlin

München, den 31. Juli 2015

Danksagung

Zusammenfassung

Die Vorhersage der möglichen epidemiologischen Effekte von neuen Impfprogramme durch mathematisch-statistische Transmissionsmodellierung ist von zunehmender Bedeutung für die Ständige Impfkommission in Deutschland. Solche Modelle erfassen in der Regel sehr große Populationen mittels Kompartmentalisierung, wobei die Flüsse zwischen den Kompartmenten durch ein System gewöhnlicher Differentialgleichungen (DGL) beschrieben werden. Die numerische Lösung dieser DGL-Systeme ist jedoch häufig mit hohem Rechenaufwand verbunden. Für sämtliche statistische Verfahren zur Schätzung zugehöriger Modellparameter anhand von Daten zur Krankheitslast stellt dies eine große Herausforderung dar. In der Praxis werden daher viele Parameter basierend auf epidemiologischen Studien fixiert, wodurch jegliche Parameterunsicherheit im Vorhinein ausgeschlossen wird. Eine vielversprechende Alternative wäre dagegen ein Bayesianisches Inferenzverfahren, welches die vorhandenen epidemiologischen Kenntnisse bei der Schätzung berücksichtigen und gleichzeitig mehr Parameterunsicherheit zulassen würde.

Diese Arbeit beschäftigt sich mit statistischen Methoden zur Bayesianischen Inferenz von DGL-basierten Modellen. Ein Ansatz zur Approximation der Posteriori-Verteilung mittels einer Gauß-Verteilung basierend auf dem Posteriori-Modus und der beobachteten Fisher-Information wird vorgestellt. Unter Anwendung einer neu entwickelten Methode zur Reskalierung der Likelihood in Form einer Power-Posteriori ist es möglich die nach der Modellanpassung verbleibende Autokorrelation in den Daten für die Unsicherheitsschätzung mit zu berücksichtigen. Als eine Alternative zum Gaußschen Approximierungsansatz wird eine adaptiver Metropolis-Hastings Algorithmus vorgestellt, welcher insbesondere auf das effiziente Sampling aus hoch-dimensionalen Posteriori-Verteilungen mit starker Parameter-Kollinearität abzielt. Zur Identifikation der wichtigsten Modell-Komponenten werden Bayesianische Modell-Selektionskriterien basierend auf der marginalen Likelihood der Daten verwendet. Die Schätzung der marginalen Likelihood erfolgt dabei mit einem neu entwickelten Ansatz unter Ausnutzung des vorhandenen Posteriori-Samples aus dem vorhergehenden Metropolis-Hastings Verfahren.

Weiterhin beinhaltet diese Arbeit eine Anwendung der vorgestellten Methoden, indem die epidemiologischen Effekte einer möglichen Rotavirus-Kinderimpfung in Deutschland vorhergesagt werden. Dazu wird ein DGL-basiertes Kompartment-Modell entwickelt, dessen Dynamik die wichtigsten Aspekte der Rotavirus-Transmission abbildet. Durch eine Erweiterung des Modells um zusätzliche Impfmechanismen ist es außerdem möglich die

Rotavirus-Impfeffektivität anhand von routinemäßig gesammelten Surveillance Daten zu schätzen. Durch die Anwendung des Bayesianischen Verfahrens werden bei der Prognose der epidemiogischen Entwicklungen infolge hoher Durchimpfungsraten die Unsicherheit hinsichtlich Modellstruktur und -Parameter mitberücksichtigt. Die Modellergebnisse sagen eine leichte Inzidenzzunahme bei älteren Kindern und Senioren voraus. Dagegen sinkt aufgrund der Routineimpfung die Krankheitslast innerhalb der Fokusgruppe der jungen Kinder drastisch, mittels Herdeneffekten sogar stärker als durch direkte Impfeffekte allein zu erwarten wäre.

Zur besseren Unterstützung von Entscheidungsfindungen unter Unsicherheit präsentiert diese Arbeit eine statistische Sichtweise auf die Modellierung der Effekte einer Routineimpfung. Darüber hinaus sind die hier vorgestellten Methoden ebenso anwendbar für die Transmissionsmodellierung von anderen Erregern wie zum Beispiel Influenza.

Abstract

Predicting the epidemiological effects of new vaccination programmes through mathematical-statistical transmission modelling is of increasing importance for the German Standing Committee on Vaccination. Such models commonly capture large populations utilizing a compartmental structure with its dynamics being governed by a system of ordinary differential equations (ODEs). Unfortunately, these ODE-based models are generally computationally expensive to solve, which poses a challenge for any statistical procedure inferring corresponding model parameters from disease surveillance data. Thus, in practice parameters are often fixed based on epidemiological knowledge hence ignoring uncertainty. A Bayesian inference framework incorporating this prior knowledge promises to be a more suitable approach allowing for additional parameter flexibility.

This thesis is concerned with statistical methods for performing Bayesian inference of ODE-based models. A posterior approximation approach based on a Gaussian distribution around the posterior mode through its respective observed Fisher information is presented. By employing a newly proposed method for adjusting the likelihood impact in terms of using a power posterior, the approximation procedure is able to account for the residual autocorrelation in the data given the model. As an alternative to this approximation approach, an adaptive Metropolis-Hastings algorithm is described which is geared towards an efficient posterior sampling in the case of a high-dimensional parameter space and considerable parameter collinearities. In order to identify relevant model components, Bayesian model selection criteria based on the marginal likelihood of the data are applied. The estimation of the marginal likelihood for each considered model is performed via a newly proposed approach which utilizes the available posterior sample obtained from the preceding Metropolis-Hastings algorithm.

Furthermore, the thesis contains an application of the presented methods by predicting the epidemiological effects of introducing rotavirus childhood vaccination in Germany. Again, an ODE-based compartmental model accounting for the most relevant transmission aspects of rotavirus is presented. After extending the model with vaccination mechanisms, it becomes possible to estimate the rotavirus vaccine effectiveness through routinely collected surveillance data. By employing the Bayesian framework, model predictions on the future epidemiological development assuming a high vaccination coverage rate incorporate uncertainty regarding both model structure and parameters. The forecast suggests that routine vaccination may cause a rotavirus incidence increase among older children and el-

derly, but drastically reduces the disease burden among the target group of young children, even beyond the expected direct vaccination effect by means of herd protection.

Altogether, this thesis provides a statistical perspective on the modelling of routine vaccination effects in order to assist decision making under uncertainty. The presented methodology is thereby easily applicable to other infectious diseases such as influenza.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Infectious diseases have always posed a constant threat to human health worldwide. Re-emerging pathogens – like the pandemic influenza in 2009, the EHEC outbreak in Middle-Europe in 2011 or the ongoing Ebola epidemic in West-Africa – attract a lot of public attention through the media. Beyond that, each year everyday diseases such as the seasonal influenza cause many severe cases, deaths and even more sick patients unable to work both in developing and industrialized countries. This yields a permanent clinical and financial burden on the national healthcare systems. Understanding of adequate control measures is thus required in order to prevent large disease outbreaks in the future and to reduce the incidence of endemic disease.

This thesis deals with the mathematical-statistical modelling of the transmission of infectious diseases and with the analysis of vaccination-based intervention methods for the containment of such infections. For this purpose we will consider models based on ordinary differential equations (ODEs), which constitute the most commonly applied model class in infectious diseases epidemiology. We will restrict our analysis to human-to-human transmittable infections, in contrast to zoonoses like salmonellosis or vector-borne diseases such as malaria. Besides presenting suitable statistical methods we will put an emphasis on the inherent uncertainty arising when inferring model parameters through disease incidence data as well as drawing epidemiological conclusions from the transmission model.

Our work is motivated by modelling the transmission of rotavirus in the German population. In 2011 the Standing Committee on Vaccination (STIKO) – which is the German decision-making body regarding the recommendation of new vaccines – was considering the introduction of childhood rotavirus vaccination via two vaccines licensed in 2006 (Rotarix® and Rotateq®) (Koch et al., 2013). To do so, the STIKO was interested in the potential epidemiological impact of a routine vaccination programme and thus the STIKO office located at the Robert Koch Institute initiated a Ph.D project on modelling the rotavirus transmission and vaccination impact, which led to the present thesis.

Hence, the aim of this thesis is to construct a suitable mathematical model for representing the rotavirus transmission dynamics within the German population. The model has to address the most relevant transmission aspects like demographic structure, age-dependent

contact behaviour, and rotavirus specific characteristics such as seasonality of the disease. Model parameters will be inferred within a Bayesian framework utilizing rotavirus case notification data according the nationwide mandatory surveillance system (Krause et al., 2007). Through the implementation of vaccination mechanisms into the model it is possible to predict the potential epidemiological effects of a vaccination programme while simultaneously accounting for uncertainty regarding parameter estimates and model structure. Thus, by considering a population-based transmission model we are able to assess not only the reduced infection risk for vaccinated individuals, but also the indirect dynamic effects in the non-vaccinated population due to an overall decreased force of infection. The main challenges of this approach include a careful assessment of parameter uncertainty in the face of dependent observations, the selection of model components and respective transmission aspects necessary to describe the observed disease pattern, and also suitable statistical algorithms when dealing with computationally complex models. Note that the methods presented in this thesis can also be applied to the transmission modelling of other infectious diseases or to any other ODE-based model.

Within this introduction we will first give a brief outline of the history and significance of disease transmission modelling. The subsequent section provides clinical and epidemiological information on rotavirus in Germany and also presents the available incidence data which will be utilized within our main application. Afterwards we will give an overview of the most prominent transmission model classes, including deterministic and stochastic models. Hereby we will also emphasize our choice of an ODE-based model complemented by a Bayesian inference framework for the epidemiological question at hand. An outline of the present thesis will conclude this introduction.

Why modelling disease transmission

The beginnings of modelling the transmission dynamics of an infectious disease is almost as old as the discovery of contagiousness of diseases itself. In the 19th century it was found that certain illnesses can be transferred through microorganisms from one person to another (Nelson and Williams, 2014). The first mathematical model capturing this transmission aspect was developed by William Hamer in 1906 describing the measles epidemic in London (Daley et al., 2001). The central idea of his model was that the contacts between infectious and healthy people would yield additional infected cases. In the early 20th century, first Ross and Hudson in 1917 and later Kermack and McKendrick in 1927 developed a corresponding continuous time model based on ODEs mimicking these interactions.

A first stochastic model for transmission dynamics was employed by Reed and Frost in 1928, who presented the occurrence of new cases as a chain-binomial process (Abbey, 1950). In mid-century Bartlett (1956) formulated a stochastic version of the deterministic model corresponding to Kermack and McKendrick (Bartlett, 1956). From then on the number of further variations for both deterministic and stochastic models grew steadily (Daley et al., 2001). However, the main goal for working with these models was to derive analytical properties of the modelled epidemic processes such as the final size of an epidemic or its equilibrium behaviour.

Due to the increasing computational power at the end of the 20th century, it became possible to compute or simulate model solutions numerically, which allowed not only a deeper quantitative analysis of model results but also more complexity in model-building. This even goes as far as to accounting for individual behaviour in agent-based models (Epstein, 2009). Thus, pathogen and population specific transmission aspects became more relevant for modelling and, in addition, the fitting of models to epidemiological data via computationally expensive statistical methods also became of increasing importance. This led to two main applications of mathematical-statistical modelling within modern infectious disease epidemiology.

Firstly, through mathematical modelling one is able to assess the importance of different epidemiological aspects within disease transmission. Thus, one can identify the key aspects with the largest effect on the spreading process. For instance, this might be the role of weather conditions in the spread of seasonal influenza as examined by Willem et al. (2012) or the network structure of sexual contacts in the case of HIV transmission, e.g. see Anderson et al. (1990). Moreover, through disease burden data it is possible to estimate crucial epidemiological model parameters, which might be otherwise difficult to assess within observational studies – like transmission probabilities as done in McBryde et al. (2008). In another example, Lunelli et al. (2013) estimated the background incidence of respiratory illnesses not attributable to influenza.

The second utilisation of transmission models is the scenario analysis of potential intervention measures. Since in the real world it is not feasible to test population wide interventions by trial and error, it is important to quantitatively assess their impact in advance based on transmission models. The primary example, which is also the topic of this thesis, is the evaluation of the epidemiological impact of a new vaccination strategy, e.g. Vynnycky et al. (2008) examined the consequences of implementing routine childhood vaccination against seasonal influenza in England and Wales. However, it is also possible to investigate the effects of alternative interventions measures such as school closures to dampen disease transmission, see e.g. Hens et al. (2009). For pandemic diseases one may be interested in the impact of even more drastic actions by imposing local or even global travel restrictions or border control, as analysed by Chong and Ying Zee (2012) and Wood et al. (2007) considering a pandemic influenza.

Such scenario analysis are crucial to support policy decision making. In particular regarding the recommendation of new vaccinations, many national immunization technical advisory Groups (NITAGs) nowadays require evidence for a beneficial impact suggested by mathematical models. The Joint Committee on Vaccination and Immunisation – the NITAG of the UK – even bases its vaccine recommendation decisions on results from mathematical models alone (Joint Committee on Vaccination and Immunisation, 2013). In Germany, the STIKO recently started to include model-based considerations into their recommendations (Koch et al., 2013; Ständige Impfkommission, 2014).

Due to this increasing importance of modelling results in health policy making, an accurate treatment of uncertainty concerning estimates and predictions from these models becomes indispensable. However, data on the modelled variables – such as the varying

number of infected people, their individual onset of symptoms and time point of recovery – are often not possible to acquire or are only partially available. For instance, disease surveillance systems only collect aggregated counts of new infections instead of information on the full transmission dynamics, and therefore available data for parameter inference or model validation is often incomplete (O'Neill, 2010). For an emerging epidemic, there is usually even no data available at all. Combined with the already complex nature of the model itself, this lack of data poses a statistical challenge which often requires customized inference tools for the specific model at hand.

1.1   Epidemiology of rotavirus in Germany

Rotavirus infection is worldwide the primary cause of acute gastroenteritis among young children. Almost every child has been infected with rotavirus at least once until the age of five years. It is estimated that rotavirus leads to annually more than 110 million episodes of diarrhoea causing 25 million clinic visits, 2 million hospitalizations, and 453,000 deaths (Tate et al., 2012).

In Germany, the number of rotavirus associated deaths is very low due to a better developed health care system. However, in many cases rotavirus infection still leads to vomiting, diarrhea, and severe dehydration requiring hospital admission. The resulting clinical and financial burden for the German health care system is substantial (Aidelsburger et al., 2014; Giaquinto et al., 2007). Since 2001, rotavirus infection is a notifiable disease in Germany according to the German Protection against Infection Act (Krause et al., 2007).
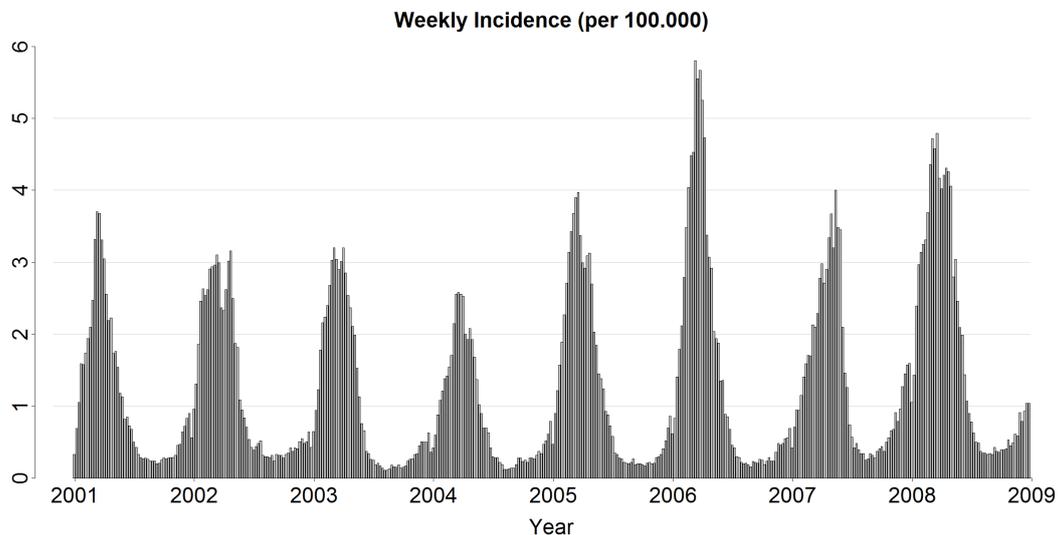
Figure 1.1: Reported rotavirus incidence in Germany according to the Infection Protection Act. Weekly incidence per 100,000 people.

Between 2001 and 2008 there were 442,199 reported cases of rotavirus infection, with

72% being detected among children under five years of age (Koch and Wiese-Posselt, 2011). With a yearly incidence of 115 cases per 10,000 children younger than 5 years rotavirus was the most frequently reported disease within this age group. Among adults and elderly above 65 years the yearly incidence is comparatively low with 0.5 and 2.1 cases per 10,000 people, respectively (Dudareva et al., 2012). Rotavirus incidence exhibits a clear seasonal trend with more than 80% of the cases occurring between January and May (see Figure 1.1). The yearly peak is usually reached in March, while the lowest incidence is reported in August. The reported incidence is considerably higher in the eastern federal states (see Figure 1.2), which is primarily caused by a different consultation seeking and diagnostic behaviour of parents and practitioners from the eastern states, respectively (Dudareva et al., 2012).
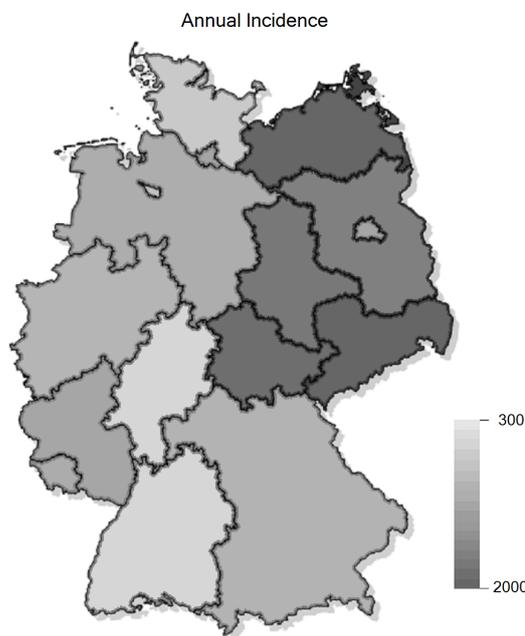
Figure 1.2: Mean annual rotavirus incidence stratified by federal state.

Data on rotavirus incidence is available as aggregated case notification counts from the German disease surveillance system *Survnet* established at the Robert Koch Institute (Krause et al., 2007). The case counts can be stratified by age, federal state, week of reporting, sex, and hospitalization status. However, further individual case information on e.g. contacts or duration of illness or infection is not available in the data.

In 2006, two live rotavirus vaccines, RotaTeq® (Merck & Co) and Rotarix® (GlaxoSmithKline), were licensed for use in Europe. Both vaccines are orally administered as a two or three-dose series within the period from 6 to 32 weeks of age. In clinical trials both vaccines demonstrated a good safety profile and high efficacy to protect infants against severe rotavirus infection (Ruiz-Palacios et al., 2006; Vesikari et al., 2006, 2007; Block et al., 2007).

In July 2013, STIKO decided to adopt rotavirus vaccination into the national vaccination schedule for children, where the results presented in this thesis provided one of the evidence components necessary for decision-making (Koch et al., 2013). In Europe, routine rotavirus vaccination has also been introduced in Austria, Belgium, the United Kingdom, Finland, and Luxembourg as of today (Vesikari, 2008).

Although reduction of the clinical disease burden among young children was the main reason for vaccination recommendation, the STIKO expected routine childhood vaccination to have a beneficial effect also on the incidence among elderly due to herd protection, i.e. an overall reduced transmission within the population. However, it was not entirely clear whether vaccination might lead to an age shift of the infant disease burden into school age, thus effectively only delaying the infection. The impact on the rotavirus seasonality is also of interest, as the occurrence of biennial patterns has also been observed, e.g. after introduction of measles vaccination (Finkenstädt and Grenfell, 2000). Predicting the potential epidemiological impact of vaccine introduction regarding those additional aspects is thus one of the main goals of this thesis.

1.2   Mathematical modelling of infectious disease transmission

In this section we give a brief review of the mathematical modelling of person-to-person transmitted diseases including its central assumptions and major model classes. More comprehensive introductions are given by, e.g., Anderson and May (1991), Keeling and Rohani (2007), Andersson and Britton (2000) and Daley et al. (2001).

Modelling of infectious diseases
The underlying assumption of most infectious disease models is that infected individuals are able to transmit the disease to other people within the population. In order to capture this transmission process, individuals are distinguished according to their health state with respect to the considered disease. In the most simple case, there are three possible health states. Firstly, people who are not infected but are capable of acquiring the infection are called *susceptible*. Secondly, individuals having the infection and being able to pass it to others are called *infectious* or *infective*. Lastly, people who are not infectious, e.g. after recovery, but are also not susceptible for further infection are referred to as *removed*. Being removed may have different meanings, for instance it might imply that people died due to infection or gained full immunity after recovery. Either way, removed individuals do not further participate in infection transmission. This approach is referred to as the *susceptible-infectious-removed-model* or in short *SIR-model*.

The typical course of disease within the simple SIR-model is that individuals start as susceptible, become infectious at some unknown time point, and recover after a certain infectious period, thus becoming removed. Thereby, the chance of getting infected for a

susceptible individual generally depends on the state of other individuals in the model, which is the major cause for the typical non-linear dynamics within disease spreading. In contrast, the timing of removal once infected, i.e. the infectious duration, is typically independent from other individuals.

The simple SIR-model is easily extendable with further states, as it may be relevant for the transmission dynamics of the disease in question. For example, one often includes an additional state for those people who acquired infected but are not yet infectious, which are called *latent* or *exposed*, with the corresponding model being called the SEIR-model. It is also possible, to allow additional or alternative state transitions such as the possibility of people losing their acquired natural immunity over time and moving back from the removed to the susceptible state, which yields the SIRS-model. If the pathogen under consideration does not induce any natural immunity after recovered infection, it is also possible to leave out the removed-state completely, resulting in the SIS-model.

Furthermore, the model in its basic form does not distinguish between infection and illness, i.e. whether an infected person develops symptoms or not, although this may be relevant for transmission, e.g. any airborne disease is more likely to be transmitted in the case of frequent coughing and sneezing. For some other low-incidence diseases, like meningococcal or pneumococcal disease, it is important to account for asymptomatic infection, i.e. carriage, as this the primary transmission driver. However, regardless of how the eventual model is specified, the basic idea remains the same: to assign each individual one among a finite set of possible health states, with this state being subject to change over time.

From deterministic to stochastic models

The first models capturing this idea of distinct disease states were the deterministic transmission models developed by Hamer (1906) and later by Kermack and McKendrick (1927), both utilizing a compartmental approach (Daley et al., 2001; Bartlett, 1956). By decomposing the whole population under consideration into susceptible, infectious and recovered people, especially the simplified continuous time model by Kermack and McKendrick provides one of the foundations of disease transmission modelling as it is nowadays known as the basic deterministic SIR-model. These first models adopted the law of mass action for modelling the occurrence of new infections, i.e. the number of new infections is proportional to the product of the numbers of infectious and susceptible people. The dynamics within the basic SIR-model covering a fixed-size population is for this purpose represented through a set of ordinary differential equations.

$$
\begin{aligned}
\frac{dS(t)}{dt} &= -\alpha S(t)I(t), \\
\frac{dI(t)}{dt} &= \alpha S(t)I(t) - \beta I(t), \\
\frac{dR(t)}{dt} &= \beta I(t),
\end{aligned}
\tag{1.1}
$$

where the parameters $\alpha > 0$ and $\beta > 0$ denote the transmission and removal rate, respec-

tively. Here, the time-dependent processes S, I, and R each give the current number of individuals in the corresponding compartment of susceptible, infected and removed, which due to the ODE-structure of the model do not necessarily have to be integer counts. Typical initial conditions for the model consist of a high initial number $S(0)$ of susceptibles, a few infectious individuals $I(0)$, and zero removed people $R(0)$. Note, that this basic ODE-system contains no influx or outflux, such that the population size $N = S(t)+I(t)+R(t)$ remains constant. Under the assumption of a fixed population the last equation for $dR(t)/dt$ becomes even redundant as R is implicitly given through S and I. Analogue ODE systems can also be formulated for the SIS- or SEIR-type model or other variations. For more details on especially ODE-based transmission model see, e.g., Anderson and May (1991) or Keeling and Rohani (2007).

While the main advantage of deterministic models compared to stochastic models is their relative simplicity regarding simulation and analysis, their biggest disadvantage is the lack of realism within certain scenarios. The spreading of a disease is an inherently random process. For large populations the above deterministic approximation based on the mass-action-principle might be justified due to the law of large numbers. However, for small populations or for quantifying the probability of certain events – like the occurance of an outbreak or the extinction of a disease – stochastic models are indispensable. For detailed introductions to stochastic transmission modelling see e.g. Bailey and Bailey (1987), Becker (1989a), Daley et al. (2001) or Andersson and Britton (2000).

The simplest stochastic model utilizing the concept of disease state compartments, which was originally developed by Bartlett (1956), can be formulated as a bivariate Markov process $(S, I) \in \mathbb{N}^2$ with the following transition rates:

Event	Rate
$(S, I) \to (S - 1, I + 1)$	αSI
$(S, I) \to (S, I - 1)$	βI

Again, by assuming a fixed population size N the number of removed people is implicitly given by $R(t) = N - S(t) - I(t)$. Reformulating these event rates into a corresponding system of transition probabilities one obtains for a small time interval

$$
\begin{aligned}
\mathbb{P}\left[(S(t + h), I(t + h)) = (s - 1, i + 1) \,|\, (S(t), I(t)) = (s, i)\right] = h\alpha si + o(h), \\
\mathbb{P}\left[(S(t + h), I(t + h)) = (s, i - 1) \,|\, (S(t), I(t)) = (s, i)\right] = h\beta i + o(h),
\end{aligned}
\tag{1.2}
$$

where $o(\cdot)$ refers to the Bachmann-Landau notation. This simple stochastic model is motivated through the same underlying assumption as the basic ODE-model, i.e. the expected number of newly occurring infections within a small time interval is proportional to the number of susceptibles times the number of infected people, e.g. see Equation (1.3), whereas the expected number of removals is proportional to only the number of infected. However, the given event rates do not determine at which time points a transition occurs – only

how likely it is – which leads to a stochastic version. Also note, that the stochastic model accounts for the integer nature of the compartment counts, while the compartment sizes vary on a continuous scale in the deterministic model given by (1.1).

Unfortunately, finding the corresponding distribution of the state vector $(S(t), I(t))$ at a given time point t imposes a difficult challenge even for this simple stochastic process. A direct approach is to formulate the corresponding Kolmogorov forward equations, i.e. a system of differential equations governing the state probabilities $p_{(s,i)}(t) = \mathbb{P}\left[(S(t), I(t)) = (s, i)\right]$ $(0 \leq s, i \leq N, s + i \leq N)$ with respect to time for given initial conditions. This ODE system can then be solved by numerical means. However, the number of ODEs scales with the number of possible states and thus the population size N, which makes this approach numerically expensive when dealing with large populations. For more information on stochastic population models and their solutions see e.g. Matis and Kiffe (2000). An alternative approach is to apply simulation based estimation using Monte Carlo techniques, which requires the repeated simulation of the stochastic process. Again, performing a proper simulation, e.g. using the Gillespie simulation algorithm (Gillespie, 1976), is very costly in the case of high population counts since the transition rates change with each jump of the process, which happens very frequently for large numbers of susceptibles and infected. Approximate solutions might be obtained by imposing constant transition rates for small time intervals which leads to the τ-leaping algorithm (Gillespie, 2001). Please refer to Wilkinson (2006) for further simulation procedures. All-in-all the repetitive simulation of stochastic models – which would be necessary for inferring parameters of those models – requires much more sophisticated methods in contrast to those needed for deterministic models.

From stochastic to deterministic models
In order to choose a suitable model for capturing the rotavirus transmission in Germany we are interested whether in such large settings the additional stochasticity actually leads to a considerably more realistic model. Here, the relation between the deterministic model solution and the *mean* stochastic solution is of special importance. Assuming that these two processes indeed coincide, this would imply that both models would yield equal parameter estimates for α and β if inference would be performed by (laxly speaking) matching the mean solution to available data. This would heavily favour the deterministic model for statistical purposes due to its comparably easier numerical treatment.

Thus, we investigate the connection between the solution $(S_{\text{det}}(t), I_{\text{det}}(t))$ of the deterministic SIR-model given through the ODE system (1.1) and the component-wise mean $(\bar{S}(t), \bar{I}(t)) = \mathbb{E}[(S_{\text{st}}(t), I_{\text{st}}(t))]$ of the solution process $(S_{\text{st}}(t), I_{\text{st}}(t))$ according to the stochastic version from equation (1.2). Intuitively one might think that the mentioned identity of $(S_{\text{det}}(t), I_{\text{det}}(t))$ and $(\bar{S}(t), \bar{I}(t))$ indeed holds since the stochastic model version was constructed such that for each component the expected infinitesimal increment corresponds

to that of the deterministic solution, e.g.

$$\lim_{h \to 0} \mathbb{E}\left[\left.\frac{S_{\text{st}}(t+h) - S_{\text{st}}(t)}{h}\right| (S_{\text{st}}(t), I_{\text{st}}(t)) = (s,i)\right] = -\alpha s i = \left.\frac{dS_{\text{det}}(t)}{dt}\right|_{(S_{\text{det}}(t), I_{\text{det}}(t))=(s,i)}.$$
(1.3)

However, this does not imply that the mean $\bar{S}(t)$ is equal to the deterministic solution $S_{\text{det}}(t)$ for all time points. In fact, it was shown by Bailey (1950), that the stochastic mean indeed deviates from the deterministic solution, i.e. the process mean $(\bar{S}(t), \bar{I}(t))$ does not fulfill the ODE system given by (1.1) given the same parameters and initial conditions. Thus, since the stochastic evolution leads to deviations from the deterministic path, it affects the expected infinitesimal behaviour and therefore also the process mean at later time points. When comparing the respective SIR-model solutions, the stochastic mean $\bar{I}(t)$, i.e. the mean epidemic curve, is flatter than its deterministic counterpart $I_{\text{det}}(t)$ – exhibiting a lower peak and a slower descend. Therefore, according to the stochastic model an epidemic is predicted to take off slower and to last longer.

This still leaves the question whether for some scenarios the two solutions are at least approximately equal, since e.g. for large populations one might expect the relative stochastic effects to become negligible in the sense of the law of large numbers. As we mentioned earlier, in particular for large populations the stochastic SIR-model is difficult to analyse. One approach to circumvent this aspect is to derive a space-continuous model approximation of the Markovian discrete space SIR-model given by equation (1.2).

Such a so-called diffusion approximation was performed by Fuchs (2013) utilizing different approximation procedures, e.g. by convergence of the Kolmogorov forward equations or the infinitesimal generator of the process. Further approximation techniques are given by Allen (2003). As a necessary step to perform the diffusion approximations the state space $(S_{\text{st}}, I_{\text{st}}) \in \{0, \ldots, N\}^2$ of the stochastic SIR-model has to be mapped onto $(s,i) = (S_{\text{st}}, I_{\text{st}})/N \in \{0, 1/N, \ldots, 1\}^2$ such that s and i represent the fraction of susceptible and infectious people among the total population of size N. Since for large populations both s and i move with small steps of size $1/N$ within the interval $[0,1]$ the idea of the diffusion approximation is to derive a stochastic differential equation (SDE) for (s,i) with drift and volatility corresponding to those of the discrete space model. As shown by Fuchs (2013), by continuation of the state space one arrives at the following SDE:

$$\begin{pmatrix} ds \\ di \end{pmatrix} = \begin{pmatrix} -\alpha s i \\ \alpha s i - \beta i \end{pmatrix} dt + \frac{1}{\sqrt{N}} \begin{pmatrix} \sqrt{\alpha s i} & 0 \\ -\sqrt{\alpha s i} & \sqrt{\beta i} \end{pmatrix} dW_t,$$
(1.4)

where W_t denotes a two-dimensional Brownian motion. The corresponding initial conditions are obtained by transforming the initial values of the discrete space process, i.e. $(s(0), i(0)) = (S_{\text{st}}(0), I_{\text{st}}(0))/N$. Note that although for the derivation of the diffusion process the population size N was assumed to be infinitely large, it still appears as a parameter in the above SDE. Thus the solution process of equation (1.4) indeed provides an approximation to that of the stochastic SIR model given in (1.2), but not the limit process

for an infinitely large population itself. However, in the limit $N \to \infty$ one obtains

$$\begin{pmatrix} ds \\ di \end{pmatrix} = \begin{pmatrix} -\alpha si \\ \alpha si - \beta i \end{pmatrix} dt, \tag{1.5}$$

which is equivalent to the deterministic SIR-model given through (1.1) subject to the transformation of the state space. This result implies, that for very large population the stochastic effects of the Markovian SIR-model become indeed negligible and that its solution process can be well approximated by the deterministic SIR-model solution.

These properties have led to the deterministic SIR-model becoming the most frequently applied model class within infectious disease epidemiology. Also for our task of modelling disease transmission and vaccination within a population of several million people, a deterministic model provides a sufficient approximation to the slightly more realistic stochastic version, while being much more easier to handle. Therefore, the model class of ODE-based models will be the focal point of the present thesis, in particular with respect to suitable statistical procedures regarding parameter and structural uncertainty.

Homogeneity versus Heterogeneity
An important aspect to pay attention to is that compartment models assume homogeneity of all individuals with respect to both, individual attributes and mixing pattern. That means, that all individuals within the same compartment are assumed to have equal susceptibility, infectiousness and recovery rates, respectively. Furthermore, homogeneous mixing implies that each two individuals in the population have the same contact frequency such that each infectious individual is equally likely to infect any one susceptible. Since this assumed homogeneity certainly does not hold when considering a nationwide population, the applied model must be able to account for a variety of individuals at least to some extent. Two possibilities to do so are the utilization of an individual-based model, a whole new model class which aims to capture the heterogeneity of a population, or to construct a finer compartmentalization when applying a compartment model.

(a) Individual-based model
This class of individual-based (or agent-based) models also utilizes distinct states to represent the disease status of a population. However, instead of modelling the size of the relevant compartments of e.g. susceptible or infectious people, agent-based models consider all individuals specifically. In doing so, this model class allows for more heterogeneity of the population and the resulting transmission process. See e.g. Railsback and Grimm (2011) for a more detailed introduction into agent-based models.

Due to the change of perspective from compartments to individuals an individual-based model captures more information than just the momentaneous counts of infected and susceptible people for any time point, namely it is also measurable which specific individuals are actually infected. These additional information may then be used to construct more realistic spreading dynamics within the population.

In an individual-based model the stochastic model process is given through an N-dimensional vector $\boldsymbol{X}(t) = (X_1(t), \ldots, X_N(t))$, where N is the population size and $X_i(t) \in \{S, I, R\}$ denotes the health state of the i-th individual at time t. Then on a continuous time scale the infection probability for a specific susceptible individual is assumed to depend on the health states of all other individuals, i.e. by

$$\mathbb{P}\left[X_i(t+h) = I \,|\, X_i(t) = S, \boldsymbol{X}_{-i}(t) = \boldsymbol{x}_{-i}\right] = h \cdot g_i(t, \boldsymbol{x}_{-i}) + o(h),$$

where $\boldsymbol{X}_{-i}(t) = (X_1(t), \ldots, X_{i-1}(t), X_{i+1}(t), \ldots, X_N(t))$ refers to the health states of all other individuals and g_i captures the agent-specific *force of infection*, which might also be time-dependent. A classical candidate for g_i is the density dependent force of infection, which scales with the proportion of currently infectious people within the population, i.e.

$$g_i(t, \boldsymbol{X}_{-i}(t)) = \frac{\alpha_i}{N} \sum_{i=1, i \neq j}^{N} \mathbb{I}_{\{I\}}(X_j(t)),$$

where α_i is an agent-specific susceptibility factor and $\mathbb{I}_{\{I\}}$ denotes the indicator function. Assuming the density-dependent force of infection would hold for all agents with equal susceptibility factor $\alpha_i = \alpha$ the resulting model would yield equivalent results to the stochastic compartmental SIR-model (1.2) as the event rate for a newly occurring infection is proportional to the number of susceptibles times the number of infected people.

The attraction of the model is founded on the flexibility in constructing the individual force of infection $g_i(t, \boldsymbol{X}_{-i}(t))$ and analogously the individual recovery rates, which are commonly assumed not to depend on $\boldsymbol{X}_{-i}(t)$ but might also differ for each individual. Exploiting this flexibility might lead to spatial models, in which the force of infection depends primarily on those infected individuals which are locally near (e.g. see Perez and Dragicevic (2009)), or to network models in which individuals are only affected through neighboured agents on a pre-specified network, see e.g. Keeling and Eames (2005) or Andersson and Britton (2000, Ch.7). Of course, it is also possible to include agent-specific attributes, e.g. age or immunity characteristics, which may be relevant for the transmission process.

Due to their complexity, simulation from individual-based models is often tedious, especially for a large population. Theoretically, the Gillespie algorithm could provide an exact simulation for this model class, but it requires permanent recalculation of every individual's reaction rate, i.e. force of infection or recovery rate, as these may change after each movement of the process $\boldsymbol{X}(t)$. Thus, this approach becomes practically unfeasible in the case of many individuals. A common circumvention is to define a time-discrete version of the model and to perform simulation by deriving jump probabilities for each individual for each time step, similarly to the τ-leaping algorithm.

However, if one aims to map, e.g., the German population consisting of over 80 million individuals, a corresponding individual-based model becomes analytically and computationally intractable, since the overall model size scales with the population. Thus, this model class may be more suitable for micro-populations like one household or the staff and patients of a specific hospital ward. Another important issue is the incidence data, which

commonly consists of aggregated counts but rarely provides information on an individual basis. Therefore, conducting inference for such agent-based models is further complicated by lacking information on which individuals within a heterogeneous model population actually became infected. Again, such detailed data is more likely to be available in much smaller settings such as single households.

(b) Compartmentalization
In order to introduce some degree of heterogeneity into compartment models, a common approach is to further split the considered compartments, e.g. susceptibles $S(t)$, into sub-groups $S^{(k)}(t)$ ($k = 1, \ldots, K$) to account for certain sub-populations, see e.g. Anderson and May (1991, Ch.12). For instance, these might be separate age groups like children, adults and elderly to allow for variation regarding age-specific susceptibility or contact behaviour. Another possibility is to introduce a spatial decomposition of the population to obtain more realistic spatial spreading patterns over a predefined set of regions. The number of model equations then grows proportionally to the number of considered sub-groups, e.g. for the susceptibles within the deterministic SIR-model we obtain the K differential equations

$$\frac{dS^{(k)}(t)}{dt} = -\sum_{j=1}^{K} \alpha_{k,j} S^{(k)}(t) I^{(j)}(t),$$

for $k = 1, \ldots, K$. Note that the number of parameters might increase overproportionally as the former transmission rate α from the unstructured deterministic SIR-model (1.1) now becomes a $K \times K$ matrix $\boldsymbol{\alpha} = (\alpha_{k,j})$. This highlights the importance of defining only as many sub-populations as necessary for answering the questions the model was designed for. Otherwise an unnecessarily detailed population decomposition not only increases computational effort to analyse the model but might also lead to identifiability problems in the context of parameter inference. Thus, in order to define suitable and necessary compartments it is very important to assess which aspects, e.g. gender, age, region, are relevant for the transmission process. A close dialogue with epidemiologists becomes essential to identify the most important sources of heterogeneity.

Since further compartmentalization provides a good alternative compared to constructing an individual-based model and since it also leads to an ODE-based model, this approach will be pursued for our rotavirus transmission modelling presented in this thesis.

Data and Inference
As in every application of statistical modelling, fitting the employed transmission model to disease burden data is a necessary step in order to obtain realistic conclusions – for example about predictions on future occurrences when evaluating interventions like a vaccination programme. Additionally, the unknown parameters, e.g. the age-specific transmission rates, often have a direct epidemiological interpretation, such that corresponding estimates yield immediate insights into important transmission aspects.

However, relating such models to available incidence data is not trivial because the data generally cover only one aspect of the whole dynamics, that is the number of newly

occurred disease cases, often aggregated along certain time intervals (usually in a weekly mode), age groups and regions. Thus, most models capturing the time-continuous infection transmission on the population level need an additional component which links the unobserved SIR-dynamics with the time-discrete case notification data. This often also touches issues regarding under-detection on every level of the surveillance pyramid, i.e. from the infected individual who has to develop symptoms before considering to seek medical treatment, where a diagnostic test has to be conducted which eventually leads to case notification at the responsible health authorities.

Alternative approaches addressing the aggregated structure of the notification data include time series models which directly target the time-discrete incidence data, e.g. the weekly number of new infections, and thus circumvent unobservable components like the number of susceptible or recovered individuals, see e.g. Held et al. (2006) or Paul et al. (2008). Another possibility is to construct a time-discrete stochastic SIR-model version, which explicitly includes the weekly occurring number of new infections as one model component. For instance, by assuming a stepwise constant number of infected individuals I_t over a certain time interval $[t, t + \Delta t]$ it follows according to the event rates from the stochastic SIR-model given by (1.2) that the observed number of new infections ΔI_t within $[t, t + \Delta t]$ is given by

$$\Delta I_t \sim \text{Bin} \left(S_t, 1 - \exp(-\alpha I_t \cdot \Delta t) \right).$$

The updated numbers of susceptible and infected people at time $t + \Delta t$ are then approximated by

$$S_{t+\Delta t} = S_t - \Delta I_t, \quad I_{t+\Delta t} = I_t + \Delta I_t - \Delta R_t,$$

where ΔR_t is the weekly number of recovered people which can be constructed similarly. This and related models belong to the class of time series SIR-models (TSIR) which were proposed by e.g. Becker (1989b), Finkenstädt and Grenfell (2000), or Klinkenberg et al. (2002). A special case of this model is the so-called chain binomial model (Becker, 1989b) in which the discrete time step corresponds to one generation time such that all previously infected recover after one iteration, i.e. $\Delta R_t = I_t$. For the opposite case of relatively small time steps Δt yielding small probabilities $1 - \exp(-\alpha I_t \cdot \Delta t)$, the binomial distributions can be approximated through Poisson distributions given that S_t and I_t are sufficiently large.

However, while ODE-based models are better suitable for capturing the hidden dynamics and also preserve the time-continuity of the transmission process, they typically do not provide the number of new infected cases directly. Thus, when applying an ODE-based model this quantity must be derived from the ODE system itself, e.g. for the basic SIR-model (1.1) the rate of new infections is given by $\alpha S(t)I(t)$ which yields for the number of new infections within a certain time interval $[t, t + \Delta t]$:

$$\Delta I_t = \int\limits_t^{t+\Delta t} \alpha S(u)I(u)du.$$

For even more complex compartment models, the rate of new infections can be derived analogously. However, when relating this to observed data, the number of new infections ΔI_t as given above takes into account neither any under-detection nor the missing stochasticity within the reporting process. These aspects have to be considered within a separate stochastic observational model component.

This also constitutes one aspect because of which we favor a Bayesian approach for performing parameter inference in an ODE-based model. For instance, due to the unknown magnitude of the hidden dynamics the actual degree of under-detection of infected cases can not be necessarily identified using a frequetist framework, whereas a Bayesian procedure utilizing the available information is better capable of providing good parameter estimates, e.g. refer to Soriano-Gabarró et al. (2006) regarding the estimates of under-detection on every level of the surveillance pyramid. This also applies to many other transmission aspects which are featured as parameters in the model but were also already investigated in clinical or epidemiological studied, such as the duration of the infectious period or the mean contact frequencies. The second major advantage of a Bayesian framework is that it provides a natural environment for handling uncertainty regarding parameter estimates and model predictions, as this is one of our main goals when it comes to evaluating a possible rotavirus vaccination programme.

An inference problem especially relevant for ODE-based models, concerns the discrepancy between modelled and true dynamics since, as we mentioned earlier, the compartment models presented above are certainly not able to perfectly mimic the complex spreading process within a population of, e.g., 80 million people. This also poses the question of how to treat correlated data points, as for instance the case numbers are likely to exceed the model prediction for several subsequent days or weeks, due a short-term higher epidemic activity for reasons which are not captured by the model.

Another important aspect when dealing with ODE models is the relatively large computational effort required for model evaluation. Since in applied infectious disease epidemiology the employed models often use more than the three basic compartments of susceptible, infectious and removed people and also account for multiple population groups, the final number of compartments often ranges into the hundreds such that the iterative numerical solution of the model poses a huge computational task. However, repeated model evaluation is almost always necessary to compute estimates when using either frequentist and Bayesian methods which both have been applied in the past for ODE-based models, see e.g. (Elderd et al., 2006; Toni et al., 2009). Therefore, applied algorithms have to account for that computational effort and decrease the necessary number of model evaluations maybe at the cost of other computational steps in order to obtain good estimates.

Altogether, the inference for transmission models through disease burden data – which constitutes one of the central topics of this thesis – requires a careful treatment of many aspects including a proper link of model and data as well as numerically efficient inference methodology.

1.3   Outline and contributions

This work is structured as follows. The first part consisting of Chapters 2 and 3 will primarily cover methodological aspects when conducting inference for ODE-based models whereas the second part compromising chapters 4 to 6 deals with the modelling of the rotavirus transmission in Germany.

In Chapter 2 we will introduce basic concepts of Bayesian inference including posterior approximation based on asymptotic properties and Bayesian model selection. Based on the asymptotic behaviour we will propose a new approach for addressing autocorrelated observations within the posterior estimation for time series models, which in particular affects the posterior variance. The new method will be investigated within the context of simulated data from a simplified SIR-model.

Posterior sampling methods based on adaptive Markov chain Monte Carlo (MCMC) procedures will be presented in Chapter 3. The presented samplers are especially suitable for treating high-dimensional ODE-based models, in which the approximate shape of the posterior distribution is rarely known in advance. Regarding the calculation of the marginal likelihood for a specific model, we will derive novel estimation algorithms, which utilize information from existing posterior samples in order to set up a more accurate estimation procedure. Marginal likelihood estimates are in particular required for later model selection.

Chapter 4 introduces our ODE-based compartment model for the dynamic transmission of rotavirus in Germany, which addresses all key epidemiological components affecting the spreading process besides vaccination. In a Bayesian framework model parameters will be inferred using the asymptotic posterior approximation accounting for autocorrelation within the reported rotavirus incidence data. This yields new insights into transmission and reporting aspects and highlights the importance of an honest treatment of data subject to an imperfect model.

In Chapter 5 the proposed transmission model will be extended by additional vaccination mechanisms. Using data on the past vaccination coverage and incidence progress in the German eastern federal states we will estimate the rotavirus vaccine effectiveness within a Bayesian setting. Here, posterior samples will be obtained by MCMC methods where the resulting samples are used to determine necessary transmission components via Bayesian model selection criteria based on the marginal likelihood. The selected model is then used to sample the predictive distribution of the future rotavirus disease burden in Germany given the recommendation of routine vaccination.

All models and methods were implemented using the statistical software *R*. A structural overview and selected parts of the code are presented in Chapter 6. Here, we also highlight some of the challenges arising when implementing inference procedures where data and parameters are linked through an ODE-system and perhaps additional components.

Finally, in Chapter 7 we summarize the most important results from our work and provide possible directions for future research.

Certain chapters and sections of this thesis are in large parts based on previously published articles. For this thesis the respective manuscripts were adapted in order to obtain a consistent notation and to eliminate redundancy.

Chapter 4 and Sections 2.2 and 2.3 are based on

Weidemann, F., M. Dehnert, J. Koch, O. Wichmann, and M. Höhle (2014). Bayesian parameter inference for dynamic infectious disease modelling: rotavirus in Germany. *Statistics in Medicine 33* (9), 1580–1599.

and its corresponding supplementary material. For this paper Michael Höhle and Manuel Dehnert gave feedback on the employed transmission and observational model whereas Judith Koch and Ole Wichmann checked its epidemiological validity. Michael Höhle proposed to utilize a Gaussian approximation of the posterior based on asymptotic arguments and to apply model averaging within the inference framework. All co-authors helped streamlining the manuscript. The content and the results presented in Chapter 4 of this thesis correspond to those in the main article and appendix of the paper. Solely the introductory section was extended by some more details on existing inference approaches for ODE-based transmission models whereas Section 4.3 includes an additional part on the accuracy of the Gaussian approximation. Sections 2.2 and 2.3 are entirely based on the supplementary material of the paper with an additional introductory part explaining the problem of dealing with autocorrelated data.

Chapter 5 is based on

Weidemann, F., M. Dehnert, J. Koch, O. Wichmann, and M. Höhle (2014). Modelling the epidemiological impact of rotavirus vaccination in Germany – A Bayesian approach. *Vaccine 32* (40), 5250 – 5257.

Michael Höhle and Manuel Dehnert reviewed the augmented transmission model used for this analysis. Judith Koch and Ole Wichmann helped with the interpretation of the epidemiological results. All co-authors revised the manuscript. Chapter 5 of this thesis presents a composition of the mentioned paper and its corresponding supplementary material. To avoid redundancy, some details regarding the adaptive MCMC sampling procedure and the marginal likelihood estimation approach, which were mentioned in the article, were omitted in this chapter as they are also outlined in Chapter 3.
The manuscripts will be cited again at the beginnings of the respective sections.

Chapter 2

Bayesian inference based on asymptotic normality of the posterior accounting for autocorrelated data

In infectious disease epidemiology – but also in other time series applications – it is a known fact that subsequent data counts are likely to be correlated since, e.g., weeks of high disease incidence are often followed by additional high incidence weeks. However, in the statistical modelling of such infectious disease data such correlation structures are sometimes neglected which might lead to overconfident parameter estimates due to the faulty assumption of independent observations. In this chapter we propose a novel approach for acknowledging the dependencies within the data by incorporating the effective data sample size into the inference procedure. The presented method provides a flexible tool to address autocorrelated data, which does not require any adjustments within the underlying statistical model.

In a preliminary first section we will give a brief introduction into Bayesian inference, which constitutes the methodological base of this and all following chapters. Here, we will especially outline Bayesian asymptotic results and model selection methods. Our new approach for addressing dependent data will be presented in Section 2.2.

2.1 Bayesian inference

Bayesian inference, named after Thomas Bayes (1702-1761), is a statistical framework for estimation of parameters through observed data and thus provides an alternative to, e.g., frequentist inference methods. The underlying idea of Bayesian statistics is to interpret the unknown parameter ϑ, which is subject of estimation, as a random variable in contrast to being a fixed value. Thereby, the probability distribution of ϑ represents what is known about the parameter (or parameter vector) at a certain stage. As a result from the inference one therefore obtains a parameter distribution instead of a single point estimate.

However, this probabilistic representation does not necessarily imply that the parameter itself has to be considered as stochastic or not unique. Instead one way to interpret the corresponding probabilility distribution is that it reflects our subjective knowledge about the unknown parameter value. This is a suitable interpretation if we assume that there exists a true parameter which generated the data. On the other hand, for models providing an abstraction of the reality the corresponding model parameter distribution indicates which parameter values are more likely to explain the observed data. For this alternative interpretation we do not assume that there exists a unique true value. For an extensive overview on Bayesian statistics see, e.g., Gelman et al. (2013). For the brief introduction within this section we were mainly inspired by Held and Bové (2014).

As a central feature, Bayesian inference allows the user to incorporate prior knowledge about the parameter into the estimation procedure before observing any data. This is specified as a probability distribution π on the parameter space Θ of ϑ, the so-called *prior distribution* or *prior*. Depending on wether the parameter space Θ is either discrete or a continuous space, the prior is given as a probability mass function or as a density function, respectively. Thereby, the role of the prior is to define in which regions of the parameter space the true parameter is thought to be located, which is therefore subject to the researchers beliefs.

The impact of the data $\boldsymbol{D}$ is measured through a *likelihood function* $f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})$, which gives the probability of observing $\boldsymbol{D}$ given the parameter $\boldsymbol{\vartheta}$. As in frequentist statistics this likelihood function is defined through the statistical model which consists of a parametric probability distribution of the data where the respective parameter is subject of estimation. The likelihood as a function of the data is either given as a probability mass function in the case of count data (or otherwise discrete data), or through a probability density if the data comes from a continuous space.

Since the likelihood $f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})$ refers to the conditional probability of $\boldsymbol{D}$ given $\boldsymbol{\vartheta}$ one can derive the reverse conditional probability $\pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D})$ for any parameter vector given the observed data by applying Bayes' rule (Gelman et al., 2013) subject to the prior distribution, i.e. in the case of a continuous space Θ we have

$$\pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}) = \frac{f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta})}{\int_{\Theta} f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta})d\boldsymbol{\vartheta}}.$$

This is called the *posterior distribution* of the parameter $\boldsymbol{\vartheta}$ (or just *posterior*) and is the central quantity in Bayesian inference as it provides the probability distribution of the parameter after observing the data. One observes that the integral in the denominator represents the necessary normalization constant such that the posterior is indeed a proper distribution. This constant is also referred to as the marginal likelihood of the data and will later be relevant for Bayesian model selection in Section 2.1.2. Not accounting for normalization one obtains that the posterior distribution is proportional to the likelihood function times the prior, i.e. one often only writes

$$\pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}) \propto f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}).$$

Thus, the probability mass of the posterior is located in parameter regions which have a high prior probability but also are well capable of explaining the data, i.e. yielding a high likelihood. The likelihood function $f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})$ also has a nice measure-theoretic interpretation, because it corresponds to the unnormalized density function of the posterior with respect to the prior distribution, which is also known as the Radon-Nikodým derivative (Cohn, 2013)

$$f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta}) \propto \frac{d\pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D})}{d\pi(\boldsymbol{\vartheta})}.$$

Thus, the likelihood function indicates in which regions of the parameter space the parameter probability increases or decreases when switching from the prior to the posterior distribution.

Although the posterior constitutes the final inference result, point estimates and corresponding uncertainty regions can be directly derived from the posterior distribution. Classical candidates for point estimates are the *posterior mean* or the *posterior mode*, i.e.

$$\hat{\boldsymbol{\vartheta}} = \int_{\Theta} \boldsymbol{\vartheta}\,\pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D})d\boldsymbol{\vartheta}, \qquad \text{or} \qquad \hat{\boldsymbol{\vartheta}} = \underset{\boldsymbol{\vartheta}\in\Theta}{\arg\max}\;\pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}),$$

respectively, whereas the componentwise *posterior median*

$$\hat{\boldsymbol{\vartheta}} = (\hat{\vartheta}_1, \ldots, \hat{\vartheta}_d): \int\limits_{\left\{\boldsymbol{\vartheta}\,\middle|\,\vartheta_i\leq\hat{\vartheta}_i\right\}} \pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D})d\boldsymbol{\vartheta} = \int\limits_{\left\{\boldsymbol{\vartheta}\,\middle|\,\vartheta_i>\hat{\vartheta}_i\right\}} \pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D})d\boldsymbol{\vartheta} = \frac{1}{2} \qquad \forall i = 1, \ldots, d$$

is also a common choice. An uncertainty region with respect to an uncertainty level $\alpha \in [0, 1]$ is given through a so-called *credibility region C*, which is a set in the parameter space with posterior probability $1 - \alpha$. However, this set is not unique and thus the set of credibility regions of level α is given by

$$\mathcal{C}_\alpha = \left\{ C \subseteq \Theta \,\middle|\, \int_C \pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D})d\boldsymbol{\vartheta} = 1 - \alpha \right\}.$$

In contrast to the confidence interval known from frequentist statistics, any credibility region of level α has the intuitive interpretation that – assuming there exists a true parameter value – it contains the unknown parameter value with probability $1 - \alpha$. Note that in the case of some discrete parameters having positive probability mass, e.g. in the case of a discrete parameter space, many of the sets $\mathcal{C}_\alpha$ might be empty and the definition of a credibility region has to be adopted (Held and Bové, 2014). However, some credibility regions are of special interest as they fulfill certain properties. For instance, in the univariate case a popular candidate is the equi-tailed credibility interval $C = [q_{\alpha/2}, q_{1-\alpha/2}]$ where, e.g., $q_{\alpha/2}$ denotes the $\alpha/2$-quantile of the univariate posterior distribution. A convenient choice which also works in higher dimensions is a so called *highest-posterior-density region* of level α which refers to the corresponding credibility region $C \in \mathcal{C}_\alpha$ for which holds

$$\forall \boldsymbol{\vartheta} \in C, \tilde{\boldsymbol{\vartheta}} \notin C : \pi(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}) \geq \pi(\tilde{\boldsymbol{\vartheta}}\,|\,\boldsymbol{D}),$$

that is it contains the region of the parameter space yielding the highest values of the posterior density, in particular it contains the posterior mode. Again note, that even the highest-posterior-density region does not have to be unique. Furthermore, a highest-posterior-density region C_{HPD} is the smallest credibility region of a given level α (Wei and Tanner, 1990), i.e. let $\Theta \subseteq \mathbb{R}^{\mathrm{d}}$ then for every set $C \in \mathcal{C}_\alpha$ it holds

$$\int_C d\boldsymbol{\vartheta} \geq \int_{C_{\mathrm{HPD}}} d\boldsymbol{\vartheta}.$$

This is a very nice property as in practice one is often interested in uncertainty regions which are as small as possible.

2.1.1 Posterior computation and asymptotic behaviour

Since all inference targets such as point estimates and credibility regions depend on the posterior, the central task within Bayesian statistics is to obtain the posterior distribution. Depending on the statistical model at hand, this might be done by, e.g., analytical computation, approximation through known distributions, or generating a sample from the posterior for further processing.

Although computation of the (unnormalized) posterior as the product of likelihood function and prior density appears to be straightforward, depending on the parametric distribution of the prior and the likelihood (as a function of the data with parameter $\boldsymbol{\vartheta}$) the resulting posterior may not always originate from a known distribution family. In that case the computation of expectation, quantiles or probabilities of specific subsets can not be performed analytically and they have to be obtained by numeric means instead.

A reasonable approach is to choose the functional form of the prior (as it may be specified by the researcher) depending on the likelihood function $f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})$, such that the prior and posterior belong to the same distribution family. In that case the respective distribution family is referred to as *conjugate* with respect to the likelihood $f(\cdot\,|\,\boldsymbol{D})$ (Held and Bové, 2014). For instance, if the observed data points $\boldsymbol{D} = (D_i)_{i=1,\dots,n}$ are assumed to be binomial distributed with parameter p, i.e. $D_i \sim \mathrm{Bin}(n, p)$, and $p \sim \mathrm{Be}(a, b)$ is a-priori beta-distributed, then the resulting posterior for p is also a beta-distribution, but of course with different parameters. See, e.g., Held and Bové (2014) for a list of likelihood functions and its conjugate distribution families.

However, depending on the underlying statistical model the likelihood function can become very complex and sometimes might not even be available in closed form as we will see in our main application in Chapters 4 and 5. In such cases it is not possible to choose a suitable prior to obtain a nice posterior distribution that is easy to work with.

There exist two common approaches in order to handle the posterior distribution in these cases. One method is to generate a sample from the posterior, which can then be used to compute approximate means and median or identify high posterior regions using, e.g., kernel density estimation (Scott, 1992). The actual sampling procedure is often some type of Monte Carlo sampling or, more specifically, Markov chain Monte Carlo methods

which requires the posterior to only be known up to a constant and some relatively weak regularity conditions (Robert and Casella, 2004). This makes this sampling procedure especially useful for posterior sampling since the unnormalized posterior is directly available as the product of prior and likelihood. We will discuss these methods in more detail in Chapter 3.

A second approach is to compute an approximation to the posterior using distributions from a well known family, e.g. through normal distributions or mixed normal distributions. Thereby, the approximation procedure may be based on minimizing the Kullback-Leibler deviance between the intractable posterior and a specific distribution family using regression techniques, e.g. see Salimans, T. and Knowles, D. (2013). Another method is motivated by the asymptotic behaviour of the posterior distribution when the number of observed data points becomes increasingly large, which is described by the Bayesian Central limit Theorem (Gelman et al., 2013) or, in a stronger version, by the Bernstein-van-Mises-Theorem (van der Vaart, 1998, Th. 10.1). Under sufficient regularity conditions, e.g. the data $\boldsymbol{D} = (D_i)_{i=1,\ldots,n}$ being independently distributed subject to

$$f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) = \prod_{i=1}^{n} f_{\boldsymbol{\vartheta}}(D_i)$$

with $f_{\boldsymbol{\vartheta}}$ being the parametric density of one data point D_i subject to $\boldsymbol{\vartheta}$, the theory then states that the posterior distribution for $n \to \infty$ converges to a Gaussian distribution with mean and covariance determined by the posterior mode $\hat{\boldsymbol{\vartheta}}$ and the curvature of the posterior density in that point. More formally, we have the following convergence in law

$$\left(-\frac{\partial^2 \log \pi_n\left(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}\right)}{\partial^2 \boldsymbol{\vartheta}} \bigg|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}_n} \right)^{\frac{1}{2}} \left(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_n \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, I_d\right),$$

where $\pi_n(\boldsymbol{\vartheta} \,|\, \boldsymbol{D})$ and $\hat{\boldsymbol{\vartheta}}_n$ refer to the posterior for n independent data points $\boldsymbol{D} = (D_i)_{i=1,\ldots,n}$ and its corresponding posterior mode, respectively, and

$$\frac{\partial^2 \log \pi_n\left(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}\right)}{\partial^2 \boldsymbol{\vartheta}} \bigg|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}_n} = \left(\frac{\partial^2 \log \pi_n\left(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}\right)}{\partial \vartheta_j \partial \vartheta_k} \bigg|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}_n} \right)_{j,k=1,\ldots,d}$$

denotes the Hessian of the log posterior evaluated at $\hat{\boldsymbol{\vartheta}}_n$. However, this convergence property requires that there exists a true parameter $\boldsymbol{\vartheta}^* \in \Theta$ such that all data points are independently and identically distributed (i.i.d.) according to the probability density $f_{\boldsymbol{\vartheta}^*}$, i.e. the reality is indeed captured by the statistical model.

In practice this is rarely the case and the single data points D_i are i.i.d. according to some density $f(D)$ not included in the parametric family, i.e. $f(D) \neq f_{\boldsymbol{\vartheta}}(D)$ for all $\boldsymbol{\vartheta} \in \Theta$. A more general asymptotic convergence property (see (Gelman et al., 2013)) is then given by

$$[n\mathscr{I}\left(\boldsymbol{\vartheta}_0\right)]^{\frac{1}{2}} \left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, I_d\right),$$

where $\boldsymbol{\vartheta}_0 \in \Theta$ refers to the parameter which minimizes the Kullback-Leibler deviance between the parametric model likelihood densities $f_{\boldsymbol{\vartheta}}(D)$ and the true underlying distribution density $f(D)$, i.e.

$$\boldsymbol{\vartheta}_0 = \arg\min_{\tilde{\boldsymbol{\vartheta}} \in \Theta} \int f(D) \log \frac{f(D)}{f(D|\tilde{\boldsymbol{\vartheta}})} dD.$$

and $\mathscr{I}(\boldsymbol{\vartheta})$ denotes the expected *Fisher-Information* evaluated at $\boldsymbol{\vartheta}_0$ which is

$$\mathscr{I}(\boldsymbol{\vartheta}) = \mathbb{E}_{f(D)}\left[-\frac{\partial^2 \log f(D|\boldsymbol{\vartheta})}{\partial^2 \boldsymbol{\vartheta}}\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}\right].$$

From this convergence property it follows immediately that the choice of the prior has no effect on the asymptotic behaviour of the posterior, which is insofar intuitive as the prior assumptions eventually are outweighed by the increasing information gained through the increasing amount of data.

These asymptotic results motivate an approximation through a Gaussian distribution for scenarios in which large data sets are available. Define the approximated posterior by

$$\hat{\pi}(\cdot|\boldsymbol{D}) \sim \mathscr{N}\left(\boldsymbol{\vartheta}^*, \left(-\frac{\partial^2 \log \pi(\boldsymbol{D}|\boldsymbol{\vartheta})}{\partial^2 \boldsymbol{\vartheta}}\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}\right)^{-1}\right), \tag{2.1}$$

with mean $\boldsymbol{\vartheta}^*$ being the posterior mode and the covariance being determined by the Hessian of $\log \pi(\boldsymbol{\vartheta}|\boldsymbol{D})$ evaluated at $\boldsymbol{\vartheta}^*$. Here, the Hessian matrix could be replaced by the expected Fisher information $\mathscr{I}(\boldsymbol{\vartheta}^*)$ at the *maximum likelihood estimator* (MLE) $\boldsymbol{\vartheta}^*$, i.e. the parameter which maximizes $f(\boldsymbol{D}|\boldsymbol{\vartheta})$, or by the corresponding *observed Fisher information* (the Hessian of $\log f(\boldsymbol{D}|\boldsymbol{\vartheta})$), i.e.

$$\mathscr{I}_{\text{obs}}(\boldsymbol{\vartheta}^*) = \frac{\partial^2 \log f(\boldsymbol{D}|\boldsymbol{\vartheta})}{\partial^2 \boldsymbol{\vartheta}}\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*},$$

since these are asymptotically equal for increasing data size (Held and Bové, 2014). In particular, the posterior mode also converges to the maximum-likelihood-estimator. However, these approximations require the likelihood function to be sufficiently analytically tractable which is not always the case such as in our application in Chapter 4. Thus, using the approximation (2.1) is easily applicable in practice, since the posterior mode $\boldsymbol{\vartheta}^*$ and the corresponding Hessian can be obtained by numerical methods using suitable algorithms for optimization and derivation.

This approximate posterior will be applied within Chapter 4's modelling of rotavirus transmission, although with the added feature of adjusting the likelihood function for autocorrelation which will be discussed in Section 2.2. For more details on the asymptotic properties under standard conditions see Gelman et al. (2013), Held and Bové (2014), and van der Vaart (1998). Respective asymptotics have also been developed for more complex scenarios in which the single data points are not i.i.d., e.g. for linear models (Ghosal, 1999) or more general discrete-time stochastic processes (Heyde and Johnstone, 1979).

2.1.2 Bayesian model selection and averaging

In time series regression modelling, as in every other statistical model class, one often considers a family $\mathcal{M}$ of many different models for describing the data at hand. Model differences may thereby arise through the structure of the underlying regression term, e.g. by choosing different maximum orders for polynomial regression or either including or excluding interaction terms of multiple explanatory variables, or by defining different observational distributions, e.g. Poisson versus negative-binomial regression. An comprehensive overview of various regression techniques is given, e.g., by Fahrmeir et al. (2013).

A suitable statistical model should be able to explain the observed data sufficiently well while being defined as simple as possible. It is natural that every model extension always yields a better fit of the data due to more degrees of freedom being available. Thus, it is a known phenomenon that too complex models incorporating many variable parameters tend to 'overfitting' of the data, i.e. such models identify patterns in the data which actually do not exist (Hawkins, 2004). Thus, a reasonable criteria for model validity should account for both of these aspects, explanatory capability and complexity.

In frequentist statistics a widely used criteria for model selection is the *Akaike information criteria* (AIC) which measures the model validity as the maximum loglikelihood penalized by the number of parameters (Akaike, 1973), i.e.

$$\text{AIC} = 2d - 2 \log f\left(\boldsymbol{D} \,\middle|\, \hat{\boldsymbol{\vartheta}} \right),$$

where $\hat{\boldsymbol{\vartheta}}$ is the MLE and $d = \dim\left(\boldsymbol{\vartheta} \right)$ is the number of free parameters. Here, the model yielding the lowest AIC value within a considered set of models $\mathcal{M}$ should be preferred.

In Bayesian statistics an alternative to select one single best model is to consider a mixture of all models contained in the model set $\mathcal{M}$ (Kass and Raftery, 1995), e.g. for modelling the predictive distribution of the data or another derived outcome. Thereby, the share of an individual model $M_i \in \mathcal{M}$ is determined by its posterior model probability $\pi^{(\mathcal{M})}(M_i | \boldsymbol{D})$ given the data $\boldsymbol{D}$. The posterior model probability $\pi^{(\mathcal{M})}(M_i | \boldsymbol{D})$ is – analogously to inference on parameter level – a combination of a factor $f^{(\mathcal{M})}(\boldsymbol{D} | M_i)$ measuring each model's capability of explaining the data, determined by the likelihood of $\boldsymbol{D}$ given model M_i, and a prior model probability $\pi^{(\mathcal{M})}(M_i)$ representing the prior estimate about each model's validity to be specified by the user, i.e.

$$\pi^{(\mathcal{M})}(M_i | \boldsymbol{D}) \propto f^{(\mathcal{M})}(\boldsymbol{D} | M_i) \cdot \pi^{(\mathcal{M})}(M_i). \tag{2.2}$$

The term $f^{(\mathcal{M})}(\boldsymbol{D} | M_i)$ is called the *marginal likelihood* of the data corresponding to model M_i, but is also (laxly) referred to as the marginal likelihood of model M_i. It is defined by

$$f^{(\mathcal{M})}(\boldsymbol{D} | M_i) = \int f^{(i)}\left(\boldsymbol{D} \,\middle|\, \boldsymbol{\vartheta}^{(i)} \right) \pi^{(i)}\left(\boldsymbol{\vartheta}^{(i)} \right) d\boldsymbol{\vartheta}^{(i)},$$

i.e. as the integral of the M_i-specific likelihood function $f^{(i)}$ times the prior $\pi^{(i)}$. Thus, it resembles the marginalized likelihood of the data by integrating out the model parameter

$\boldsymbol{\vartheta}^{(i)}$ with respect to its prior. The marginal likelihood therefore has the very intuitive interpretation of providing the probability of observing the data $\boldsymbol{D}$ given model M_i. Note also, that the marginal likelihood depends on the choice of prior distribution since two models with an identical parametric structure but different prior distributions usually yield different marginal likelihoods. Thereby the model, which places high prior probability in parameter regions yielding high likelihoods, yields also the higher marginal likelihood. This becomes clear, because the marginal likelihood can also be written as the expectation of the model specific likelihood function with respect to the prior distribution, i.e.

$$f^{(\mathcal{M})}(\boldsymbol{D}\,|\,M_i) = \mathbb{E}_{\pi^{(i)}}\left[f^{(i)}\left(\boldsymbol{D}\,|\cdot\right)\right].$$

The ratio of the marginal likelihood of two models is called a *Bayes-factor* which provides a measure for the strength of evidence of one model against another, not accounting for prior model probabilities $\pi^{\mathcal{M}}(M_i)$, i.e.

$$B_{M_i,M_j} = \frac{f^{(\mathcal{M})}(D\,|\,M_i)}{f^{(\mathcal{M})}(D\,|\,M_j)},$$

where a Bayes-factor $B_{M_i,M_j} > 1$ means that model M_i is preferred over model M_j according to the data. Kass and Raftery (1995) provided a scale for interpreting the Bayes-factor in order to obtain a degree of evidence for/against a specific model (see Table 2.1.2). However, these interpretations should be considered as a guideline since fixed thresholds for grading the evidence from the Bayes-factor are highly subjective.

Table 2.1: Overview of how to interpret various values of the Bayes-factor according to Kass and Raftery (1995).

range of B_{M_1,M_2}	degree of evidence
$B_{M_1,M_2} \in [1,3]$	barely worth to mention
$B_{M_1,M_2} \in [3,20]$	positive evidence for M_1
$B_{M_1,M_2} \in [20,150]$	strong evidence for M_1
$B_{M_1,M_2} > 150$	very strong evidence for M_1

If one is just interested in identifying one best model instead of model-specific weights, another criteria which can be derived from the marginal likelihood is the *Bayesian information criterion* (BIC) (Schwarz, 1978). The BIC has the advantage that, due to its form similar to the AIC, it can be applied in both, frequentist and Bayesian settings. It is motivated by the asymptotic behaviour of the marginal likelihood if the model-specific likelihood functions $f^{(i)}(\boldsymbol{D}|\boldsymbol{\vartheta}^{(i)})$ belong to the exponential family (Held and Bové, 2014) and it is defined for a specific model by

$$\text{BIC} = \log(n)\,d - 2\log f\left(\boldsymbol{D}\,\Big|\,\hat{\boldsymbol{\vartheta}}\right),$$

where n refers to the sample size of the data set $\boldsymbol{D} = (D_i)_{i=1,\dots,n}$ and d to the number of model parameters. As with AIC, the model yielding the lowest BIC score should be

preferred. Hereby, the penalization of model complexity, i.e. the number of free parameters, is even more pronounced when using the BIC, as the penalization factor $\log(n)$ depends on the size of the data set. In the case of AIC the penalization factor is fixed at 2 which becomes negligible for large data sets as minor model extensions may still yield large increases of the maximum likelihood value.

Another strictly Bayesian measure for model validity, motivated by information-theoretic arguments, is the deviance-information-criterion (DIC) (Spiegelhalter et al., 2002), which is defined by the effective number of parameters p_D in the model, i.e.

$$p_D = \mathbb{E}_{\pi(\boldsymbol{\vartheta}|\boldsymbol{D})}\left[-2\log f\left(\boldsymbol{D}\,|\cdot\right)\right] + 2\log f\left(\boldsymbol{D}\,\Big|\hat{\boldsymbol{\vartheta}}\right),$$

where $\hat{\boldsymbol{\vartheta}}$ denotes the posterior mean. Here, the effective number of parameter may be lower than the actual size of the parameter vector due to implicit restrictions contained in the prior, e.g. if the prior imposes a strong correlation of some model parameters or if certain model parameters do not affect the likelihood. The DIC is then given by

$$\begin{aligned}
\text{DIC} = &-2\log f\left(\boldsymbol{D}\,\Big|\hat{\boldsymbol{\vartheta}}\right) + 2p_D \\
= &2\mathbb{E}_{\pi(\boldsymbol{\vartheta}|\boldsymbol{D})}\left[-2\log f\left(\boldsymbol{D}\,|\cdot\right)\right] + 2\log f\left(\boldsymbol{D}\,\Big|\hat{\boldsymbol{\vartheta}}\right).
\end{aligned}$$

One can see, that if the parameter restrictions through the prior are very low and the posterior mean is near the MLE, the DIC becomes approximately equivalent to the AIC.

One main difference in practice is the computability of the three selection criteria AIC, BIC and DIC. While evaluation of AIC and BIC requires the MLE, the DIC involves computation of the expected likelihood and the posterior mean. Thus, the DIC can be easily obtained if one has a posterior sample (and corresponding likelihood values) available, which is the case when sampling procedures such as MCMC methods are used for posterior computation. The AIC and BIC in any case require maximization of the likelihood function which, depending on the model, may need more or less additional effort.

However, all three criteria are only suitable to compare distinct models against each other and to identify one best model among them. If one is interested to assign a probability to each model of a certain model set $\mathcal{M}$ under consideration, one has to utilize posterior model probabilities as defined in (2.2). Since one often deals with a finite set of model, i.e. $|\mathcal{M}| < \infty$, the normalized probabilities can be easily calculated for all $M_i \in \mathcal{M}$ by

$$\pi^{(\mathcal{M})}(M_i\,|\boldsymbol{D}) = \frac{f^{(\mathcal{M})}(\boldsymbol{D}\,|M_i)\pi^{(\mathcal{M})}(M_i)}{\displaystyle\sum_{M\in\mathcal{M}} f^{(\mathcal{M})}(\boldsymbol{D}\,|M)\pi^{(\mathcal{M})}(M)}.$$

Still, computing the posterior model probabilities requires the calculation of the marginal likelihoods $f^{(\mathcal{M})}(\boldsymbol{D}\,|M_i)$. For some models, these can be computed analytically as the marginal likelihood coincides with the normalizing constant of the model specific posterior distribution $\pi^{(i)}(\cdot\,|\boldsymbol{D})$. Conversely, if the model's posterior distribution is difficult to obtain analytically then so is the marginal likelihood. In such situations it has to be either

approximated (e.g. by Laplace-approximation or by the BIC which is already an asymptotic approximation), computed by numerical integration, or estimated using Monte Carlo methods. Among the latter, Monte-Carlo based estimation approaches utilizing posterior samples are covered within Chapter 3.

In the case of having an infinite number of models available, evaluating the marginal likelihood for all models is less obvious. For that scenario, sampling procedures which sample simultaneously from the posterior model distribution $\pi^{(\mathcal{M})}(\cdot\,|\boldsymbol{D})$ and the corresponding parameter posteriors $\pi^{(i)}(\cdot\,|\boldsymbol{D})$ $(M_i \in \mathcal{M})$ are able to approximate the respective distribution (Toni et al., 2009; Green, 1995).

Having computed the posterior model probabilities, it is possible to calculate averaged values for quantities which are defined for each model in $\mathcal{M}$. For instance, when modelling infectious disease transmission, we are often interested in the predictive distribution of future case counts. However, different models yield different predictive distributions, but we are interested in a joint prediction from all models. More formally, suppose one is interested in a random variable with distribution function Q, which is defined within each model $M_i \in \mathcal{M}$. The distribution function Q, however, is defined differently for each model in $\mathcal{M}$ as it for instance may depend on the parameter $\boldsymbol{\vartheta}^{(i)} \in \Theta^{(i)}$ corresponding to M_i, where the parameter spaces $\Theta^{(i)}$ might not coincide for all considered models. Thus, let $Q^{(i)}$ denote the distribution function of Q when defined within a specific model M_i, i.e. $Q|_{M=M_i} = Q^{(i)}$. Then, the averaged distribution according to the posterior model distribution $\pi^{(\mathcal{M})}(\cdot\,|\boldsymbol{D})$ is defined via

$$Q = \sum_{M_i \in \mathcal{M}} \pi^{(\mathcal{M})}(M_i\,|\boldsymbol{D}) \cdot Q^{(i)}.$$

A common example is given through averaging the predictive distribution of the data $\boldsymbol{D}$ (here assuming it has an absolutely continuous distribution). For each model M_i the (mostly multi-dimensional) probability density $f^{(i)}(\cdot)$ of $\boldsymbol{D}$ is defined via the likelihood function $f^{(i)}(\cdot\,|\boldsymbol{\vartheta}^{(i)})$ and the parameter posterior $\pi^{(i)}(\cdot\,|\boldsymbol{D})$ by

$$f^{(i)}(\boldsymbol{D}) = \int_{\Theta^{(i)}} f^{(i)}\left(\boldsymbol{D}\,\middle|\,\boldsymbol{\vartheta}^{(i)}\right) \pi^{(i)}(\boldsymbol{\vartheta}^{(i)}\,|\boldsymbol{D})d\boldsymbol{\vartheta}^{(i)},$$

i.e. as the (with respect to the posterior) averaged likelihood of observing $\boldsymbol{D}$ in model M_i. The averaged predictive probability density is then given by

$$f(\boldsymbol{D}) = \sum_{M_i \in \mathcal{M}} \pi^{(\mathcal{M})}(M_i\,|\boldsymbol{D}) \cdot f^{(i)}(\boldsymbol{D})$$

Such averaging techniques can also be applied to, e.g., the distribution of a specific parameter component ϑ which is contained in every model in $\mathcal{M}$ by averaging over the respective model specific marginal distributions $\pi^{(i)}(\vartheta\,|\boldsymbol{D})$. Moreover, if the posterior model distribution is heavily degenerated in the favour of one single model, i.e. $\pi^{(\mathcal{M})}(M_i\,|\boldsymbol{D}) > 0.99$ for one M_i, it might be worthwhile to only use this single model for any further analyses

instead of including the marginal effect of all other improbable models through averaging processes.

Altogether, there exists a multitude of tools for measuring and comparing the validity of a set of models, which enables the identification of a best model to explain the data. For evaluation each of the different criteria requires different objects such as a posterior sample or the MLE. By computing posterior model probabilities, it is also possible to jointly process the whole ensemble of models with respective model shares. Thus, one can compute averaged values for quantities, which are common to all considered models. These could be, e.g., the averaged predictive distribution of the data or averaged parameter distributions.

2.2 A new approach for addressing autocorrelated observations in time series models

> The content of Sections 2.2 and 2.3 is based on the supplementary material published in Weidemann et al. (2014a).

When fitting statistical models to time series data $\boldsymbol{D} = (D_i)_{i=1,\dots,n}$ the default assumption is to take single observations D_i as independent given their respective expected values. This is due to the corresponding likelihood function $f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})$ being easy to handle, i.e. for some functions f_i and the parameter vector $\boldsymbol{\vartheta}$ it holds

$$f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta}) = \prod_{i=1}^{n} f_i(D_i\,|\,\boldsymbol{\vartheta}),$$

whereas deviation from the independence assumption would require to specify a certain dependency structure within the model's joint likelihood function $f(\boldsymbol{D}\,|\,\boldsymbol{\vartheta})$, which might not be known in advance. In particular when dealing with infectious disease transmission models based on ordinary differential equations the assumption of independence is frequently made, although a certain degree of correlation in subsequent observation residuals appears to be very likely. Thus, because of mismatches between the model and the real underlying dynamics the assumption of independence is often too strong. This can, e.g., be seen by checking the residuals between observed and predicted values which might exhibit a strong autocorrelation. Ignoring this dependency leads to bias during inference since the actual information content of the data is overrated when assuming independence. A consequence of this in frequentist statistics is an underestimation of the variance of the parameter estimates or correspondingly in a Bayesian framework an underestimation of the parameter's posterior variance.

To address this issue, we developed an approach based on rescaling the model likelihood function using the residual autocorrelation structure within the time series data. To do so, we aim to adjust the likelihood by some exponent $\mathrm{CA} \in (0, 1]$, which yields a new

likelihood function $f(\boldsymbol{D}\,|\boldsymbol{\vartheta})^{\mathrm{CA}}$ implying a so-called power posterior. The idea behind this adjustment is to reduce the impact of the data according to its actual information content, which is summarized within CA. Such an adjustment would favour the impact of the prior distribution in a Bayesian framework, but generally leads to higher uncertainty regarding parameter estimates due to decreasing the impact of the data. Moreover, our proposed procedure provides a suitable alternative to integrating the autocorrelation structure directly into the likelihood function $f(\boldsymbol{D}\,|\boldsymbol{\vartheta})$. The method is thereby derived based on Gaussian approximations of the posterior distribution or likelihood function as presented in Section 2.1.1. The new methodology was applied for Bayesian inference of the rotavirus transmission model presented in Chapter 4 and thus constitutes a suitable framework for obtaining more honest uncertainty assessments of parameters within ODE-based models. Such models could otherwise considerably suffer from the additional complexity imposed by a direct inclusion of dependency structures within the corresponding observational component.

Aim of this section is to provide the analytical rationale for this procedure as well as performance results from a simulation study. To do so we will first analyse maximum likelihood estimators corresponding to models using the assumption of either independent or autocorrelated residuals in a simple setting in Section 2.2.1. It will be shown that the observed Fisher information, whose inverse represents the asymptotic covariance matrix of the model parameters, differs by a factor equal to the cumulative autocorrelation (CA) within the data. This result motivates the approach of adjusting the observed Fisher information by rescaling the likelihood function by the CA. To evaluate the developed inference approach, we will conduct a simulation study in Section 2.3 by using a reduced transmission model framework that captures the essential problems of parameter inference for such models but still remains computationally tractable. In Section 2.3.5 we will conclude this work with a discussion of our gained insights and the transferability of the likelihood based adjustment procedure into a Bayesian context.

2.2.1   Adjusting observed Fisher information using the cumulative autocorrelation

In this section it will be shown that rescaling the observed Fisher information based on the cumulative autocorrelation within the data yields a more accurate assessment of uncertainty than ignoring residual dependencies in the observations. For this approach, we were inspired by the idea given in (Thiébaux and Zwiers, 1984) where the impact of autocorrelation on the sample variance was studied using the notion of effective sample size. For the simple analysis in this section, we follow a similar approach within the setting of estimating the sample mean of a time series using a maximum likelihood approach. We will analyse point estimators and corresponding observed Fisher information from likelihood models assuming either independent residuals or residuals following an autoregressive process of order p, i.e. AR(p), in a Gaussian setup. Both models yield the same point estimator for the sample mean but different results for the corresponding standard errors, where the

latter are computed based on the observed Fisher information matrices evaluated at their respective maximum. It will be shown that the ratio of the observed Fisher information solely depends on estimators for the AR-coefficients and corresponding autocorrelations. As main result we will show that this ratio can be identified as the cumulative autocorrelation of an AR(p)-process having the same estimated coefficients. This finding suggests a new approach for approximating standard errors and hence Wald-based confidence regions around the MLE of a sample mean by adjusting the standard errors originating from the model assuming independent observations with the cumulative residual autocorrelation inherent in the data. Note, that within this analysis we utilize frequentist inference methods, but the results are easily transferable to a Bayesian framework assuming a (possibly improper) uniform prior, where the posterior variance is also approximated via the observed Fisher information which yields equivalent results.

Independent observations

Let $\boldsymbol{X} = (X_i)_{i=1,\dots,N}$ be a stationary time series of observations of Gaussian variables with stationary mean μ. Our aim is the estimation of μ and to provide a confidence region for that estimate. One estimation approach is the maximisation of the loglikelihood function $\mathrm{LL}(\mu, \sigma^2)$ corresponding to the model assuming independent observations with unknown error variance σ^2

$$\mathrm{LL}\left(\boldsymbol{X} \,\middle|\, \mu, \sigma^2\right) = \log f\left(\boldsymbol{X} \,\middle|\, \mu, \sigma^2\right) = -\frac{N}{2} \log\left(2\pi\sigma^2\right) - \frac{\sum_{i=1}^{N}\left(X_i - \mu\right)^2}{2\sigma^2}.$$

Finding the score function and computing its roots yields the well known maximum likelihood estimators

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}, \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(X_i - \hat{\mu})^2}{N}.$$

To assess the corresponding uncertainty of these estimates one possibility is to use the property that the MLE is asymptotically multivariate normally distributed with mean μ and covariance matrix $\Sigma = H^{-1}$, where

$$H = -D^2 \,\mathrm{LL}\left(\mu, \sigma^2 \,\middle|\, \boldsymbol{X}\right)\Big|_{\mu=\hat{\mu}, \sigma^2=\hat{\sigma}^2}$$

is the Hessian of the negative loglikelihood function evaluated at the MLE (Efron and Hinkley, 1978), i.e. D^2 denotes the operator

$$D^2 g(\boldsymbol{x} = (x_1, \dots, x_d)) = \left(\frac{\partial^2 g(\boldsymbol{x})}{\partial x_i \partial x_j}\right)_{i,j=1,\dots,d}.$$

It is well known that μ and σ^2 are information-orthogonal given this model (Lindsey, 1996). Hence, if we are only interested in the uncertainty regarding $\hat{\mu}$, it is sufficient to

only consider the second derivative of the loglikelihood function with respect to μ evaluated at the MLE:

$$\left.\frac{\partial^2}{\partial \mu^2} \, \text{LL}\left(\mu, \sigma^2 \, |\boldsymbol{X}\right)\right|_{\mu=\hat{\mu}, \sigma^2=\hat{\sigma}^2} = -\frac{N}{\hat{\sigma}^2} = \frac{N^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}. \tag{2.3}$$

Note that this approach ignores any dependencies within the time series of observations $\boldsymbol{X}$. In the following we want to examine the effects of potential autocorrelation within the data on the estimators and their confidence regions. In order to do so simple models accounting for autocorrelated observations will be analysed.

AR(p) observations

We want to compare the above results with those of a model assuming that the time series $\boldsymbol{X}$ follows an AR(p) process with mean μ and autocorrelation coefficients $(\phi_1, \dots, \phi_p)$. Thus, we drop the assumption of independent observations. Instead the distribution and hence the observation X_i depends on the p previous observations as follows

$$(X_i - \mu) - \phi_1(X_{i-1} - \mu) - \dots - \phi_p(X_{i-p} - \mu) \sim \mathcal{N}\left(0, \sigma^2\right),$$

where $(X_i - \mu)_{i=1,\dots,N}$ will be referred to as the observational residuals. Assuming the past p values of the data time series $(X_{1-p}, \dots, X_0)$ to be known this yields the loglikelihood function

$$\text{LL}\left(\boldsymbol{X} \,\Big|\, \mu, \sigma^2, \phi_1, \dots, \phi_p\right) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{\sum_{i=1}^{N}\left(X_i - \left(\mu + \sum_{j=1}^{p}\phi_j\left(X_{i-j} - \mu\right)\right)\right)^2}{2\sigma^2}. \tag{2.4}$$

Given the case where the past of the data is unknown, the term above defines the loglikelihood function conditioned on $(X_{1-p}, \dots, X_0)$. To derive the loglikehood from there, we have to additionally consider the marginal likelihood of the vector $(X_{1-p}, \dots, X_0)$, which is asymptotically negligible if $N >> p$ (Brockwell and Davis, 1991, Ch. 8.8). Therefore, it is sufficient to look at the loglikelihood function as stated in (2.4). We again obtain the MLEs by computing the roots of the score function, the first derivative of the loglikelihood, which yields

$$\hat{\mu} = \frac{\left(\sum_{i=1}^{N} X_i - \sum_{j=1}^{p}\phi_j \sum_{i=1}^{N} X_{i-j}\right)}{N\left(1 - \sum_{j=1}^{p}\phi_j\right)},$$

$$\begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\gamma}_1 \\ \hat{\gamma}_{p-1} & \cdots & \hat{\gamma}_1 & \hat{\gamma}_0 \end{pmatrix} \hat{\boldsymbol{\phi}} = \begin{pmatrix} \hat{\gamma}_1 \\ \vdots \\ \vdots \\ \hat{\gamma}_p \end{pmatrix} \quad \text{with} \quad \hat{\gamma}_k = \frac{\sum_{i=1}^{N}\left(X_{i-k} - \hat{\mu}\right)\left(X_i - \hat{\mu}\right)}{N},$$

and

$$\hat{\sigma}^2 = \frac{1}{N}\left[\sum_{i=1}^{N} X_i - \hat{\mu} - \sum_{j=1}^{p} \hat{\phi}_j\left(X_{i-j} - \hat{\mu}\right)\right]^2 = \frac{\sum_{i=1}^{N}\left(X_i - \hat{\mu}\right)^2}{N}\left(1 - \sum_{j=1}^{p} \hat{\phi}_j \frac{\hat{\gamma}_j}{\hat{\gamma}_0}\right).$$

These MLEs play an important role in the theory of linear regression with time lags (Christensen, 2001, Ch. 5.5.2), while the matrix equation for $\hat{\boldsymbol{\phi}}$ is known as the Yule-Walker equations (Brockwell and Davis, 1991, Ch. 8.1). Maximum likelihood estimators for more general ARMA-models with zero mean can also be found in (Brockwell and Davis, 1991, Ch. 8.7). However, since we are mainly interested in the estimation of μ and its corresponding uncertainty via Fisher information some of these results are recapped here. Firstly, note that

$$\hat{\mu} \overset{\mathrm{a}}{=} \frac{\bar{X} - \sum_{j=1}^{p} \phi_j \bar{X}}{1 - \sum_{j=1}^{p} \phi_j} = \bar{X}$$

is asymptotically equal to the point estimator from the setting assuming independent observations, where $x \overset{\mathrm{a}}{=} y$ denotes that $x = y + \mathcal{O}(\frac{1}{N})$. The MLE for σ^2 can be rewritten using the notion of empirical autocorrelation $\hat{\boldsymbol{\rho}}$, i.e.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}\left(X_i - \hat{\mu}\right)^2}{N}\left(1 - \sum_{j=1}^{p} \hat{\phi}_j \hat{\rho}_j\right), \quad \hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}.$$

Also in this case, μ is asymptotically information-orthogonal with respect to σ^2 and the correlation coefficients ϕ_k due to

$$\frac{\partial^2}{\partial\mu\partial\phi_k} \mathrm{LL}\left(\boldsymbol{X}\,\middle|\,\mu,\sigma^2,\boldsymbol{\phi}\right)\bigg|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2,\phi=\hat{\phi}}$$

$$= -\left(1 - \sum_{j=1}^{p} \hat{\phi}_j\right)^{-1}\frac{\partial}{\partial\mu} \mathrm{LL}\left(\boldsymbol{X}\,\middle|\,\mu,\sigma^2,\boldsymbol{\phi}\right)\bigg|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2,\phi=\hat{\phi}} + \left(1 - \sum_{j=1}^{p} \hat{\phi}_j\right)\left(N\hat{\mu} - \sum_{i=1}^{N} X_{i-k}\right)$$

$$\overset{\mathrm{a}}{=} 0,$$

which follows from $(\hat{\mu}, \hat{\sigma}^2, \hat{\boldsymbol{\phi}})$ being the root of the derivative of LL with respect to μ and also

$$\hat{\mu} \overset{\mathrm{a}}{=} \frac{\sum_{i=1}^{N} X_{i-k}}{N}.$$

Thus, to obtain uncertainty of the estimator $\hat{\mu}$ it is again sufficient to evaluate the second derivative of LL with respect to only μ at the MLE, which yields

$$\frac{\partial^2}{\partial\mu^2} \mathrm{LL}\left(\boldsymbol{X}\,\middle|\,\mu,\sigma^2,\phi_1,\ldots,\phi_p\right)\bigg|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2,\phi_1=\hat{\phi}_1,\ldots,\phi_p=\hat{\phi}_p} = \frac{\left(1 - \sum_{j=1}^{p} \hat{\phi}_j\right)^2}{1 - \sum_{j=1}^{p} \hat{\phi}_j \hat{\rho}_j}\frac{N^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}.$$

$$(2.5)$$

Due to information-orthogonality, we can compute the ratio of the variances of the MLE $\hat{\mu}$ from the two models by computing the inverted ratio of the corresponding loglikelihood's second derivative with respect to μ evaluated at their respective MLEs as given in (2.3) and (2.5). Thus, we obtain a factor that solely depends on the estimates $(\hat{\phi}_1, \ldots, \hat{\phi}_p)$ and $(\hat{\rho}_1, \ldots, \hat{\rho}_p)$:

$$\frac{\operatorname{Var}_{\text{IND}}(\hat{\mu})}{\operatorname{Var}_{\text{AR}}(\hat{\mu})} = \frac{\left(\frac{\partial^2}{\partial \mu^2} \operatorname{LL}_{\text{IND}(\boldsymbol{X}|\mu,\sigma^2)}\Big|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2}\right)^{-1}}{\left(\frac{\partial^2}{\partial \mu^2} \operatorname{LL}_{\text{AR}(\boldsymbol{X}|\mu,\sigma^2,\phi_1,\ldots,\phi_p)}\Big|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2,\phi_1=\hat{\phi}_1,\ldots,\phi_p=\hat{\phi}_p}\right)^{-1}} = \frac{\left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2}{1 - \sum_{j=1}^p \hat{\phi}_j \hat{\rho}_j}.$$

(2.6)

As expected, the additional consideration of autocorrelation has a significant impact on the observed Fisher information regarding the sample mean, as the ratio stated in (2.6) tends to zero for highly autocorrelated samples. This demonstrates the substantial under-estimation of the estimate's variance when autocorrelation is erroneously ignored, as the variance is asymptotically equal to the inverted observed Fisher information. However, the following proposition introduces the cumulative autocorrelation (CA) of a time series which in the case of AR(p)-processes can be identified with the inverted variance ratio from (2.6) regardless of the order p assumed in the underlying model accounting for autocorrelation.

Proposition 1. *Let $\boldsymbol{Y} = (Y_t)$ be an AR(p) process with coefficients $\phi_1, \ldots, \phi_p$ and σ^2 such that*

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0 \quad \forall |z| \leq 1 \tag{2.7}$$

and let ρ_k be the lag-k autocorrelation of $\boldsymbol{Y}$. Then for the cumulative autocorrelation it holds:

$$\operatorname{CA} := \sum_{k=-\infty}^{\infty} \rho_k = \frac{1 - \sum_{j=1}^p \phi_j \rho_j}{\left(1 - \sum_{j=1}^p \phi_j\right)^2}$$

Proof. Let $G(z)$ denote the autocovariance generating function of the process $\boldsymbol{Y}$

$$G(z) = \sum_{k=-\infty}^{\infty} \gamma_k \cdot z^k$$

where γ_k denotes the lag-k autocovariance of $\boldsymbol{Y}$. Due to condition (2.7) it holds ((Brockwell and Davis, 1991, Ex. 3.5.1))

$$G(z) = \frac{\sigma^2}{\phi(z)\,\phi(z^{-1})}.$$

Using the identity ((Brockwell and Davis, 1991, Eq. 3.3.8))

$$\sigma^2 = \gamma_0 - \sum_{j=1}^p \phi_j \gamma_j$$

it follows immediately that

$$\mathrm{CA} = \frac{G(1)}{\gamma_0} = \frac{\gamma_0 - \sum_{j=1}^{p} \phi_j \gamma_j}{\gamma_0 \phi(1)\phi(1)} = \frac{1 - \sum_{j=1}^{p} \phi_j \rho_j}{\left(1 - \sum_{j=1}^{p} \phi_j\right)^2}.$$

$\square$

Our analysis showed that the estimate's variance regarding $\hat{\mu}$ based on the inverted observed Fisher information from the two different models differs by a factor equal to the inverted cumulative autocorrelation:

$$\frac{\mathrm{Var}_{\mathrm{AR}}(\hat{\mu})}{\mathrm{Var}_{\mathrm{IND}}(\hat{\mu})} = \frac{\left.\frac{\partial^2}{\partial \mu^2}\,\mathrm{LL}_{\mathrm{IND}(\boldsymbol{X}|\mu,\sigma^2)}\right|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2}}{\left.\frac{\partial^2}{\partial \mu^2}\,\mathrm{LL}_{\mathrm{AR}(\boldsymbol{X}|\mu,\sigma^2,\phi_1,\dots,\phi_p)}\right|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2,\phi_1=\hat{\phi}_1,\dots,\phi_p=\hat{\phi}_p}} = \sum_{k=-\infty}^{\infty} \hat{\rho}_k = \widehat{\mathrm{CA}}.$$

This result suggest the following new approach for parameter inference of general mean structure models by accounting for otherwise unconsidered dependency structures within the observations: First estimate model parameters $\boldsymbol{\vartheta}$ using a likelihood function $\mathrm{LL}_{\mathrm{IND}}$ corresponding to a model assuming independent observations. Then, afterwards, adjust the resulting asymptotic variance $\mathrm{Var}_{\mathrm{IND}}\left(\hat{\boldsymbol{\vartheta}}\right)$ from the observed Fisher information for the relevant parameters by an estimate for the cumulative autocorrelation of the observation residuals, i.e.

$$\widehat{\mathrm{CA}} = \sum_{k=-\infty}^{\infty} \hat{\rho}_k,$$

which finally yields an adjusted variance

$$\mathrm{Var}_{\mathrm{adjust}}\left(\hat{\boldsymbol{\vartheta}}\right) = \widehat{\mathrm{CA}}\,\mathrm{Var}_{\mathrm{IND}}\left(\hat{\boldsymbol{\vartheta}}\right).$$

This approach has three major advantages:

1. In order to obtain a more realistic assessment of uncertainty for parameters regarding the sample mean, it is not necessary to estimate additional parameters regarding the autocorrelation, which may result in a high computational effort especially if one expects autocorrelation of high order.

2. There is no need to settle for a model incorporating the autocorrelation in advance, e.g. to specify a maximum order of possible autoregressive coefficients to be accounted for. This makes the approach very flexible regarding the underlying autocorrelation structure inherent in the data. Thus, the procedure also avoids the problem of model selection in order to find an accurate joint model for both, the underlying means and the correlation structure.

3. It is possible to adapt this approach to a Bayesian framework by estimating the residual autocorrelation based either on the posterior mode or a posterior sample

and by adjusting only the likelihood part of the posterior distribution. This leads to a two-step procedure, where the posterior incorporating the adjusted likelihood is computed within a second step as it will be done in the main application in Chapter 4.

One remaining question is how to obtain an appropriate estimator for the cumulative autocorrelation. The empirical autocorrelation within a time series suffers from bias and a large variance with increasing lag k (Thiébaux and Zwiers, 1984), which makes it an unsuitable estimator for our approach. One possibility to avoid this problem is to dampen the impact of the higher order autocorrelation estimates by defining

$$\widehat{\mathrm{CA}} = \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) \hat{\rho}_k$$

with N being the sample size. This corresponds to the definition of the effective sample size of a data sample as it was given in Thiébaux and Zwiers (1984). However, this term underestimates the cumulative autocorrelation and therefore again leads to an underestimation of the model parameter variance. An alternative approach is to fit an $\mathrm{AR}(p)$ or $\mathrm{ARMA}(p,q)$ process to the time series of residuals and use some model selection criterion, e.g. AIC, to choose the optimal parameters p and q. Based on the best fitting process the cumulative autocorrelation may then be easily estimated, since the ρ_k are implicitly given as functions of the autoregressive and moving average coefficients in an $\mathrm{ARMA}(p,q)$ model (see Brockwell and Davis (1991, Ch. 3.2)). By following that approach one still obtains a best time series model (within the given model class) and corresponding parameters for the residual time series, but without having to conduct simultaneous maximum likelihood estimation for all parameters, i.e. those from the underlying mean structure model and those from the prespecified observation residuals model.

Generalizing the results of this section, it is worth pointing out that a similar analysis can be performed for more comprehensive models where the expectation of each observation may vary with time depending on a model parametrised by a parameter vector $\boldsymbol{\theta}$ such that

$$\mu_t = \mu(\boldsymbol{\theta}, t), \quad t = 1, \ldots, N.$$

Defining the analogue likelihood function with or without autoregressive coefficients and conducting the same computations leads to equivalent results regarding the asymptotic variance of the parameters $\boldsymbol{\theta}$ based on the observed Fisher information. This is especially crucial considering inference of ODE-based disease transmission models where the expected incidences are computed by an ODE-system given a vector of parameters as it is the case for our dynamic transmission model described in Chapters 4 and 5.

In the following section we want to study the performance of the proposed method in the context of dynamic transmission modelling. Thereby, our focus lies on examining uncertainty assessment of various methods in a more complex environment which is heavily inspired by our rotavirus transmission model presented in Chapter 4.

2.3   Simulation study: Cumulative autocorrelation

Within a simulation study we now aim to investigate the accuracy of the newly proposed adjustment procedure based on the CA using a setup of ODE-based transmission models. By fitting a simple SIRS-model to simulated surveillance time series data we compare the performance of this approach with those of established methods for parameter inference. Furthermore, using Markov Chain Monte Carlo it will be checked whether a multivariate normal distribution centered at the MLE or posterior mode, respectively, yields an appropriate approximation. Finally, we examine whether the results also hold in a count data setup based on negative binomially distributed observations.

Transmission model and data simulation

In the present study we examined a simplified version of the SIRS-model compared to the one used in the main rotavirus application from Chapters 4 and 5. This simple model neither accounts for age structure nor for multiple susceptibility levels, but still includes seasonal transmission. The population of size N is divided into the three classes of susceptibles (S), infected (I), and recovered (R) with $S + I + R = N$ at all times. The transmission dynamics are governed by the following system of ordinary differential equations:

$$\frac{dS}{dt} = -R_0 \beta \left\{ 1 + a \cos\left(\frac{2\pi t}{52}\right) \right\} \frac{SI}{N} + \delta(N - S),$$
$$\frac{dI}{dt} = R_0 \beta \left\{ 1 + a \cos\left(\frac{2\pi t}{52}\right) \right\} \frac{SI}{N} - \beta I - \delta I,$$
$$\frac{dR}{dt} = \beta I - \delta R.$$

The parameters R_0 and β represent the basic reproduction number and the recovery rate (1/mean duration of infection), respectively. The population size $N = 82,000,000$ remains constant due to both birth and death rate being equal to the same parameter δ. The parameter a controls the amplitude of the seasonal variation in transmission, which was modelled as one harmonic oscillation for the sake of simplicity. For the purpose of this study, we fixed the latter two parameters at $a = 0.2$ and $\delta = 4000$ such that R_0 and β remain as the only two transmission model parameters to be estimated from data. The model runs for a prespecified time T where one time unit represents one week. In order to get a data setting corresponding to routine public health surveillance, and similar to the data setting of our rotavirus application discussed in Chapter 4, we translated the output of the model into a record of new infections within a weekly time interval. We calculated this number of new weekly infections X_t by integrating the infectious inflow over the course of one week t, i.e.

$$X_t = R_0 \beta \int_{t-1}^{t} \left(1 + a \cos\frac{2\pi u}{52}\right) \frac{S(u)I(u)}{N} du.$$

To additionally incorporate observational noise we added a mean-zero noise process $(\epsilon_t)_{t=0,\ldots,T}$ to the time series $\log(X_t)_{t=0,\ldots,T}$. Our interest was now on investigating the consequences of autocorrelation within the data, which was given through the process $(\epsilon_t)_{t=0,\ldots,T}$. We hence chose the noise process to be governed by either an $\mathrm{ARMA}(p,q)$ process or alternatively by a Brownian bridge process W_t, which returns to zero with the end of each year $(t = k \cdot 52, \, k \in \mathbb{N})$, i.e.

$$W_{t+52k} = B_{t+52k} - B_{52k} - \frac{t}{52}\left(B_{52+52k} - B_{52k}\right), \quad t \in \{0,\ldots,51\},$$

where B is a standard Brownian motion (Chung and Williams, 2014). In both cases, process realisations are straightforward to simulate. The resulting simulated time series $(D_t)_{t=0,\ldots,T}$ of the observed number of new infections is then given by

$$\log D_t = \log X_t + \sigma \cdot \epsilon_t,$$

where $\sigma > 0$ controls the magnitude of the additional noise. Within this setting the noise process $(\epsilon_t)_{t=0,\ldots,T}$ is the only source of randomness while the mean process $(X_t)_{t=0,\ldots,T}$ is deterministic, depending only on the parameters of the ODE system and its initial conditions. In contrast to our main application in Chapter 4, the data generated within this simulation study do not have a discrete distribution on $\mathbb{N}$ but are continuously distributed on $\mathbb{R}_+$ instead. This has the advantage that we can use established inference methods for continuous data as a benchmark for comparison. In Section 2.3.4 extensions to the count data case will be discussed.

2.3.1 Inference methods

Within this setting we compared three different methods for parameter estimation. Each approach estimated the three parameters R_0, β, σ, and further parameters regarding the residual autocorrelation where necessary. The first method is based on using the loglikelihood function corresponding to a model assuming independent observations

$$\mathrm{LL}_{\mathrm{IND}}\left(\boldsymbol{D}\,|R_0,\beta,\sigma\right) = \log f_{R_0,\beta,\sigma}^{\mathrm{IND}}\left(D_1,\ldots,D_T\right) = \sum_{t=1}^{T} \log \phi\left(\log(D_t); \mu_t = \log\left(X_t(R_0,\beta)\right), \sigma\right),$$

where $\phi(x;\mu,\sigma)$ denotes the pdf of the normal distribution with expectation μ and standard deviation σ and $X_t(R_0,\beta)$ denotes the model predicted expected number of observed cases in week t as described above. This method assumes no residual autocorrelation of the observations given subject to the model, since all observations given their respective expected values are independent from each other. We computed $(1-\alpha) \cdot 100\%$-confidence regions for the unknown parameters determining the model mean, i.e. R_0 and β. Again note, that point estimates and uncertainty regions are here derived within a frequentist framework. However, equivalent results would be obtained by Bayesian methods when assuming that the prior is uniform and therefore improper in this particular setting.

By assuming asymptotic normality of the MLE we obtain

$$\mathrm{CI}_{1-\alpha} = \left\{ \boldsymbol{\vartheta} = (R_0, \beta) \left| \left(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}} \right) \hat{\Sigma}^{-1} \left(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}} \right)^T \leq q_{1-\alpha} \right. \right\}, \tag{2.8}$$

where $\hat{\boldsymbol{\vartheta}} = (\hat{R}_0, \hat{\beta})$ is the maximum likelihood estimator, $\hat{\Sigma}^{-1}$ the observed Fisher information matrix with respect to (R_0, β), i.e.

$$\hat{\Sigma}^{-1} = - \left. D^2_{R_0, \beta} \mathrm{LL}_{\mathrm{IND}} \left(\boldsymbol{D} \left| R_0, \beta, \dots \right. \right) \right|_{\left(R_0 = \hat{R}_0, \beta = \hat{\beta}, \dots \right)}$$

and $q_{1-\alpha}$ is the (1-α)-quantile of the chi-squared χ^2_2-distribution with two degrees of freedom.

The second inference approach is our newly proposed method which is based on the same loglikelihood function defined above, but additionally adjusting the loglikelihood, and thus the resulting Fisher information, by the cumulative residual autocorrelation of the data. The cumulative autocorrelation of a data set $\boldsymbol{D}$ given the best fitting model parameters $(\hat{R}_0, \hat{\beta})$ using $\mathrm{LL}_{\mathrm{IND}}$ is defined as the accumulated autocorrelations of the time series of log residuals

$$R_t = \log D_t - \log X_t(\hat{R}_0, \hat{\beta}).$$

To obtain an estimator for the autocorrelations $\hat{\rho}_i$ of the time series $\boldsymbol{R} = (R_1, \dots, R_T)$ we determined the best fitting ARMA(p, q) process by using the AIC criteria to find the optimal orders p^* and q^*. Using the corresponding coefficient estimates $\hat{\phi}_1, \dots, \hat{\phi}_{p^*}$ and $\hat{\theta}_1, \dots, \hat{\theta}_q^*$ we computed the autocorrelation function $\hat{\rho}(\tau)$ for lag τ as described in (Brockwell and Davis, 1991, Ch. 3.2). Using $\hat{\rho}(\tau)$ as an estimator for the residual autocorrelation the estimated cumulative autocorrelation $\widehat{\mathrm{CA}}(\boldsymbol{D})$ was then computed by

$$\widehat{\mathrm{CA}}(\boldsymbol{D}) = \sum_{\tau=-T}^{T} \hat{\rho}(\tau).$$

Finally, we adjusted the loglikelihood function $\mathrm{LL}_{\mathrm{IND}}$ with the cumulative autocorrelation of the data $\widehat{\mathrm{CA}}(\boldsymbol{D})$ yielding

$$\mathrm{LL}_{\mathrm{CA}} = \widehat{\mathrm{CA}}(\boldsymbol{D})^{-1} \cdot \mathrm{LL}_{\mathrm{IND}}.$$

This modified loglikelihood function yields the same point estimators for R_0, β and σ. However, the confidence regions based on the inverted observed Fisher information are adjusted according to the residual correlation structure inherent in the data since

$$D^2_{R_0, \beta} \mathrm{LL}_{\mathrm{CA}} = \widehat{\mathrm{CA}}(\boldsymbol{D})^{-1} D^2_{R_0, \beta} \mathrm{LL}_{\mathrm{IND}},$$

such that confidence region of the estimates scales with the estimated cumulative autocorrelation, i.e.

$$\mathrm{CI}^{\mathrm{CA}}_{1-\alpha} \left(\hat{R}_0, \hat{\beta} \right) = (\hat{R}_0, \hat{\beta}) + \widehat{\mathrm{CA}}(\boldsymbol{D})^{\frac{1}{2}} \left(\mathrm{CI}^{\mathrm{IND}}_{1-\alpha} \left(\hat{R}_0, \hat{\beta} \right) - (\hat{R}_0, \hat{\beta}) \right)$$
$$= \left\{ (R_0, \beta) \left| \exists (R_0^*, \beta^*) \in \mathrm{CI}^{\mathrm{IND}}_{1-\alpha} \left(\hat{R}_0, \hat{\beta} \right) : (R_0, \beta) = (\hat{R}_0, \hat{\beta}) + \widehat{\mathrm{CA}}(\boldsymbol{D})^{\frac{1}{2}} \left((R_0^*, \beta^*) - (\hat{R}_0, \hat{\beta}) \right) \right. \right\}.$$

Analogously, within a Bayesian framework assuming uniform priors the posterior covariance scales with the cumulative autocorrelation as derived in Section 2.2.1

$$\text{Cov}\left(\hat{R}_0, \hat{\beta}\right)_{\text{CA}} = \widehat{\text{CA}}(\boldsymbol{D}) \, \text{Cov}\left(\hat{R}_0, \hat{\beta}\right)_{\text{IND}}.$$

This leads to larger confidence regions and variances in the case of $\text{CA}(\boldsymbol{D}) > 1$ and vice versa if $\text{CA}(\boldsymbol{D}) < 1$. Since the estimated cumulative autocorrelation is a function of (R_0, β) itself, it might appear convenient to compute the MLE using the likelihood function adjusted by $\widehat{\text{CA}}(\boldsymbol{D})(R_0, \beta)$ from the start, yielding a simultaneous estimation of CA and (R_0, β). However, the cumulative autocorrelation is only well-defined given the MLE $(\hat{R}_0, \hat{\beta})$ yielding the best fit and produces misleading results for other values of (R_0, β), since the raw residuals suffer from an imperfect model calibration in that case. Thus, since the loglikelihood takes negative values in many cases, this approach might maximise the CA instead of the original loglikelihood which yields counter-intuitive results. Thus, a two-step procedure as described above is necessary for the CA-based approach.

For the third method, which functions as benchmark method, we used a likelihood function corresponding to a model assuming the time series of residuals to follow an AR(1) process, i.e.

$$\log(D_t) - \log(X_t) = \phi_1 \left(\log\left(D_{t-1}\right) - \log\left(X_{t-1}\right)\right) + \sigma_t, \qquad \sigma_t \sim \mathcal{N}\left(0, \sigma\right) \text{ i.\,i.\,d},$$

which requires the estimation of an additional autocorrelation parameter ϕ_1. Hence, the corresponding loglikelihood function for the parameter vector $(R_0, \beta, \sigma, \phi_1)$ is

$$\text{LL}_{\text{AR}}\left(\boldsymbol{D} \,|\, R_0, \beta, \sigma, \phi\right) = \log f_{R_0, \beta, \sigma, \phi}^{\text{AR}}\left(D_1, \ldots, D_T\right)$$

$$= \sum_{t=0}^{T} \log \phi \left(\log(D_t); \mu = \log\left(X_t(R_0, \beta)\right) + \phi_1 \left(\log(D_{t-1}) - \log(X_{t-1})(R_0, \beta)\right), \sigma\right).$$

Also for this model corresponding confidence intervals were defined using the Hessian of the loglikelihood at the resulting MLE.

2.3.2 Simulation setup

To each simulated data set $\boldsymbol{D}_i$ $(i = 1, \ldots, M)$ and for each of the three inference methods $\bullet \in \{\text{IND,CA,AR}\}$ (independent residuals, cumulative autocorrelation, AR(1)-residuals) we computed the maximum likelihood estimators $\hat{\boldsymbol{\vartheta}}_{i,\bullet} = (\hat{R}_0, \hat{\beta})_{i,\bullet}$. To check the precision of each inference method we computed the componentwise squared error $\text{SE}_{i,\bullet}$ for each generated data set $\boldsymbol{D}_i$ and the corresponding MLE $(\hat{R}_0, \hat{\beta})_{i,\bullet})$ by

$$\text{SE}_{i,\bullet} = \left(\left(\hat{R}_{0\,i,\bullet} - R_0^*\right)^2, \left(\hat{\beta}_{i,\bullet} - \beta^*\right)^2\right)$$

with R_0^* and β^* being the true values from which the data was generated. The resulting squared error score (SES) was then computed as the averaged squared errors for each

inference method via

$$\text{SES}_\bullet = \frac{1}{M} \sum_{i=1}^{M} \text{SE}_{i,\bullet}.$$

To additionally examine the coverage of the estimated confidence regions we applied a probability integral transformation (PIT) (Czado et al., 2009). For each simulated data set $\boldsymbol{D}_i$ and each inference method $\bullet \in \{\text{IND,CA,AR}\}$ we computed the smallest confidence level $(1 - \alpha_{i,\bullet})$ such that the true value $\boldsymbol{\vartheta}^* = (R_0^*, \beta^*)$ lies within the corresponding confidence region $\text{CI}_{1-\alpha_{i,\bullet}}^\bullet$. Following our approach the $(1 - \alpha)$-confidence region around the point estimator $\hat{\boldsymbol{\vartheta}}_{i,\bullet} = (\hat{R}_0, \hat{\beta})_{i,\bullet}$ corresponds to the smallest region of probability $(1 - \alpha)$ from a 2-dimensional normal distribution with mean $(\hat{R}_0, \hat{\beta})_{i,\bullet}$ and covariance matrix $\hat{\Sigma}_{i,\bullet}$, where $\hat{\Sigma}_{i,\bullet}^{-1}$ is the observed Fisher information matrix. Based on the normality assumption we then computed the effective distance of the true value with respect to the MLE, i.e.

$$d_{i,\bullet} = \left(\hat{\boldsymbol{\vartheta}}_{i,\bullet} - \boldsymbol{\vartheta}^*\right) \hat{\Sigma}_{i,\bullet}^{-1} \left(\hat{\boldsymbol{\vartheta}}_{i,\bullet} - \boldsymbol{\vartheta}^*\right)^T$$

which follows a χ_2^2 distribution such that we obtained $(1 - \alpha_{i,\bullet})$ by

$$1 - \alpha_{i,\bullet} = F_{\chi_2^2}\left(d_{i,\bullet}\right),$$

where $F_{\chi_2^2}$ denotes the CDF of the χ_2^2 distribution. If the confidence regions generated by an inference method are accurate, then the $(1-\alpha_{i,\bullet})$'s should follow a uniform distribution on $[0, 1]$, which can be checked for each method $\bullet \in \{\text{IND,CA,AR}\}$ by plotting histograms of $(1 - \alpha_{i,\bullet})_{i=1,\ldots,M}$.

2.3.3 Results and interpretation

We fixed the time horizon to $T = 8 \cdot 52$ observations such that one data set consisted of weekly observations over the course of 8 years as for the rotavirus data described in Chapter 4.1.1. The initial condition was set to $(S(0), I(0), R(0)) = (49241, 16, 32743) \cdot 10^3$ which add up to 82 million – the size of the German population. The model parameters used for data simulation were set to $R_0^* = 5/3$ and $\beta^* = 0.3$. The standard error of the residual error process was set to $\sigma = 0.1$.

We ran three different data setups within our study, i.e. three different processes for simulating the residual errors. In the first setup we simulated $M = 1,000$ data sets using an AR(1) process with correlation coefficient $\phi = 0.6$ as the observational error process. For the second setup we used an ARMA(2,1) process with $\phi = (0.4, 0.4)$ and $\vartheta = -0.6$. In the third setup the observational errors were generated by a Brownian bridge process, where the standard error was set to $\sigma = 0.05$ to better relate the observational error to the underlying expected values. See Figure 2.1 for a trajectory from each of the three simulation setups.

The squared error scores (see Table 2.3.3) show that there is no notable difference in the precision of the three inference methods, where the MLEs yielded by the IND and CA
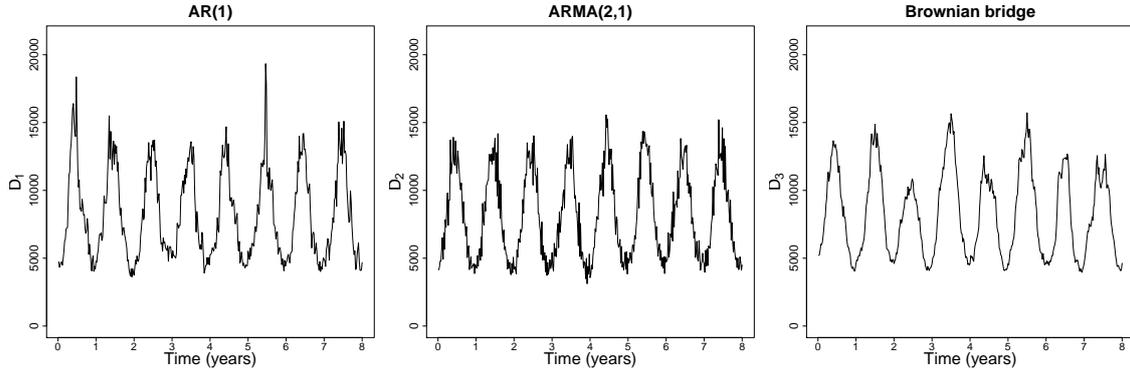
Figure 2.1: Sample data from each of the three simulation setups, which differ by the observational noise generating process: AR(1), ARMA(2,1) and Brownian bridge.

approach are identical by definition anyway. All point estimators are near the true values R_0^*, β^* from which the data was generated, while neither the AR-inference nor the other two methods (IND, CA) perform substantially better. This confirms the analytical results for the simple model examined in Section 2.2.1, which suggested asymptotic equality of the MLEs.

Residual	Inference method	
generating process	IND/CA	AR
AR(1)	(0.9339,0.9339)	(0.9339,0.9337)
ARMA(2,1)	(0.9340,0.9340)	(0.9340,0.9340)
BB	(0.9342,0.9335)	(0.9341,0.9344)

Table 2.2: Componentwise squared error score regarding the estimates for R_0 and β from the three simulation setups AR(1), ARMA(2,1) and Brownian bridge (BB). Note that the two inference methods IND and CA yield the same MLEs and therefore also the same scores.

However, differences arise when considering the respective uncertainty regions. The PIT-histograms of the three methods in Figure 2.2 show the accuracy of the estimated confidence regions. As we see in the first row of the figure, the AR-inference yields the best performance for setup 1 with the CA-inference being equally suitable. This is an expected result since the residual noise was generated exactly from an AR(1) process. Thereby, the distribution of the PIT-histogram is more or less skewed to the right for all inference methods, i.e. higher values of $1 - \alpha$, which means that the inference procedures yields confidence regions which tend to be too small, depending on the degree of skewness. When data are generated from a more complex ARMA(2,1) process (see second row of the figure), the CA-inference outperforms both other methods. This is due to the CA-inference being able to also capture partial autocorrelations of higher order as it searches for the best fitting ARMA process to fit the observed residuals, whereas the other two inference procedures only check for first order autocorrelation (AR) or do not consider autocorrelation at all

(IND). In the third setting, where the residual noise comes from a Brownian bridge process, the CA and the AR-method yield comparable results (see third row of Figure 2.2). The slightly better performance of the AR-inference may be due to the fact that the Brownian bridge is a Markov process such that all linear dependencies are already captured within the lag-1-autocorrelation.

Altogether, we see that the CA-method performs well compared to the established AR(1)-method, which directly incorporates the lag-1-autocorrelation into the estimation. More complex likelihood functions allowing for higher order correlation coefficients might outperform all methods mentioned above. However, the consideration of more parameters not only comes at the price of a higher computational effort, but also requires the specification of an adequate model, which is often not known in advance. In particular there is no obvious extension of the AR-modelling to count data settings, which would require an autoregressive process on $\mathbb{N}$. Thus, the CA-based approach is a robust method easily extendable to non-Gaussian data, as it only needs to estimate the autocorrelation of the residual process, but without making assumptions on the underlying process structure. Hence, it is also able to assess the autocorrelation structure within the data without estimating additional parameters during the actual model calibration. Also note the comparably bad performance of the inference assuming independent observations, which is commonly used for this kind of ODE-based transmission models. As we have seen using the CA-method, we obtain much more realistic confidence regions due to a simple adjustment of the classical likelihood function without further modifying the underlying model.

2.3.4 Investigating asymptotic normality

In a Bayesian framework the likelihood function on a prespecified support corresponds to the unnormalized posterior distribution when assuming uniform priors on that support. Thus, our frequentist inference setting may be easily translated into a Bayesian setting by assuming a uniform prior on a wide range around the presumed parameter values, e.g. $[0; 100]^2$ as support for (R_0, β). In many cases, it might be possible to define an improper prior $\pi(R_0, \beta) \propto 1$ on the whole set $\mathbb{R}^2$, if the resulting posterior is still integrable. However, for the employed ODE-model, for $R_0 \to \infty$ the model produces an endemic situation and thus the likelihood does not decay to zero which yields an improper posterior. Therefore, a bounded parameter space is necessary for well-definedness of the posterior.

Defining asymptotic confidence regions based on the MLE $\hat{\boldsymbol{\vartheta}}$ – which in the Bayesian case corresponds to the posterior mode – and the observed Fisher information as in equation (2.8) is equivalent to approximate the posterior by a normal distribution with mean $\hat{\boldsymbol{\vartheta}}$ and covariance based on the observed Fisher information. To check whether this yields a proper approximation to the actual posterior distribution of our model we computed a posterior sample using MCMC methods (see Section 3.2) from the true posterior and compared it to the asymptotic normal distribution. The posterior sample was obtained by sampling from a distribution proportional to the likelihood function $f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta})$ restricted to $[0; 100]^2$.
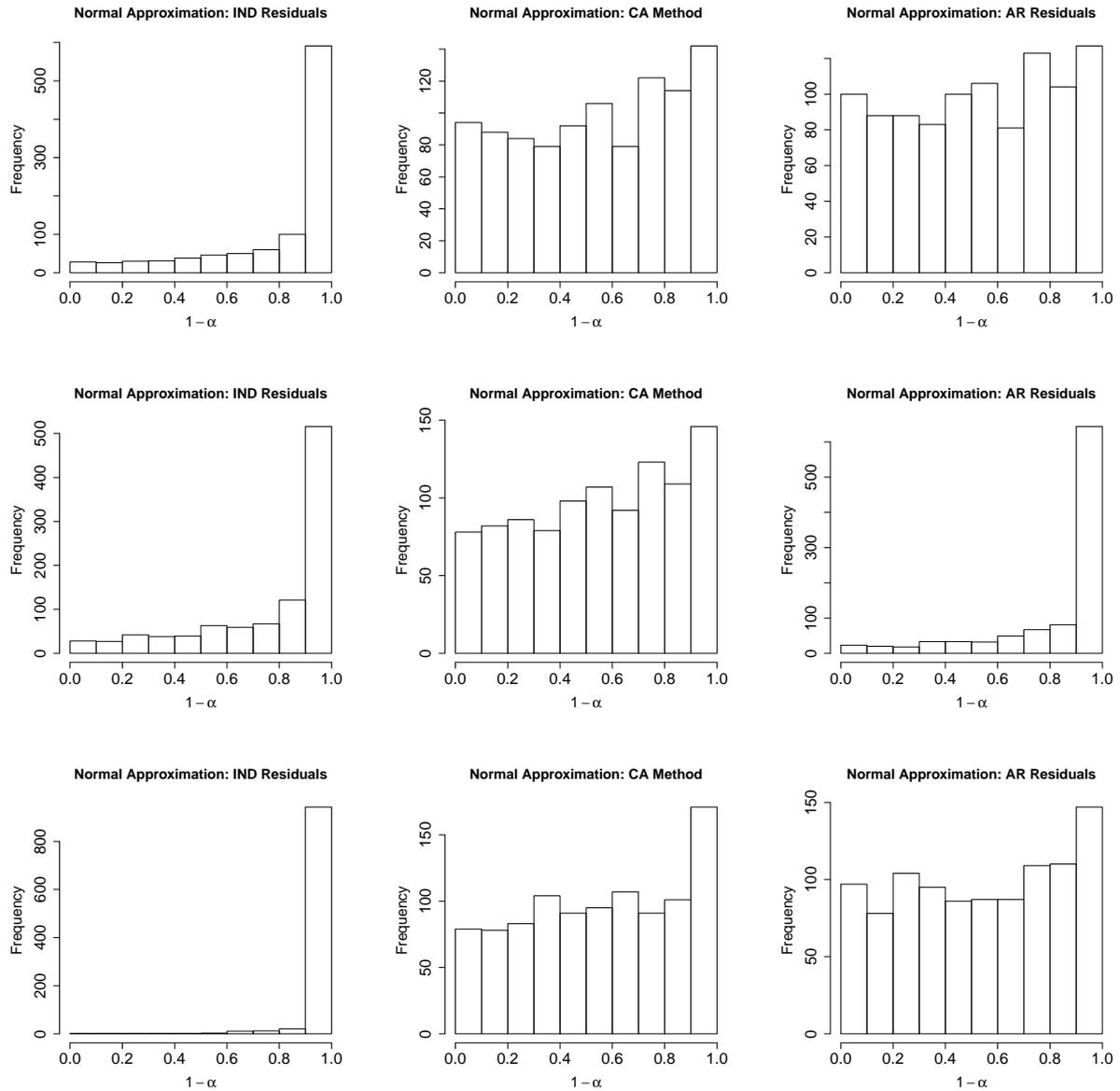
Figure 2.2: Histograms of the PIT confidence level $(1 - \alpha)$ for the methods IND, CA, AR based on 1,000 simulated data sets from each of three simulation setups. In the first row data were simulated with observational errors generated by an AR(1) process with $\phi = 0.6$, in the second row by an ARMA(2,1) process with $\phi = (0.4, 0.4)$ and $\theta = -0.6$, and for the third row simulation of observational errors was based on a Brownian bridge process.

Given one simulated data set and using the loglikelihood function LL_{AR} we computed the maximum likelihood estimator $\hat{\boldsymbol{\vartheta}}_{AR}$ and the observed Fisher information $\hat{\Sigma}_{AR}^{-1}$. We then computed a MCMC sample of length 100,000 with the first 10,000 being discarded as burn-in and only every 20th value taken for the final sample. We used $\hat{\Sigma}_{AR}$ as the covariance matrix of the centered Gaussian Metropolis-Hastings transition density and $\hat{\boldsymbol{\vartheta}}_{AR}$ as the starting point of the chain. For each entry $\boldsymbol{\vartheta}_i$ ($i = 1, \ldots, 4500$) from the final posterior sample we computed the smallest level $(1 - \alpha_i)$ such that $\boldsymbol{\vartheta}_i$ is contained in the approximated $(1 - \alpha_i)$-confidence region $\text{CI}_{1-\alpha_i}$ as explained in Equation (2.8). If the confidence regions are appropriate and the posterior sample is representative, the $(1 - \alpha_i)_{i=1,\ldots,4500}$ will be approximately uniformly distributed on $[0, 1]$. The histogram displayed in Figure 2.3 contains the PIT-Transform obtained by this procedure. For the data we used the dataset given in the leftmost graphic of Figure 2.1.
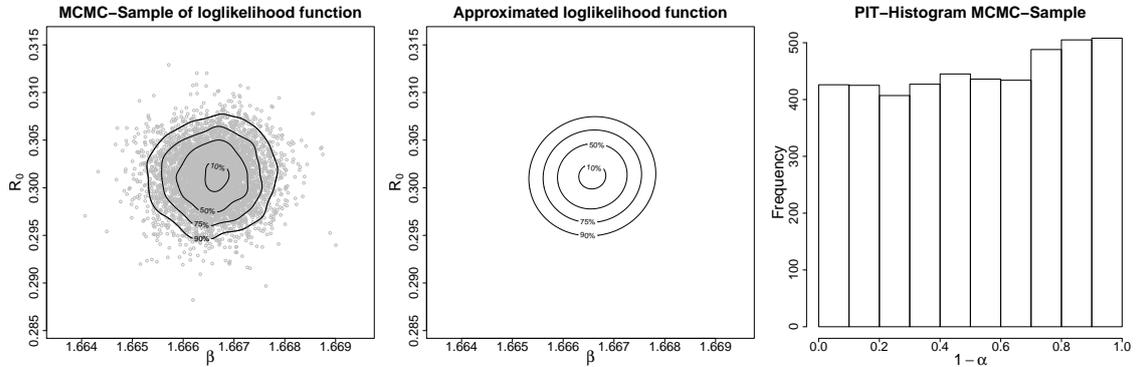
Figure 2.3: Comparison of the confidence regions generated from a posterior sample and from the approximate normal distribution. The data was generated using an AR(1) observational residual process. Posterior sample and normal approximation of the likelihood originated from the AR-inference method.

As we can see the asymptotic confidence regions give a good approximation to the confidence regions coming from the sample, which justifies the assumption of asymptotic normality. Furthermore, it is worth pointing out that computing such a large MCMC sample uses significantly more computational time compared to finding the posterior mode and its Hessian. Thus, a Gaussian approximation of the posterior while accounting for residual autocorrelation provides a faster inference procedure, while still yielding accurate results. This is especially relevant in more complex model settings as for our later rotavirus transmission application where finding the posterior mode alone is an involved computational effort.

Negative binomial observations

As a final analysis we tested our method in a discrete response setting with integer counts representing new weekly cases. This corresponds to the data setup according to the German

notification system which provides the disease burden through the absolute number of occurred cases (Krause et al., 2007). A discrete response setting was also applied to the rotavirus transmission models presented in Chapters 4 and 5. Here, we examine the effects of using the CA to adjust the likelihood function corresponding to a model, which assumes an independent negative binomial distribution of the weekly data given the expected weekly counts from the model.

To simulate data $\boldsymbol{D}$ using the expected number of weekly cases $\boldsymbol{X} = (X_t)_{t=0,\ldots,T}$ we constructed a discrete data analogue of an AR(1)-process by defining

$$D_t \sim \text{NegBin}\left(\mu_t = X_t\left(1 + \phi_1 \frac{D_{t-1}}{X_{t-1}}\right), \sigma\right)$$

where $\phi_1 \in [0,1)$ is an autocorrelation parameter and σ is the size parameter of the negative binomial distribution. In the hypothetical case of $X_t \equiv x > 0$ being constant for all t this results in a stationary process with values in $\mathbb{N}_0$, mean x, and autocorrelated components.

We again used three different models and corresponding loglikelihood functions to estimate the transmission parameters (R_0, β) . Firstly, we assumed the data to be independently distributed, i.e.

$$\text{LL}_{\text{IND}}\left(\boldsymbol{D}\,|R_0, \beta, \sigma\right) = \sum_{t=0}^{T} \log f_{\text{NegBin}}\left(D_t; \mu_t = X_t(R_0, \beta), \sigma\right)$$

with f_{NegBin} being the probability mass function of the negative binomial distribution. Based on LL_{IND} we define

$$\text{LL}_{\text{CA}}\left(\boldsymbol{D}\,|R_0, \beta, \sigma\right) = \widehat{\text{CA}}(\boldsymbol{D})^{-1}\,\text{LL}_{\text{IND}}\left(\boldsymbol{D}\,|R_0, \beta, \sigma\right),$$

where the cumulative autocorrelation $\widehat{\text{CA}}(\boldsymbol{D})$ given the first estimator $\hat{\boldsymbol{\vartheta}}_{\text{IND}} = (\hat{R}_0, \hat{\beta})_{\text{IND}}$ is again computed by fitting an $\text{ARMA}(p, q)$ to the time series of residuals. In contrast to the case of the lognormally distributed response described in Section 2.3.1, the raw residuals $R_t = D_t - X_t$ in the count data setting are far from a Gaussian distribution since the D_t are discrete, bounded from below and asymmetrically distributed. Therefore, we performed an Anscombe transformation in order to obtain approximately normally distributed residuals (Hardin and Hilbe, 2007) before fitting an ARMA-process to the resulting time series of residuals. For the necessary details on Anscombe transformation of residuals from a negative binomial distribution, see Section 4.2.2 where the same procedure was applied to the rotavirus data subject to the transmission model from our main application. As a third loglikelihood function we define

$$\text{LL}_{\text{AR}}\left(\boldsymbol{D}|R_0, \beta, \sigma\right) = \sum_{t=0}^{T} \log f_{\text{NegBin}}\left(D_t; \mu_t = X_t(R_0, \beta)\left[1 + \phi_1 \frac{D_{t-1}}{X_{t-1}(R_0, \beta)}\right], \sigma\right)$$

which assumes the same data distribution which was used for simulation of the data. Confidence regions were obtained as in the continuous space case by computing the observed Fisher information matrix, i.e. $-D^2_{R_0, \beta}\,\text{LL}(\boldsymbol{D}\,\big|\hat{R}_0, \hat{\beta}, \ldots)$.

We simulated 500 data sets and calculated maximum likelihood estimators and confidence regions for each of the three likelihood functions $LL_\bullet$, $\bullet \in \{IND, CA, AR\}$. The squared error score $SES_\bullet$ for estimation accuracy was also computed for each model. We again calculated the minimal level $(1 - \alpha_{i,\bullet})$ such that the true value $\boldsymbol{\vartheta}^*$ is included in the respective $(1 - \alpha_{i,\bullet})$-confidence regions. The corresponding PIT-histograms are shown in Figure 2.4.
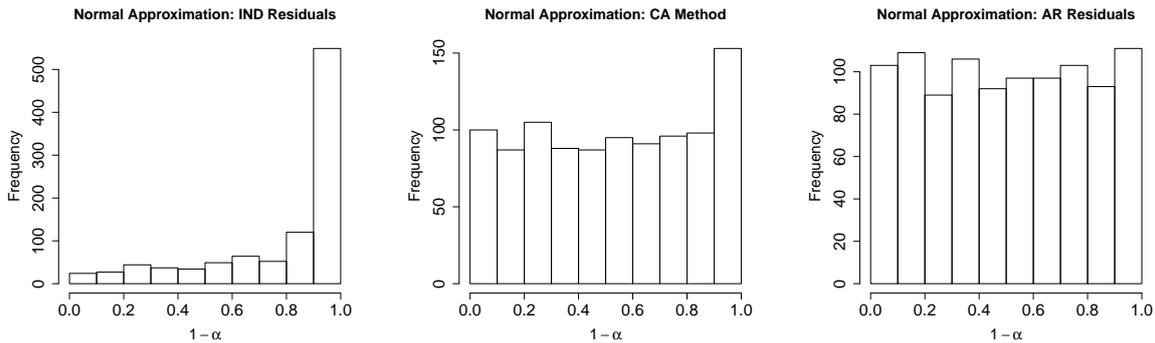
Figure 2.4: Histograms of the PIT confidence level $(1 - \alpha)$ corresponding to the loglikelihoods LL_{IND} (left), LL_{CA} (middle) and LL_{AR} (right) based on 500 simulated data sets. Data was simulated with observational errors generated by a discrete space autoregressive process with parameter $\phi_1 = 0.6$.

We observe that the three inference methods yield results consistent with what we saw in the continuous space model, i.e. poor performance of the model assuming independent residuals and comparable performance of the CA- and AR-model. Thereby, the AR-model based on the same distribution for data observations which was also used to simulate the data, performed slightly better. The estimation accuracy expressed via the square error scores was virtually equal for all methods with relative differences of $< 0.1\%$. Altogether, this suggests that the usage of the cumulative autocorrelation approach is also justified in the discrete response setting. Hereby, one has to be careful when estimating the autocorrelation structure of the residuals required for the adjustment procedure since in the discrete setting the residual process generally does not yield a Gaussian process. However, by performing a suitable transformation of the residuals one may apply the same technique as in the Gaussian residuals setting, i.e. via fitting a tractable process like ARMA to the residuals time series.

Studying higher order autoregressive processes within this setting remains difficult since a discrete space analogue of the $AR(p)$ process is not well established. This also speaks in favor of our CA-based approach as it also accounts for even higher order autocorrelation structures, where a direct modelling of the autocorrelation via the here presented discrete version of the AR(1)-process does not necessarily work well.

2.3.5 Conclusion

Our simulation study has shown that adjusting the likelihood function – or final confidence regions, respectively – based on the cumulative autocorrelation of the data yields a more realistic assessment of estimation uncertainty compared to assuming independence of all observations. This remains also true if the underlying mean structure is modelled as a parametric process with the respective parameters being subject of estimation, as it was the case for our here examined ODE-model. The presented adjustment approach was thereby motivated by analytical properties of AR-processes suggesting that the relevant ratio of the respective observed Fisher information could be identified with the cumulative autocorrelation of the observational residuals.

Although the simulation study was performed within a reduced model setting, not accounting for age structure and a more detailed representation of epidemiological aspects, we were still able to capture the essential problems that arise during inference for ODE-based transmission models. These are in particular multivariability and potential parameter collinearity, approximation of the confidence regions based on asymptotic normality, and dependency within the observational errors. Moreover, we obtained the same results when accounting for the discrete nature of the incidence data, which is important since in practice disease data is often available as absolute counts per time unit considered.

In all simulation setups our proposed approach performs comparable to or even better than direct likelihood methods based on models incorporating an autoregressive component. Due to its flexibility regarding the underlying autocorrelation structure in the data, even autoregressive and moving average components of higher order can be captured within the adjustment procedure. A major advantage of our method is that the computational burden remains comparably low, since there is no need for introducing additional parameters during the model formulation, which might otherwise even further increase complexity of already computational expensive models.

It should be pointed out that inference based solely on the likelihood assuming independent observations is the most commonly used approach, especially in applied disease transmission modelling based on ODEs (Atchison et al., 2010; Pitzer et al., 2009). As we have seen this method heavily underestimates uncertainty in the estimated model parameters. For example, less than 50% of the 90%-confidence regions computed within the simulation actually covered the true parameters (see left column of Figure 2.2) – and this fact arose not due to a lack of asymptotic normality. Considering the straightforward adjustment of the estimated variance our proposed approach offers an easy to apply inference procedure in addition to established methods in infectious disease modelling.

Certainly, AR-processes which we used within our study are not the only possibility to directly model the dependency structure of the observations. A time-discrete Markov chain, with its transition matrix being determined by the underlying model mean, thereby provides a general framework to model autocorrelated count data. Further advanced examples regarding a continuous state space include models based on stochastic differential equations, e.g. modelling data as discretely observed Ornstein-Uhlenbeck process around

a dynamic mean (Chung and Williams, 2014). Ornstein-Uhlenbeck processes might also be utilized if the time span between subsequent observations is not constant, e.g. not on a weekly basis but sporadically instead, or if the observed quantity, e.g. a diseases prevalence, can be observed continuously, although this is rarely the case.

Extensions of the CA-based approach include an alternative estimation of the CA itself. This could by improved by assuming a more general model class for the time series of residuals subject to the corresponding autocorrelations still being computable. Of course, straightforward estimation of the empirical autocorrelation is also possible, but one has to pay attention to the higher lag terms as these suffer from high variance. This might be avoided by truncating the autocorrelation sum at a prespecified maximum time lag or by deweighting higher order terms, e.g. by using the effective sample size instead (Thiébaux and Zwiers, 1984).

Moreover, the CA-based approach is also extendable into a Bayesian framework, which may be achieved by conducting a two-step inference approach, which we will further describe in Section 4.2 addressing the Bayesian inference for the full rotavirus transmission model. The basic concept of this procedure is to first compute the posterior mode of a given model and to estimate the residual cumulative autocorrelation, and then, in a second step, to adjust the likelihood function accordingly. In such a scenario reducing the likelihood impact has even further implications, since even the posterior mode is affected by the adjustment of the likelihood. Thus, accounting for potential residual autocorrelation might be even more important in a Bayesian setting as it affects the balance between likelihood and prior, i.e. the value of new insights versus existing knowledge.

Chapter 3

MCMC methods for high dimensional models based on ordinary differential equations

3.1 Introduction

In Bayesian inference the use of Markov chain Monte Carlo (MCMC) methods is widely acknowledged as a numerical method to sample from the posterior distribution. Depending on the structure of the model one can choose from a variety of possible MCMC algorithms, e.g. standard Metropolis-Hastings, Gibbs sampling, slice sampling, or sequential algorithms (Robert and Casella, 2004). Unfortunately, the necessary additional model information – like full posterior conditionals or derivatives – required for applying these specialized algorithms are rarely available in the case of infectious disease transmission models, which are often defined through a large system of ODEs, possibly supplemented by a stochastic observational model. This makes the construction of an efficient MCMC-based sampling algorithm as well as the computation of further model quantities, e.g. the marginal likelihood required for model selection, a difficult task.

Aim of this chapter is to provide MCMC methods for posterior sampling and marginal likelihood estimation which are particularly suitable for such ODE-based models. To do so we will present existing sampling approaches, but we will also propose new algorithms for marginal likelihood estimations as current methods often struggle with the statistical challenges inherent to such models.

In applied infectious disease epidemiology most modern dynamic transmission models often share two main characteristics which have to be accounted for when choosing an appropriate sampling algorithm for parameter inference from data (Pitzer et al., 2011; Atchison et al., 2010; Presanis et al., 2011; Birrell et al., 2011). Firstly, evaluating the likelihood for a specific parameter vector requires an excessive computational effort when compared to evaluating the prior probability or drawing new parameter candidates from a

proposal distribution within a sampling algorithm. This disproportion comes due to models being commonly formulated via a very large ODE system, which has to be numerically solved in order to calculate the likelihood of available incidence data. Suitable algorithms for posterior sampling or the computation of other model quantities therefore should require as few likelihood evaluations as possible. Secondly, many applied transmission models are based on a relatively large number of input parameters governing relevant aspects regarding the transmission and epidemiology of the pathogen, the demographic development of the population, its contact behaviour, and potential vaccination mechanisms. This leads to inference about high-dimensional parameters. The increasing model parameter dimension can lead to slow convergence rates of many algorithms. Furthermore, given the high dimension and the non-linearity inherent in dynamic models, it is very difficult to define a suitable proposal distribution for any random walk algorithm as the multivariate parameter correlations are almost impossible to assess in advance.

The structure of this chapter is as follows. Within the next section we will outline the basic concepts and pitfalls of MCMC sampling for ODE-based models illustrated by the general Metropolis-Hastings algorithm. Following this, an adaptive Metropolis-Hastings algorithm proposed by Haario et al. (2001) which addresses the above mentioned challenges of Bayesian inference for ODE-based models will be presented. In Section 3 concerning the estimation of the marginal likelihood, we will introduce an approach suggested by Chib and Jeliazkov (2001) which makes use of the MCMC posterior sample. After deriving an analytical assessment of the estimator in a high-dimensional Gausian posterior setting, we will propose new modified versions of the marginal likelihood estimation algorithm. Their performance will be investigated within a simulation-based study. The gained insights and potential alternatives will be discussed in Section 4.

3.2 Posterior computation for ODE-based models

When modelling disease transmission dynamics as in the case of the rotavirus transmission presented in Chapters 4 and 5, models are commonly defined using a set of ordinary differential equations to represent the disease spread within a given population. The aim of this ODE system is to deliver the expectation structure $\boldsymbol{\mu} = (\mu_t)_{t=1,\dots,T}$ for the available disease incidence time series data $\boldsymbol{D} = (D_t)_{t=1,\dots,T}$, which may be in the form of case counts or incidence rates. Thus, the likelihood of the incidence data is defined as a function $f(\boldsymbol{D} \,|\, \boldsymbol{\mu}, \boldsymbol{\vartheta_f})$ with

$$f(\boldsymbol{D} \,|\, \boldsymbol{\mu}, \boldsymbol{\vartheta_f}) = \prod_{t=1}^{T} f_t \left(D_t \,|\, \mu_t, \boldsymbol{\vartheta_f} \right),$$

where $t = 1, \dots, T$ denotes the index of each observation, e.g. a stratification by time, region, age or a combination of these, and f_t denotes the distribution of each observation D_t given its expectation μ_t and possible further parameters $\boldsymbol{\vartheta_f}$ defining this distribution. As already mentioned, the expectation structure $\boldsymbol{\mu}$ is derived by the solution of the ODE

system and therefore depends on the parameters $\boldsymbol{\vartheta}_O$ determining the ODEs. Thus, the full likelihood as a function of the total parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_O, \boldsymbol{\vartheta}_f)$, i.e.

$$f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) = f(\boldsymbol{D} \,|\, \boldsymbol{\mu}(\boldsymbol{\vartheta}_O), \boldsymbol{\vartheta}_f),$$

as well as the posterior $\pi(\boldsymbol{\vartheta} \,|\, \boldsymbol{D})$, which represents an updated version of the prior distribution $\pi(\boldsymbol{\vartheta})$ through Bayes' theorem

$$\pi(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}) \propto f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}),$$

are generally not available in closed form, but instead have to be evaluated using numerical ODE solvers such as the Runge-Kutta scheme (Press et al., 2007). This implies that additional information on the posterior shape, e.g. parameter covariance, or on the conditional distributions among components of $\boldsymbol{\vartheta}$ are rarely known in advance.

3.2.1 Posterior sampling by MCMC methods

A common approach to sample from an intractable distribution π, which can not be sampled directly from, is to construct a Markov chain $\boldsymbol{X} = (X_n)_{n \in \mathbb{N}}$ which has the target distribution π as its stationary distribution. For comprehensive overview of MCMC techniques please refer to e.g. Robert and Casella (2004). A very general method to construct a suitable Markov chain is the Metropolis-Hastings algorithm, which only requires the desired density to be known up to proportionality (Hastings, 1970). This makes the algorithm particularly useful for Bayesian statistics as the posterior density is commonly available only as unnormalized density $\pi(\cdot \,|\, \boldsymbol{D}) \propto f(\boldsymbol{D} \,|\, \cdot)\pi(\cdot)$. In the following $\pi(\cdot \,|\, \boldsymbol{D})$ will refer to the normalized posterior density, but we will point out the cases where it is sufficient to only have the unnormalized posterior density $f(\boldsymbol{D} \,|\, \cdot)\pi(\cdot)$ available. The Metropolis-Hastings algorithm is outlined in Algorithm 1 (Robert and Casella, 2004).

The idea of the algorithm thus is to generate a Markov chain $\boldsymbol{\Theta}$ by randomly proposing parameter vectors according to a (possibly state-dependent) proposal distribution q and accepting these proposals with probability depending on the unnormalized posterior density at the proposed points. One can immediately observe, that the acceptance probability increases for parameters yielding a higher posterior density such that the chain primarily moves along high posterior regions. The construction of the acceptance probability within Algorithm 1 thereby guarantees that the stationary distribution of the resulting chain is indeed the posterior distribution. Moreover, the space distribution of $\boldsymbol{\vartheta}^{(j)}$ converges to the stationary distribution for increasing chain length j. This convergence is the so-called ergodicity of the chain, which is governed by the following theorem, e.g. for a finite state

---

Algorithm 1: Metropolis-Hastings

Input: $\pi(\boldsymbol{\vartheta}|\boldsymbol{D})$: (unnormalized) posterior density

Input: $q(\boldsymbol{\vartheta},\cdot)$: transition proposal distribution, which may depend on the current
   state $\boldsymbol{\vartheta}$

Input: $\boldsymbol{\vartheta}^{(0)}$: initial parameter value of the chain

Input: J: length of chain

Output: $\boldsymbol{\Theta} = (\boldsymbol{\vartheta}_j)_{j=1,\dots,J}$: sample from the posterior distribution

for $j = 1$ **to** J **do**

   1. Propose a vector $\boldsymbol{\vartheta}^*$ using the proposal distribution $q(\boldsymbol{\vartheta}^{(j-1)},\cdot)$

   2. Evaluate the acceptance probability $\alpha = A\left(\boldsymbol{\vartheta}^{(j-1)},\boldsymbol{\vartheta}^*\right)$ with

   $$A\left(\boldsymbol{\vartheta}^{(j-1)},\boldsymbol{\vartheta}^*\right) = \min\left\{1, \frac{\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})\, q(\boldsymbol{\vartheta}^*,\boldsymbol{\vartheta}^{(j-1)})}{\pi(\boldsymbol{\vartheta}^{(j-1)}|\boldsymbol{D})\, q(\boldsymbol{\vartheta}^{(j-1)},\boldsymbol{\vartheta}^*)}\right\}$$

   3. Set $\boldsymbol{\vartheta}^{(j)} = \boldsymbol{\vartheta}^*$ with probability α and $\boldsymbol{\vartheta}^{(j)} = \boldsymbol{\vartheta}^{(j-1)}$ otherwise

end

---

space see Levin et al. (2008, Theorem 4.16)

Theorem 1. *Let g be a real-valued function defined on a finite state space Ω. If $\boldsymbol{X} = (X_n)$ is an irreducible Markov chain on Ω, then for any initial distribution $X_0 \sim \mu$ it holds*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} g(X_i) = \mathbb{E}_\pi\left[g(\cdot)\right] \quad \textit{almost surely,}$$

where π is the stationary distribution of $\boldsymbol{X}$.

Based on the ergodicity one can utilize the resulting Markov chain to estimate desired quantities such as the posterior mean, i.e. $g(x) = x$, or the cdf, i.e. $g_a(x) = \mathbb{I}_{\{(-\infty,a]\}}(x)$. A general issue when using Metropolis-Hastings is the choice of distribution q to propose new parameter candidates. Using state-independent proposal densities $q(\boldsymbol{\vartheta},\cdot) = q(\cdot)$ often leads to chains that get stuck at high posterior density points, because the probability to move away from those points can become very small, which yields a very inefficient algorithm.

To circumvent this problem, one often utilizes random walk proposals, i.e. a symmetric proposal distribution centred around the current parameter vector $\boldsymbol{\vartheta}$, e.g. $q(\boldsymbol{\vartheta},\cdot)$ is a multivariate normal or uniform distribution with mean $\boldsymbol{\vartheta}$. One advantage of such proposals is that due to the symmetry of q the acceptance probability only depends on the posterior values at the current and proposed vector. Furthermore, one can achieve a high acceptance probability by choosing the variance of the proposal distribution sufficiently small, which

secures that the chain indeed moves around the parameter space. However, choosing a too small variance can lead to very small step sizes, such that the chain moves very slowly in the parameter space and yields highly autocorrelated chain components. Another issue when sampling from high-dimensional parameter spaces is that the proposal q should account for possible parameter correlations, as otherwise it might frequently propose parameters within low posterior regions which again yields low acceptance rates. Therefore, a Metropolis-Hastings algorithm with an in-advance fixed symmetric proposal distribution is likely to work inefficiently in our context, as transmission parameters are often correlated, while this correlation structure as well as the overall posterior variance which might determine a suitable proposal step size are rarely known in advance. For more information on the Metropolis-Hastings algorithm including some historical remarks see e.g. Richey (2010).

3.2.2 Further MCMC sampling procedures and alternatives

A special case of the Metropolis-Hastings algorithm is the Gibbs sampler (Casella and George, 1992) which utilizes knowledge of the conditional posterior distribution $\pi\left(\boldsymbol{\vartheta}_1 \,|\, \boldsymbol{\vartheta}_2, \boldsymbol{D}\right)$ of two (or even more) parameter blocks $(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) = \boldsymbol{\vartheta}$ with respect to each other. The sampler then generates a Markov chain by alternately drawing from the conditional distributions, conditional to the current values of the remaining parameters, and updating the respective block of the drawn parameter. Thus, the chain always moves and depending on the correlation between the parameter blocks may even exhibit relatively large jumps in the parameter space. However, in the case of an ODE model the required conditional distributions are rarely given as even the full unnormalized posterior is rarely available in closed-form.

Another MCMC procedure is so-called Slice sampling (Neal, 2003) which aims to produce large step sizes and guaranteed acceptance by first uniformly sampling an unnormalized posterior density value $y^{(j)}$ from $[0, \pi(\boldsymbol{\vartheta}^{(j-1)}|\boldsymbol{D})]$ and then uniformly sampling a parameter from the region with unnormalized posterior density of at least $y^{(j)}$, i.e. from

$$S^{(j)} = \left\{ \boldsymbol{\vartheta} \,\middle|\, \pi\left(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}\right) \geq y^{(j)} \right\}.$$

However, computing the set $S^{(j)}$ requires a large number of posterior evaluations (and therefore repeated ODE solving) which grows exponentially with the dimension $d = \dim(\boldsymbol{\vartheta})$ of the underlying parameter space and is therefore practically infeasible in high-dimensional settings such as in our application.

Alternative approaches designed for MCMC sampling in high dimensions include the multiple-try method by Liu et al. (2000), which addresses the problem of a high dimension by proposing multiple points $\boldsymbol{\vartheta}_k^{(j)}$, for $k = 1, \ldots, K$ according to the the proposal $q(\boldsymbol{\vartheta}^{(j-1)}, \cdot)$ and choosing one candidate, i.e. k^*, among those at random using the unnormalized weights

$$w(\boldsymbol{\vartheta}_k^{(j)}, \boldsymbol{\vartheta}^{(j-1)}) = \pi\left(\boldsymbol{\vartheta}_k^{(j)}\,|\,\boldsymbol{D}\right) q(\boldsymbol{\vartheta}_k^{(j)}, \boldsymbol{\vartheta}^{(j-1)}) \lambda(\boldsymbol{\vartheta}_k^{(j)}, \boldsymbol{\vartheta}^{(j-1)}),$$

where λ is some freely chosen positive symmetric function, i.e. $\lambda(x,y) = \lambda(y,x)$. By sampling a new set of points $\tilde{\boldsymbol{\vartheta}}_k$, $(k = 1, \ldots, K-1)$ from $q(\boldsymbol{\vartheta}_{k*}^{(j)}, \cdot)$ the probability for accepting $\boldsymbol{\vartheta}_{k*}^{(j)}$ as the new chain value is then computed by

$$A(\boldsymbol{\vartheta}^{(j-1)}, \boldsymbol{\vartheta}_{k*}^{(j)}) = \min\left\{1, \frac{\sum_{k=1}^{K} w(\boldsymbol{\vartheta}_k^{(j)}, \boldsymbol{\vartheta}^{(j-1)})}{w(\boldsymbol{\vartheta}^{(j-1)}, \boldsymbol{\vartheta}_{k*}^{(j)}) + \sum_{k=1}^{K-1} w(\tilde{\boldsymbol{\vartheta}}_k, \boldsymbol{\vartheta}_{k*}^{(j)})}\right\}.$$

Therefore, one may choose q such that proposed points deviate stronger from the old parameter vector $\boldsymbol{\vartheta}^{(j-1)}$ which leads a larger mean step size, while still achieving high acceptance rates. However, the algorithm requires the in total $2K-1$ posterior evaluations for all proposed points $\boldsymbol{\vartheta}_k^{(j)}$ but also the reference points $\tilde{\boldsymbol{\vartheta}}_k$ which is a very large drawback in our setting, given that the majority of proposed parameter vectors are discarded in any case.

Further examples for generating informed chains include the approach inspired by the Langevin diffusion presented by Roberts and Tweedie (1996). This random walk algorithm makes use of the posterior log density's gradient $\nabla \log \pi(\cdot | \boldsymbol{D})$ as it attempts to move the chain into high posterior regions, instead of searching "blindly" according to a fixed proposal distribution. Unfortunately, the density's gradient is not directly available for the class of ODE models considered in this thesis, but would have to be computed numerically at each step. However, numerical computation requires even further costly evaluations of the posterior density, which makes this algorithm unsuitable for our purposes.

As an alternative to MCMC algorithms targeting the posterior density, a simulation based procedure to sample from the posterior distribution, e.g. suggested by McKinley (2009), does not require evaluation of the likelihood function. Such an approach simulates data from a prespecified parameter and defines its goodness of fit by measuring (according to some metric) the deviation of the simulated and observed data. This method is particularly useful, if many unobserved random variables are incorporated in the model, which would otherwise had to be marginalized out when doing likelihood-based inference. However, in the case of an ODE-based model the hidden variables, i.e. the incidence expectation structure, is in fact a deterministic process depending on the parameters to be estimated. Thus, hidden random variables do not exist and there is no gain in simulating from a parameter instead of directly evaluating its likelihood, which makes simulation-based approach unsuitable for our application.

3.2.3 An adaptive Metropolis-Hastings algorithm for posterior sampling

A possible candidate for posterior sampling in our setting is a so-called adaptive algorithm proposed by Haario et al. (2001), which is a modification of the classical Metropolis-Hastings algorithm. The intuition behind the approach is to adjust the transition proposal distribution q on the run according to the so far generated samples. A natural way of doing this using the standard Metropolis-Hastings algorithm is to first generate a rough

sample of the posterior using a possibly inefficient proposal q and afterwards define a new informed proposal function based on the generated sample to achieve a better performance in a second comprehensive run. The adaptive method is putting this approach to the next level as the proposal function q is adjusted at each step. Thus, the chain is continuously "learning" to more frequently propose parameter vectors within regions of high posterior density. Furthermore, the step size is adjusted to obtain a reasonable acceptance rate. This approach provides a decreasing autocorrelation of the MCMC sample and thus a faster convergence rate to the posterior distribution compared to the classical Metropolis-Hastings as it was shown via simulation by Haario et al. (2001). Therefore, the algorithm requires less costly evaluations of the posterior and likelihood function subject to generating a representative sample. The adaptive procedure works as illustrated in Algorithm 2.

The adaptive part of the algorithm is to be found in Step 1 of Algorithm 2. During an initial period ($j \leq K$) the transition kernel has a fixed covariance matrix Σ_0 which has to be provided in advance. Classical candidates for this could be the prior distribution's covariance matrix or, if available, an educated guess of the posterior's covariance matrix. After the initial period has passed the transition density is defined depending on the chain's past values. The reasoning for this definition is that the chain itself already yields a rough sample from the posterior and therefore provides information regarding the posterior covariance. Using this information ensures that the algorithm is more likely to search within areas of high posterior density and, thus, avoids improbable parameter vectors due to a badly defined proposal density.

It is important to point out that the adaptive algorithm does not yield a Markov chain. However, the ergodic properties are still fulfilled as shown by Roberts and Rosenthal (2007, Corollary 6) and Haario et al. (2001, Theorem 2) such that the resulting sample, if sufficiently large, is still representative for the posterior distribution .

The scaling parameter s_d controls the acceptance ratio of the chain and should secure an efficient search. According to Gelman et al. (1996) an efficient choice is given by

$$s_d = \frac{2.4^2}{d},$$

with d being the model parameter dimension. This scaling parameter achieves a large step size while still maintaining high acceptance rates for the case where the unscaled proposal covariance is of similar magnitude as the posterior covariance. There are approaches to also compute the scaling factor adaptively during the run to maximise the mean squared jump size $\mathbb{E}[\|\boldsymbol{\vartheta}^{(j+1)} - \boldsymbol{\vartheta}^{(j)}\|^2]$, see e.g. Pasarica and Gelman (2010). However, this method is so far only investigated for the univariate case.

Altogether, the adaptive algorithm proposed by Haario et al. (2001) works very efficiently compared to a classical Metropolis-Hastings procedure, as they have shown within a simulation-based analysis. The superior performance is most pronounced in cases where the initial proposal density does not match the correlation structure within the posterior distribution or where the proposed step sizes would be far too small or too large compared to the posterior variance. In these cases the adaptive method is able to considerably

---

Algorithm 2: Adaptive Metropolis-Hastings

Input: $\pi\left(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}\right)$: (unnormalized) posterior density

Input: Σ_0: initial covariance matrix of the proposal distribution

Input: s_d: scaling factor for the sample covariance matrix

Input: $\boldsymbol{\vartheta}^{(0)}$: initial parameter value of the chain

Input: J: length of the chain

Input: K: length of the initial period

Output: $\boldsymbol{\Theta} = \left(\boldsymbol{\vartheta}_j\right)_{j=1,\dots,J}$: sample from the posterior distribution

for $j = 1$ **to** J **do**

> 1. Set the Gaussian proposal density q_j to be
>
> $$q_j(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*) = \phi_{\boldsymbol{\vartheta},\Sigma_j}(\boldsymbol{\vartheta}^*)$$
>
> where ϕ is the multivariate normal density with
>
> $$\Sigma_j = \left\{ \begin{array}{ll} \Sigma_0 & , \text{if } j \leq K \\ s_d \widehat{\text{Cov}}\left(\boldsymbol{\vartheta}^{(0)}, \dots, \boldsymbol{\vartheta}^{(j-1)}\right) & , \text{if } j > K \end{array} \right.$$
>
> where $\widehat{\text{Cov}}\left(\boldsymbol{\vartheta}^{(0)}, \dots, \boldsymbol{\vartheta}^{(j-1)}\right)$ denotes to the empirical covariance and s_d is a predefined scaling factor.
>
> 2. Generate a candidate vector $\boldsymbol{\vartheta}^*$ using the proposal distribution $q_j(\boldsymbol{\vartheta}^{(j-1)}, \cdot)$
>
> 3. Compute the acceptance probability $\alpha = A\left(\boldsymbol{\vartheta}^{(j-1)}, \boldsymbol{\vartheta}^*\right)$ with
>
> $$A\left(\boldsymbol{\vartheta}^{(j-1)}, \boldsymbol{\vartheta}^*\right) = \min\left\{1, \frac{\pi\left(\boldsymbol{\vartheta}^*\,|\,\boldsymbol{D}\right)}{\pi\left(\boldsymbol{\vartheta}^{(j-1)}\,|\,\boldsymbol{D}\right)}\right\}$$
>
> 4. Set $\boldsymbol{\vartheta}^{(j)} = \boldsymbol{\vartheta}^*$ with probability α and $\boldsymbol{\vartheta}^{(j)} = \boldsymbol{\vartheta}^{(j-1)}$ otherwise

end

---

improve the proposal distribution on the run. This makes the presented procedure particularly suitable for Bayesian inference in complex ODE-based settings such as in the context of modelling dynamic disease transmission, because in these models the parameters to be estimated often exhibit high correlation and the marginal posterior variances are difficult to guess in advance. This would make the construction of a suitable fixed proposal distribution practically impossible.

3.3   Marginal likelihood and model selection

One aim within the task of estimating the rotavirus vaccine effectiveness trough a transmission model in Chapter 5 was to find the model among an ensemble of models $\mathcal{M}$ that is most suitable to explain the observed rotavirus incidence data.

As outlined in Chapter 2 in Bayesian statistics the validity of a model $i \in \mathcal{M}$ is measured by the marginal likelihood of the data $f^{(\mathcal{M})}(\boldsymbol{D}\,|\,i)$ given the model, i.e.

$$ f^{(\mathcal{M})}(\boldsymbol{D}\,|\,i) = \int f^{(i)}\left(\boldsymbol{D}\,\middle|\,\boldsymbol{\vartheta}^{(i)}\right) \pi^{(i)}\left(\boldsymbol{\vartheta}^{(i)}\right) d\boldsymbol{\vartheta}^{(i)}, $$

which can be interpreted as the expectation of the model specific likelihood function $f^{(i)}$ with respect to the prior distribution $\pi^{(i)}$.

Within this section we will present methods for computing or estimating this marginal likelihood for any given model. In this context, an estimation approach proposed by Chib and Jeliazkov (2001) based on the detailed-balance condition will be explained in more detail. Based on their method we will derive novel marginal likelihood estimation algorithms which better address the challenges of ODE-based models, i.e. a high parameter dimension as well as the computation effort for evaluating these models. The new algorithms' performance is later investigated within a simulation-based setting.

Since in the following we will talk about how to compute the marginal likelihood for a specific model, the index (i) will be dropped and the model specific likelihood and prior and posterior functions will be denoted by $f\left(\boldsymbol{D}\,|\,\cdot\right)$, $\pi(\cdot)$ and $\pi\left(\cdot\,|\,\boldsymbol{D}\right)$, respectively.

Similarly to sampling from the posterior distribution it is often difficult to compute the marginal likelihood of a given model analytically, since it involves an integration problem which can be solved only for certain classes of prior distribution and likelihood function, e.g. members of the exponential family and their conjugate prior. In other cases the marginal likelihood also has to be approximated by numerical methods (Kass and Raftery, 1995). Due to the same reasons stated in Section 3.2.3 we are interested in an approach that requires only few evaluations of the posterior. Thus, an estimation procedure which could utilize an existing posterior sample would be preferable.

Raftery et al. (2007) outlaid an approach, which is solely based on an existing posterior sample and does not at all require any additional posterior evaluations. Their procedure utilizes the harmonic mean identity of the marginal likelihood, i.e. the inverted expectation

of the inverted likelihood function with respect to the posterior

$$f^{(M)}(\boldsymbol{D}) = \left(\int f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta})^{-1} \pi(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}) d\boldsymbol{\vartheta} \right)^{-1}.$$

Thus, the marginal likelihood can easily be estimates using the likelihood values from a posterior sample, which yields the Monte Carlo estimate

$$\hat{f}^{(M)}(\boldsymbol{D}) = \left(K^{-1} \sum_{j=1}^{K} f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}_j)^{-1} \right)^{-1},$$

where $(\boldsymbol{\vartheta}_j)_{j=1,\dots,K}$ is a sample from the posterior. However, for many applications the resulting estimator has infinite variance and is often unstable as the integrand is large in regions where the posterior, i.e. the integrator, is small such that the corresponding sample is very sparse in those regions. While Raftery et al. (2007) propose parameter dimension reduction methods to stabilize the estimator, these methods are based on knowing the marginal posterior distribution of a reduced parameter vector, which is unfortunately not available within a complex ODE-based model.

A sampling-based estimator for the marginal model likelihood, as suggested by Friel and Pettitt (2008), utilizes the identity

$$\log f^{(M)}(\boldsymbol{D}) = \int_0^1 \mathbb{E}_t \left[\log f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) \right] dt,$$

where the expectation $\mathbb{E}_t$ is with respect to the so-called power posterior $\pi_t(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}) \propto f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta})^t \pi(\boldsymbol{\vartheta})$ (the concept of the power posterior was already utilized for the likelihood adjustment procedure presented in Chapter 2.2). An existing posterior sample may be used to approximate a part of the integral with $t \in [1 - \epsilon, 1]$, e.g. by using importance weights. However, for calculation of the remaining integral one needs additional samples from the power posteriors π_t for various $t \in [0, 1 - \epsilon]$, which would require the same computational effort as for generating the original posterior sample.

An approach that samples across several competing models and its corresponding parameter spaces is the reversible jump MCMC method presented by Green (1995). This method does not require marginal likelihood calculation as it directly estimates the joint distribution of models and parameters. Unfortunately, finding efficient proposal distribution to jump between different models are difficult to construct, especially when probability mass is very dense within the singular models (Brooks et al., 2003).

3.3.1 Marginal likelihood estimation based on the detailed-balance equation

One approach for computing the marginal likelihood based on the MCMC sampling output is the method presented by Chib and Jeliazkov (2001). Their approach utilizes the fact that

the marginal likelihood $f^{(M)}(\boldsymbol{D})$ is in fact just the normalizing constant of the unnormalized posterior density. This normalizing constant could easily be computed if we know the value of the normalized posterior density $\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})$ for some parameter vector $\boldsymbol{\vartheta}^*$ by re-arranging

$$f^{(M)}(\boldsymbol{D}) = \frac{f(\boldsymbol{D}|\boldsymbol{\vartheta}^*)\,\pi(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})}.$$

Hence, the problem can be reformulated into an estimation of $\pi(\boldsymbol{\vartheta}^*|D)$ for a any fixed $\boldsymbol{\vartheta}^*$ with positive posterior density. The estimation approach from Chib and Jeliazkov (2001) utilizes the reversibility of the Metropolis-Hastings transition kernel, i.e. for the acceptance probability A and transition proposal kernel q as defined in Algorithm 1 of Section 3.2.3 it holds

$$A(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)\,q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)\,\pi(\boldsymbol{\vartheta}|\boldsymbol{D}) = \pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})\,A(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta})\,q(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta}). \tag{3.1}$$

for all $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}^*$. Equation (3.1) is also know as the detailed-balance condition and is a sufficient condition for $\pi(\cdot|\boldsymbol{D})$ being a stationary distribution of the corresponding Markov chain generated by Algorithm 1 (see e.g. Levin et al. (2008, Prop. 1.19)). Integrating with respect to $\boldsymbol{\vartheta}$ and isolating $\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})$ then leads to the following representation

$$\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D}) = \frac{\int A(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)\,q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)\,\pi(\boldsymbol{\vartheta}|\boldsymbol{D})\,d\boldsymbol{\vartheta}}{\int A(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta})\,q(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta})\,d\boldsymbol{\vartheta}}, \tag{3.2}$$

which can be reformulated into

$$\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D}) = \frac{\mathbb{E}_{\pi(\cdot|\boldsymbol{D})}\left[A(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)\,q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)\right]}{\mathbb{E}_{q(\boldsymbol{\vartheta}^*, \cdot)}\left[A(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta})\right]},$$

where the upper and lower expectation is with respect to the the posterior distribution $\pi(\cdot|\boldsymbol{D})$ and the proposal distribution around $\boldsymbol{\vartheta}^*$, respectively. Using a posterior sample $\boldsymbol{\Theta} = (\boldsymbol{\vartheta}_j)_{j=1,\dots,J}$ and a sample $(\breve{\boldsymbol{\vartheta}}_n)_{n=1,\dots,N}$ from the proposal kernel $q(\boldsymbol{\vartheta}^*, \cdot)$, one can thus obtain a Monte-Carlo estimator for the pointwise posterior density

$$\hat{\pi}(\boldsymbol{\vartheta}^*|\boldsymbol{D}) = \frac{J^{-1}\sum_{j=1}^{J} A(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)\,q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)}{N^{-1}\sum_{n=1}^{N} A(\boldsymbol{\vartheta}^*, \breve{\boldsymbol{\vartheta}}_n)}. \tag{3.3}$$

Thus, after generating a posterior sample $\boldsymbol{\Theta} = (\boldsymbol{\vartheta}_j)_{j=1,\dots,J}$ via the standard Metropolis-Hastings algorithm (Algorithm 1) using proposal distribution q, the natural approach for estimating the posterior point density $\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})$ at some point $\boldsymbol{\vartheta}^*$ is to use Estimator(3.3) with the same proposal density q, the acceptance function A, the (possibly autocorrelated) MCMC sample $\boldsymbol{\Theta} = (\boldsymbol{\vartheta}_j)_{j=1,\dots,J}$ and a newly drawn sample from $q(\boldsymbol{\vartheta}^*, \cdot)$.

However, if one uses an adaptive algorithm, e.g. Algorithm 2, for posterior sampling the proposal distribution q_j is different for each iteration step j such that there is no immediate configuration for the subsequent marginal likelihood estimation using Estimator (3.3), which requires a fixed proposal distribution q. Fortunately, although it was not

stated in the original work by Chib and Jeliazkov (2001), the proposal distributions used for generating the posterior sample using MCMC and for marginal likelihood estimation do not necessarily have to coincide. In fact, as already mentioned the only necessary objects for evaluating the point posterior estimator (3.3) are (i) a posterior sample Θ, (ii) a proposal density q and the corresponding acceptance function A which fulfill the detailed-balance condition (3.1) and (iii) a sample $\breve{\Theta}$ from $q\left(\boldsymbol{\vartheta}^*, \cdot\right)$. Hereby, the choices for q and A are restricted by the detailed-balance whereas the sample $\breve{\Theta}$ obviously depends on the choice of q. However, q and A can be chosen independently from the posterior sample Θ in the sense that if Θ origins from a Metropolis-Hastings algorithm the proposal and acceptance function used for marginal likelihood estimation do not have to coincide with those used in the preceeding sampling algorithm (but they may of course). It should also be remarked, that this allows applying the marginal likelihood estimator (3.3) by using posterior samples which were generated via adaptive algorithms or any other sampling method.

This poses the question on which choice of q is best subject to an accurate estimation of the likelihood. Again, the only restriction is that for some acceptance function A the detailed-balance condition (3.1) with respect to the posterior density must be fulfilled, in which case A can be explicitly derived from q, i.e.

$$A\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*\right) = \min\left\{1, \frac{\pi\left(\boldsymbol{\vartheta}^* \mid \boldsymbol{D}\right) q(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta})}{\pi\left(\boldsymbol{\vartheta} \mid \boldsymbol{D}\right) q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)}\right\}. \tag{3.4}$$

Since the detailed-balance condition holds for a very wide class of possible proposal densities, e.g. including all symmetric proposal kernels with domain $\mathbb{R}^d$, this allows many options for choosing q. On that same subject, it might be interesting to examine, whether choices for q, which are considered optimal with respect to MCMC sampling, are also optimal with respect to estimating the marginal likelihood.

As Chib and Jeliazkov (2001) did not further consider the effect of different choices of proposal distributions q on the Estimator (3.3), in the following section we aim to look more closely on the analytical properties of the estimator within a Gaussian posterior setting. Especially, we will investigate the effect of rescaling the proposal density's covariance and consider the importance of at which point $\boldsymbol{\vartheta}^*$ the estimator is evaluated.

3.3.2 Properties of the marginal likelihood estimator in a Gaussian setting

Within this section we take a look at the properties of marginal likelihood estimation via the point posterior estimator (3.3) using the example of a multivariate Gaussian posterior distribution. In particular we will consider how rescaling the covariance of the proposal distribution q may affect the accuracy of the estimator, while also accounting for the dimension d of the underlying parameter space. As the possibility to use different proposal distributions for posterior sampling and posterior density estimation via (3.3) was not pursued by Chib and Jeliazkov (2001) these results provide new insights into how the

estimator's performance might be improved through an appropriate choice of the proposal distribution.

Firstly, note that the suggested estimator is biased, i.e. by letting U and L refer to the numerator and denominator of Estimator (3.3), respectively, we obtain

$$\mathbb{E}\left[\hat{\pi}(\boldsymbol{\vartheta}^{*}\,|D)\right] = \mathbb{E}\left[\frac{J^{-1}\sum_{j=1}^{J} A\left(\boldsymbol{\vartheta}_{j}, \boldsymbol{\vartheta}^{*}\right) q\left(\boldsymbol{\vartheta}_{j}, \boldsymbol{\vartheta}^{*}\right)}{N^{-1}\sum_{n=1}^{N} A\left(\boldsymbol{\vartheta}^{*}, \breve{\boldsymbol{\vartheta}}_{n}\right)}\right] = \mathbb{E}\left[\frac{U}{L}\right] \leq \frac{\mathbb{E}\left[U\right]}{\mathbb{E}\left[L\right]}, \qquad (3.5)$$

where the latter relation follows from Jensen's inequality (Cohn, 2013). Hereby, the degree of bias is determined by the variance of the denominator such that the bias vanishes for decreasing variance. In addition to minimizing the bias, we would like to have the overall variance of the estimator as small as possible, where the total variance depends on both the numerator's and the denominator's distribution and especially the corresponding variances. Thus, let us consider these single variances in a simple setting.

For that purpose we suppose that the posterior is d-dimensional normal distribution with expectation $\boldsymbol{\mu}_{P}$ and covariance matrix $\boldsymbol{\Sigma}_{P}$. The assumption of an Gaussian distribution is motivated by the asymptotic properties of the posterior distribution as given in Section 2.1.1. Furthermore, let q be a Gaussian proposal density based on an available posterior sample $\boldsymbol{\Theta}$, i.e. q takes the empirical covariance $\widehat{\mathrm{Cov}}(\boldsymbol{\Theta})$ multiplied by a scalar s_d as its covariance matrix $\boldsymbol{\Sigma}_{q}$. For a comprehensive posterior sample we have that $\widehat{\mathrm{Cov}}(\boldsymbol{\Theta}) \approx \boldsymbol{\Sigma}_{P}$, i.e. the posterior covariance. Thus, for the purpose of keeping this analysis simple we assume that the Gaussian density q has mean $\boldsymbol{\vartheta}$ and covariance matrix $\boldsymbol{\Sigma}_{q} = s_d\boldsymbol{\Sigma}_{P}$, i.e. q is given by the multivariate normal density

$$q(\boldsymbol{\vartheta}, \cdot) = \phi_{\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_{q}}(\cdot). \qquad (3.6)$$

However, in practice – since $\boldsymbol{\Sigma}_{P}$ is generally not known – one would take $\boldsymbol{\Sigma}_{q} = s_d\widehat{\mathrm{Cov}}(\boldsymbol{\Theta})$ instead.

Within this Gaussian setting we now want to investigate the variances of the numerator and denominator of the pointwise posterior density estimator $\hat{\pi}(\boldsymbol{\vartheta}^{*}\,|\boldsymbol{D})$ given by (3.3). Thereby, we are especially interested in the effect of the scaling parameter s_d. We recall that for efficiency regarding Metropolis-Hastings MCMC sampling a choice of $s_d^{(M)} = 2.4^2/d$ was considered optimal for scaling the proposal covariance (see Section 3.2.3).

For the assessment of the numerator's and denominator's variance the pointwise posterior density will be estimated at the posterior mode $\boldsymbol{\vartheta}^{*} = \boldsymbol{\mu}_{P}$ since taking a high density point for evaluation of the estimator was also recommended by Chib and Jeliazkov (2001). Again, in practice the posterior mode $\boldsymbol{\mu}_{P}$ is rarely known and other suitable points are used, e.g. the empirical posterior mode or the posterior sample mean, but for this illustration we assume that $\boldsymbol{\mu}_{P}$ is available.

First, let us consider the estimator's numerator U from equation (3.5). For clarification please note that the stochasticity of the random variable U is solely inherited by the random nature of the posterior sample draws $\boldsymbol{\vartheta}_{j}$ whereas A and q should be understood

as deterministic functions. Since $\boldsymbol{\vartheta}^*$ is the posterior mode and q is symmetric, for the acceptance probability A defined via (3.4) evaluated at $\boldsymbol{\vartheta}^*$, we obtain

$$A(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*) = 1 \quad \forall \boldsymbol{\vartheta}.$$

Furthermore, as $\boldsymbol{\Theta} = (\boldsymbol{\vartheta}_j)_{j=1,\ldots,J}$ is a (possibly autocorrelated) sample from the posterior distribution we have that

$$\boldsymbol{\vartheta}_j \sim \mathcal{N}(\boldsymbol{\vartheta}^*, \Sigma_P)$$

which yields

$$\boldsymbol{\Sigma}_P^{-\frac{1}{2}}(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*) \sim \mathcal{N}(0, \boldsymbol{I}_d) \quad \text{and hence} \quad (\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*)^T \boldsymbol{\Sigma}_P^{-1}(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*) \sim \chi_d^2,$$

where $\boldsymbol{I}_d$ refers to the d-dimensional unit matrix and χ_d^2 to the central chi-squared distribution with d degrees of freedom. Using this representation we can express the remaining factors of U by

$$
\begin{aligned}
q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*) &= \phi_{\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_q}(\boldsymbol{\vartheta}^*) = \phi_{0, \boldsymbol{\Sigma}_q}(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*) \\
&= \left((2\pi)^d |\boldsymbol{\Sigma}_q|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*)^T \boldsymbol{\Sigma}_q^{-1}(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*)\right) \\
&= \left((2\pi)^d |s_d \boldsymbol{\Sigma}_P|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2 s_d}(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*)^T \boldsymbol{\Sigma}_P^{-1}(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*)\right) \\
&= \left((2\pi s_d)^d |\boldsymbol{\Sigma}_P|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2 s_d} X_j\right)
\end{aligned}
\tag{3.7}
$$

with X_j being χ_d^2-distributed. Then for the variance of the single summands within the numerator U it holds

$$\text{Var}(A(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*) q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)) = \text{Var}(q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)) = \mathbb{E}\left[q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)^2\right] - \mathbb{E}[q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)]^2,$$

where the two expectations can be computed using the Laplace-transformation of the chi-squared distribution (Cairns, 2004), i.e.

$$
\begin{aligned}
\mathbb{E}[q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)] &= \left((2\pi)^d |\boldsymbol{\Sigma}_P|\right)^{-\frac{1}{2}} (1 + s_d)^{-\frac{d}{2}}, \\
\mathbb{E}\left[q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)^2\right] &= \left((2\pi)^d |\boldsymbol{\Sigma}_P|\right)^{-1} \left(s_d^2 + 2 s_d\right)^{-\frac{d}{2}},
\end{aligned}
\tag{3.8}
$$

which leads to

$$\text{Var}(A(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*) q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)) = \left((2\pi)^d |\boldsymbol{\Sigma}_P|\right)^{-1} g_d(s_d) \quad \text{with} \quad g_d(s) = (s^2 + 2s)^{-\frac{d}{2}} - (1 + s)^{-d}.$$

As $g_d(s)$ tends to infinite for $s \to 0$ and $\partial g_d / \partial s < 0$ for $s > 0$ (see Figure 3.1), it would be optimal to choose s_d as large as possible to achieve the smallest possible variance for the single summands. Furthermore, as the Figure shows any value $s < 0.5$ already leads to drastic increases of the variance, which is especially severe for high parameter space
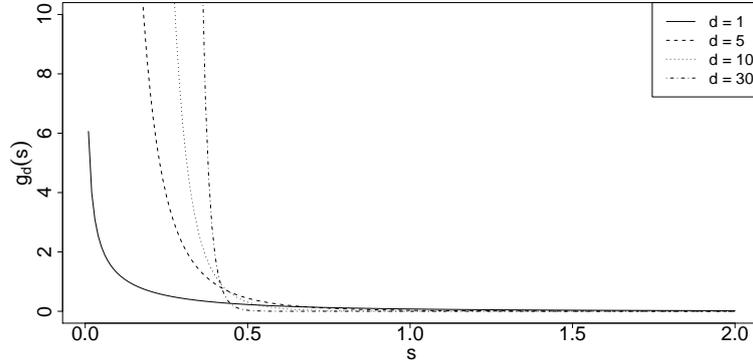
Figure 3.1: Plot of the function $g_d(s)$, which represents the impact of the covariance scalar s_d on the numerator's variance. The function is displayed for different model dimensions $d \in \{1, 5, 10, 30\}$ over the range $s \in [0.1, 2]$.

dimensions d. The dependence on the scaling factor s_d remains, when considering the overall variance of the numerator U, i.e.

$$
\begin{aligned}
\mathrm{Var}(U) &= \mathrm{Var}\left(J^{-1} \sum_{j=1}^{J} A\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right) q\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right)\right) \\
&= J^{-1} \left((2\pi)^d \left|\boldsymbol{\Sigma}_P\right|\right)^{-1} g_d(s_d) \left\{1 + 2 \sum_{j=1}^{J-1} \left(1 - \frac{j}{J}\right) \rho_j(s_d)\right\},
\end{aligned}
\tag{3.9}
$$

where $\rho_j(s_d)$ denotes the lag-j autocorrelation of the single summands

$$
\rho_j(s_d) = \mathrm{Cor}\left(A\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*\right) q\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*\right), A\left(\boldsymbol{\vartheta}_{1+j}, \boldsymbol{\vartheta}^*\right) q\left(\boldsymbol{\vartheta}_{1+j}, \boldsymbol{\vartheta}^*\right)\right),
\tag{3.10}
$$

which due to the construction of the proposal density q (Equation (3.6)) again depends on the choice of s_d. Since in practice the parameter vectors $\boldsymbol{\vartheta}_j$ often origin from an MCMC algorithm or other sampling procedures, the autocorrelations $\rho_j(s_d)$ may be non-zero in those cases. However, by sufficient thinning of the preceding MCMC sample (which lessens the autocorrelation among $\boldsymbol{\Theta}$) the effect of the $\rho_j(s_d)$ can be reduced. A remaining question concerns the effect of the choice of s_d on the autocorrelation which we will address later in Section 3.3.2. Assuming that $\rho_j(s_d)$ is not affected by s_d for all $j = 1, \ldots, J-1$, we obtain that it would be still best to choose s_d as large as possible if we aim to just minimize $\mathrm{Var}(U)$, which corresponds to a flat proposal distribution.

However, we also have to account for the denominator L of the posterior point estimator (3.5) on page 63. For that purpose, let $\breve{\boldsymbol{\Theta}} = (\breve{\boldsymbol{\vartheta}}_n)_{n=1,\ldots,N}$ refer to an independent sample drawn from the proposal kernel $q(\boldsymbol{\vartheta}^*, \cdot)$, i.e.

$$
\breve{\boldsymbol{\vartheta}}_n \overset{\text{i.i.d.}}{\sim} \mathscr{N}\left(\boldsymbol{\vartheta}^*, \boldsymbol{\Sigma}_q\right).
$$

For the denominator it holds that the acceptance probabilities $A\left(\boldsymbol{\vartheta}^*, \cdot\right) < 1$ almost surely, since $\boldsymbol{\vartheta}^*$ is the posterior mode. Recall that the posterior is assumed to be normally distributed with expectation $\boldsymbol{\mu}_P$ and covariance Σ_P and therefore

$$A\left(\boldsymbol{\vartheta}^*, \breve{\boldsymbol{\vartheta}}_n\right) = \frac{\phi_{\boldsymbol{\mu}_P, \Sigma_P}(\breve{\boldsymbol{\vartheta}}_n)}{\phi_{\boldsymbol{\mu}_P, \Sigma_P}(\boldsymbol{\vartheta}_*)} = \exp\left(-\frac{1}{2}\left(\breve{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\right)^T \Sigma_P^{-1}\left(\breve{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\right)\right).$$

Using analogue arguments as above one can show that

$$A\left(\boldsymbol{\vartheta}^*, \breve{\boldsymbol{\vartheta}}_n\right) = \exp\left(-\frac{s_d}{2}X_n\right),$$

with X_n being independent χ_d^2-distributed random variables and, again by applying the Laplace-transformation, that the variance of L in Equation (3.5) is given by

$$\mathrm{Var}(L) = \frac{1}{N}\left[(1 + 2s_d)^{-\frac{d}{2}} - (1 + s_d)^{-d}\right].$$

Although the variance of L is bounded by one due to the acceptance probabilities $A\left(\boldsymbol{\vartheta}^*, \bullet\right)$ being bounded, the variance of L^{-1} can get out of hand for large s_d. This is due to the expectation of L converging to zero for $s_d \to \infty$ as

$$\mathbb{E}\left[L\right] = \mathbb{E}_q\left[A\left(\boldsymbol{\vartheta}^*, \breve{\boldsymbol{\vartheta}}_1\right)\right] = \mathbb{E}\left[\exp\left(-\frac{s_d}{2}X_1\right)\right] = (1 + s_d)^{-\frac{d}{2}},$$

which follows from the independence of the sample $\breve{\boldsymbol{\Theta}}$. For instance, in the simple case of $N = 1$, i.e. $L = A\left(\boldsymbol{\vartheta}^*, \breve{\boldsymbol{\vartheta}}_1\right)$, we have that $\mathrm{Var}(L^{-1}) = \infty$ for $s_d > 1/2$. Since the overall posterior point estimator (3.3) is given by $\hat{\pi}(\boldsymbol{\vartheta}^*|D) = U \cdot L^{-1}$ this illustrates that s_d should not be chosen arbitrarily large, which one might suggest when only looking at the numerator U.

However, computing the overall variance of $\hat{\pi}(\boldsymbol{\vartheta}^*|D)$ analytically for arbitrary N remains a difficult task. Therefore, in order to assess the overall estimator's distribution, we numerically computed the mean and quantiles of the estimator's distribution for different scaling parameters s_d and high sample sizes for the posterior sample $J = |\boldsymbol{\Theta}|$ and the proposal density sample $N = \left|\breve{\boldsymbol{\Theta}}\right|$. To do so, we generated 10,000 samples of size $|\boldsymbol{\Theta}|$ and $\left|\breve{\boldsymbol{\Theta}}\right|$ from the χ_d^2-distribution to obtain 10,000 realizations of the numerator U and denominator L for each choice of s_d. Thus, for each s_d we also obtained 10,000 realizations of the overall estimator $\hat{\pi}(\boldsymbol{\vartheta}^*|D)$. To investigate the role of the underlying parameter space's dimension d we examined two scenarios in which we set $d = 5$ or $d = 30$, respectively. In this setting the posterior variance was set to $\boldsymbol{I}_d$, i.e. the unit matrix. The corresponding results are given in Figure 3.2.

These numerical estimates reveal that the scaling factor s_d determining the proposal kernel q has a crucial effect on the mean and variance of the posterior density estimator $\hat{\pi}(\boldsymbol{\vartheta}^*|D)$, especially when the space dimension d becomes large. According to these
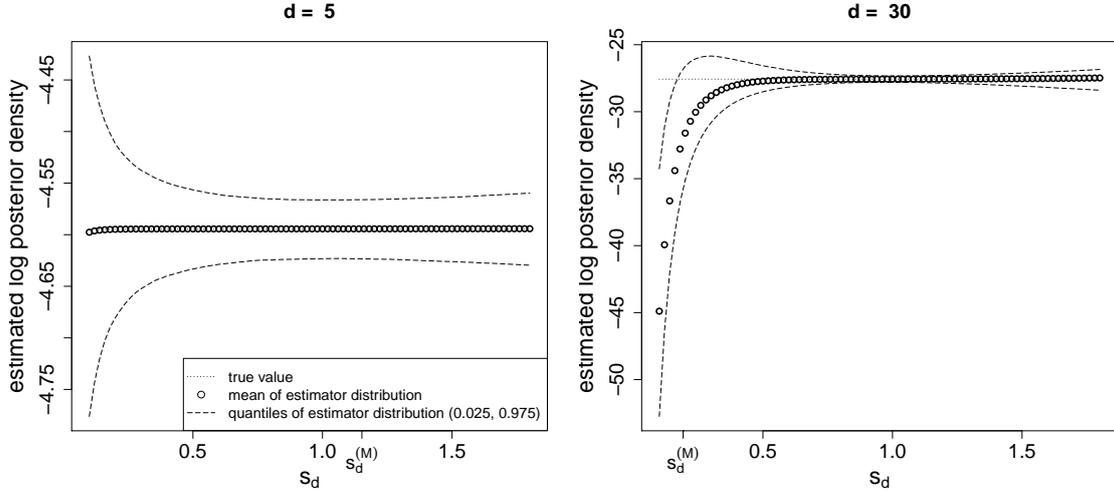
Figure 3.2: Distribution of the log posterior density estimator evaluated at the posterior mode with an underlying parameter space dimension d of 5 (left) or 30 (right), respectively. The transition proposal q used within the estimation algorithm were defined as the sample covariance matrix multiplied by different scalars s_d which ranged from 0.1 to 2.0 also including the MCMC-optimal scalar $s_d^{(M)} = 2.4^2/d$. Note the different scaling of the y-axes providing the log posterior density mode subject of estimation.

results choosing a scalar s_d near to one, i.e. the transition proposal covariance coincides with the covariance of the posterior sample, yields in any case good results providing an effectively unbiased estimator with small variance. As we explained above, too small values of s_d increases the numerator's variance whereas to large values increase the inverted denominator's variance.

Moreover, it becomes apparent that using the scalar $s_d^{(M)} = 2.4^2/d$, although providing efficiency with respect to Metropolis-Hastings sampling, leads to high variance and large bias of the posterior density estimator in a high dimensional setting when compared to not adjusting for dimensionality, i.e. setting $s_d = 1$. This phenomenon most likely originates from the skewness of the distribution of $q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)$ in these cases. However, recall that the proposal distribution q, and thus also the covariance scalar s_d, used within the estimator (3.3) can deviate from the one that was used for generating the posterior sample within a preceding MCMC run.

Accounting for correlation within the MCMC sample

So far we did not consider the autocorrelation structure of the underlying posterior sample Θ and its interplay with the choice of scaling factor s_d. Since in practice such posterior samples are commonly generated by MCMC procedures, these samples are by construction autocorrelated. Here, we would like to assess the impact on the presented estimator's

variance arising from the inherent correlation structure, which becomes apparent when looking at the numerator (3.9) due to the lag-j autocorrelations $\rho_j(s_d)$ of the summands.

Considering a very simple setting, we look more closely on these autocorrelations $\rho_j(s_d)$ and investigate its dependence on the scaling factor s_d. Thus, let us assume that the posterior corresponds to the standard normal distribution on $\mathbb{R}$ and $\boldsymbol{\Theta}$ represents a corresponding posterior sample. Further, we assume that the ordered sample $\boldsymbol{\Theta}$ originated from the stochastic process given by

$$\boldsymbol{\vartheta}_{j+1} \,|\, \boldsymbol{\vartheta}_j, \ldots, \boldsymbol{\vartheta}_1 \sim \mathcal{N}\left(\rho\boldsymbol{\vartheta}_j, 1 - \rho^2\right) \quad \text{with} \quad \boldsymbol{\vartheta}_1 \sim \mathcal{N}(0, 1) \tag{3.11}$$

i.e. an AR(1)-process with coefficient ρ. Thus, the higher order autocorrelations are determined by ρ as well as $\mathrm{Cor}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_{1+j}) = \rho^j$ and the Markov property is in place. Moreover, it can be easily checked that the posterior distribution, i.e. the standard normal distribution, is indeed the stationary distribution.

It should be pointed out that this dependency structure is only an approximation to that of a sample generated by an MH-algorithm (Algorithm 1), where the conditional density of the single chain components are given by

$$\boldsymbol{\vartheta}_{j+1} \,|\, \boldsymbol{\vartheta}_j, \ldots, \boldsymbol{\vartheta}_1 \sim (1 - P_A(\boldsymbol{\vartheta}_j))\, \delta_{\{\boldsymbol{\vartheta}_j\}} + A(\boldsymbol{\vartheta}_j, \cdot)q(\boldsymbol{\vartheta}_j, \cdot) \tag{3.12}$$

where $P_A(\boldsymbol{\vartheta}_j) = \mathbb{P}\left[\boldsymbol{\vartheta}_{j+1} \neq \boldsymbol{\vartheta}_j \,|\, \boldsymbol{\vartheta}_j\right]$ denotes the probability of moving away from $\boldsymbol{\vartheta}_j$, i.e.

$$P_A(\boldsymbol{\vartheta}_j)) = \int\limits_{\mathbb{R}} A(\boldsymbol{\vartheta}_j, x)q(\boldsymbol{\vartheta}_j, x)dx$$

and δ refers to the Dirac delta function. However, it appears to be infeasible to obtain closed-form terms for the relevant autocorrelations $\rho_j(s_d)$ given by Equation (3.10), and even for $\mathrm{Cor}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_{1+k})$ itself, according to the conditional distribution (3.12).

Therefore, for the purpose of investigating the impact of autocorrelation within the posterior sample, we will continue with the simpler but yet reasonable model for the correlation structure given by (3.11). In order to obtain $\rho_1(s_d)$ we first have the calculate $\mathbb{E}[q(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*) \cdot q(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*)]$ where $q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)$ is given by

$$q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*) = \exp\left(-\frac{1}{2s_d}\left(\boldsymbol{\vartheta}^2\right)\right),$$

by analogy with (3.7) on page 64 with $\boldsymbol{\vartheta}^* = 0$ being the posterior mode. Thus, we obtain

$$\begin{aligned}
\mathbb{E}\left[q(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*) \cdot q(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*)\right] &= \mathbb{E}\left[\exp\left(-\frac{1}{2s_d}\left(\boldsymbol{\vartheta}_1^2 + \boldsymbol{\vartheta}_2^2\right)\right)\right] \\
&= \mathbb{E}\left[\exp\left(-\frac{1}{2s_d}\left(\boldsymbol{\vartheta}_1^2 + \left(\rho\boldsymbol{\vartheta}_1 + \sqrt{1-\rho^2}z\right)^2\right)\right)\right],
\end{aligned} \tag{3.13}$$

where z denotes a standard-normally distributed random variable independent from $\boldsymbol{\vartheta}_1$, which follows from the conditional distribution of $\boldsymbol{\vartheta}_2$ with respect to $\boldsymbol{\vartheta}_1$. Thus, the above

expectation of the product can be computed by subsequently integrating with respect to z and $\boldsymbol{\vartheta}_1$, i.e.

$$\mathbb{E}\left[q\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*\right) \cdot q\left(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*\right)\right]$$

$$= (2\pi s_d)^{-1} \int_{\mathbb{R}} \exp\left(-\frac{1}{2s_d}\boldsymbol{\vartheta}_1^2\right) \int_{\mathbb{R}} \exp\left(-\frac{1-\rho^2}{2s_d}\left(\frac{\rho}{\sqrt{1-\rho^2}}\boldsymbol{\vartheta}_1 + z\right)^2\right) \phi(z)\phi(\boldsymbol{\vartheta}_1)\,dz\,d\boldsymbol{\vartheta}_1$$

$$= (2\pi s_d)^{-1} \int_{\mathbb{R}} \exp\left(-\frac{1}{2s_d}\boldsymbol{\vartheta}_1^2\right)\left(1 + \frac{1-\rho^2}{s_d}\right)^{-\frac{1}{2}} \exp\left(-\frac{\rho^2}{2\left(s_d + 1 - \rho^2\right)}\boldsymbol{\vartheta}_1^2\right) \phi(\boldsymbol{\vartheta}_1)\,d\boldsymbol{\vartheta}_1$$

$$= (2\pi s_d)^{-1}\left(1 + \frac{1-\rho^2}{s_d}\right)^{-\frac{1}{2}} \int_{\mathbb{R}} \exp\left(-\frac{s_d(1+\rho^2) + 1 - \rho^2}{2s_d\left(s_d + 1 - \rho^2\right)}\boldsymbol{\vartheta}_1^2\right) \phi(\boldsymbol{\vartheta}_1)\,d\boldsymbol{\vartheta}_1$$

$$= (2\pi s_d)^{-1}\left(1 + \frac{1-\rho^2}{s_d}\right)^{-\frac{1}{2}}\left(1 + \frac{s_d(1+\rho^2) + 1 - \rho^2}{s_d\left(s_d + 1 - \rho^2\right)}\right)^{-\frac{1}{2}}.$$

By using the expectation and variance computed earlier in (3.8), it follows for the correlation

$$\mathrm{Cor}\left(q\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*\right), q\left(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*\right)\right) = \frac{\mathbb{E}\left[q\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*\right) \cdot q\left(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*\right)\right] - \mathbb{E}\left[q\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*\right)\right]\mathbb{E}\left[q\left(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*\right)\right]}{\sqrt{\mathrm{Var}\left(q\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*\right)\right)\mathrm{Var}\left(q\left(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*\right)\right)}}$$

$$= \frac{\left(\frac{s_d + 1 - \rho^2}{1 + s_d} \cdot \frac{s_d^2 + s_d(1+\rho^2) + (1+s_d)(1-\rho^2)}{(1+s_d)(s_d + 1 - \rho^2)}\right)^{-\frac{1}{2}} - 1}{\left(1 + s_d^{-1}\right)\left(1 + 2s_d^{-1}\right)^{-\frac{1}{2}} - 1}.$$

$$(3.14)$$

For validation it can be easily checked that the autocorrelation $\mathrm{Cor}(q(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*), q(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}^*))$ given by Equation (3.14) is zero or one, given the (artificial) case of the correlation ρ of the MCMC chain components being zero or one, respectively. Furthermore, since we assumed the autocorrelation structure of the MCMC chain to corresponds to that of an AR(1)-process, the lag-j autocorrelations

$$\rho_j(s_d) = \mathrm{Cor}(q(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}^*), q(\boldsymbol{\vartheta}_{1+j}, \boldsymbol{\vartheta}^*)) \tag{3.15}$$

can be obtained by plugging in ρ^j instead of ρ. Thus, due to continuity of (3.14) with respect to ρ the higher lag autocorrelation converges to the zero for $k \to \infty$ as one would expect. Computing the corresponding approximated autocorrelation for more complex models, e.g. a higher dimension d with $\mathrm{Cor}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) = \boldsymbol{\Sigma}_C$ being some $d \times d$ correlation matrix, could be certainly done using a similar approach and presumable yield similar results. However, for the purpose of illustration we will stick to the one-dimensional case to study the impact of the choice of scalar s_d.

Using the above representation (3.14) we are able compute the cumulative autocorrelation which is the eventual adjustment factor of the estimator's variance as given in (3.9), i.e.

$$1 + 2 \sum_{j=1}^{J-1} \left(1 - \frac{j}{J}\right) \rho_j(s_d). \tag{3.16}$$

The cumulative autocorrelation for different values of s_d assuming different autocorrelations ρ of the underlying posterior sample is provided in Figure 3.3. As we see the adjustment
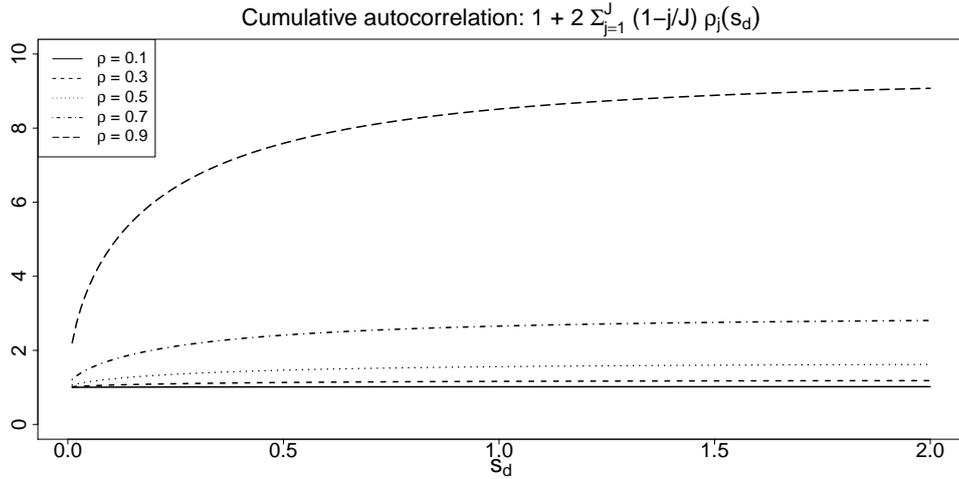
Figure 3.3: Cumulative autocorrelation (3.16) of the summands within the estimator's numerator (3.9) based on the autocorrelation derived in (3.14). The cumulative autocorrelations are evaluated for different autocorrelations within the underlying posterior sample ($\rho \in \{0.1, 0.3, 0.5, 0.9\}$) over the range of $s_d \in [0.1, 2]$

.

factor (3.16) is smallest for small values of s_d, independently of the underlying correlation ρ. Thereby, the cumulative autocorrelation of course does not fall below one for any s_d. Furthermore, the impact of the scaling factor s_d is more pronounced for higher correlation coefficients ρ, where the cumulative rapidly decreases for $s_d < 0.5$. However, recall that this is also the range in which the overall estimator's variance is considerably higher and heavy bias is induced for higher dimensions as seen in Figure 3.2.

One important insight taken from this analysis is that the cumulative autocorrelation appears to be monotonically increasing in s_d and also converging, where the asymptotic behaviour may be assessed by a Taylor approximation of $q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*)$ as

$$\exp\left(-\frac{1}{2s_d}\boldsymbol{\vartheta}^2\right) \approx 1 - \frac{1}{2s_d}\boldsymbol{\vartheta}^2$$

for very large s_d and therefore for the lag-j autocorrelation (3.15) it holds

$$\rho_j(s_d) = \mathrm{Cor}\left(\exp\left(-\frac{1}{2s_d}\boldsymbol{\vartheta}_1^2\right), \exp\left(-\frac{1}{2s_d}\boldsymbol{\vartheta}_{1+j}^2\right)\right) \stackrel{s_d \to \infty}{\longrightarrow} = \mathrm{Cor}\left(\boldsymbol{\vartheta}_1^2, \boldsymbol{\vartheta}_{1+j}^2\right) = \rho^{2j},$$

where the latter equation follows directly from applying the conditional distribution of $\boldsymbol{\vartheta}_{1+j}$ with respect to $\boldsymbol{\vartheta}_1$ as done in (3.13). Thus, as the important take-away we can assure that the autocorrelations $\rho_j(s_d)$ within the summands of U are lower than the autocorrelations within the underlying posterior sample for every choice of s_d. In that regard again note, that the autocorrelation within posterior samples resulting from MCMC methods can be decreased in advance by sufficient thinning of the sample, although this effectively reduces the size $|\boldsymbol{\Theta}| = J$ of the available sample. Altogether this analysis shows, that drastic negative effects on the overall estimator's variance $\mathrm{Var}(\hat{\pi}(\boldsymbol{\vartheta}^*|D))$ due to possibly 'bad' choices of the scaling factor s_d can be excluded, when considering only the autocorrelation adjustment (3.16).

Choice of evaluation point

Another important aspect is the choice of point $\boldsymbol{\vartheta}^*$ for evaluation of the posterior density estimator $\hat{\pi}(\boldsymbol{\vartheta}^*|\boldsymbol{D})$. According to Chib and Jeliazkov (2001) any high density point is suitable for evaluation. To comprehend the impact of choosing a low density point, suppose we choose for evaluation of the estimator not the posterior mode but some other point $\boldsymbol{\vartheta}^* \neq \boldsymbol{\mu}_P$. Considering the numerator U in this case we have that $A(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*) < 1$ for many points of the posterior sample $\boldsymbol{\Theta} = (\boldsymbol{\vartheta_j})_{j=1,\dots,J}$, e.g. considering our Gaussian setting and assuming the posterior covariance $\boldsymbol{\Sigma}_P = \sigma I_d$ being a scaled unit matrix we obtain

$$
\begin{aligned}
A\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right) = {}& \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j - \boldsymbol{\mu}_P\|^2 \leq \|\boldsymbol{\vartheta}^* - \boldsymbol{\mu}_P\|^2\}} \exp\left(-\frac{1}{2\sigma}\left(\|\boldsymbol{\vartheta}^* - \boldsymbol{\mu}_P\|^2 - \|\boldsymbol{\vartheta}_j - \boldsymbol{\mu}_P\|^2\right)\right) \\
& + \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j - \boldsymbol{\mu}_P\|^2 > \|\boldsymbol{\vartheta}^* - \boldsymbol{\mu}_P\|^2\}}.
\end{aligned}
$$

where $\|\cdot\|$ refers to the Euclidian metric. Secondly, $\boldsymbol{\vartheta}^*$ might lie within a region in which the posterior sample is not particularly dense compared to high posterior regions such that $q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)$ is close to zero for many points of the sample which again depends on the choice of s_d, i.e.

$$
q\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right) = (2\pi s_d \sigma)^{-\frac{d}{2}} \exp\left(-\frac{1}{2s_d\sigma}\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*\|^2\right),
$$

where

$$
\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*\|^2 \sim \chi^2(d, \|\boldsymbol{\vartheta}^* - \boldsymbol{\mu}_P\|^2)
$$

follows a non-central χ^2 distribution (Abramowitz and Stegun, 1964). Without loss of generality let us suppose that $\boldsymbol{\mu}_P = 0$. Then for the summands within the numerator U it

follows

$$
(2\pi s_d \sigma)^{\frac{d}{2}} A\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right) q\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right)
$$

$$
= \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j\|^2 \leq \|\boldsymbol{\vartheta}^*\|^2\}} \exp\left(-\frac{1}{2\sigma}\left(\|\boldsymbol{\vartheta}^*\|^2 - \|\boldsymbol{\vartheta}_j\|^2\right) - \frac{1}{2s_d\sigma}\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*\|^2\right)
$$

$$
+ \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j\|^2 > \|\boldsymbol{\vartheta}^*\|^2\}} \exp\left(-\frac{1}{2s_d\sigma}\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*\|^2\right)
$$

$$
\stackrel{s_d \neq 1}{=} \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j\|^2 \leq \|\boldsymbol{\vartheta}^*\|^2\}} \exp\left(\frac{s_d}{2(1-s_d)\sigma}\|\boldsymbol{\vartheta}^*\|^2\right) \exp\left(-\frac{1-s_d}{2s_d\sigma}\left\|\boldsymbol{\vartheta}_j - \frac{1}{1-s_d}\boldsymbol{\vartheta}^*\right\|^2\right)
$$

$$
+ \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j\|^2 > \|\boldsymbol{\vartheta}^*\|^2\}} \exp\left(-\frac{1}{2s_d\sigma}\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*\|^2\right),
$$

or for the other case of $s_d = 1$

$$
(2\pi s_d \sigma)^{\frac{d}{2}} A\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right) q\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*\right)
$$

$$
\stackrel{s_d = 1}{=} \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j\|^2 \leq \|\boldsymbol{\vartheta}^*\|^2\}} \exp\left(\frac{1}{2\sigma}\left(\|\boldsymbol{\vartheta}^*\|^2 - 2\langle\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta}_j\rangle\right)\right)
$$

$$
+ \mathbb{I}_{\{\|\boldsymbol{\vartheta}_j\|^2 > \|\boldsymbol{\vartheta}^*\|^2\}} \exp\left(-\frac{1}{2\sigma}\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*\|^2\right).
$$

Therefore, the terms $A(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*)q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*) \in [0, (2\pi s_d\sigma)^{\frac{-d}{2}}]$ are within the same range as in the simple case $\boldsymbol{\vartheta}^* = \boldsymbol{\mu}_P$ reaching its maximum when $\boldsymbol{\vartheta}_j = \boldsymbol{\vartheta}^*$ and decay exponentially with increasing distance $\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}^*\|^2$.

Also recall that the elements $\boldsymbol{\vartheta}_j$ of the posterior sample are normally distributed around $\boldsymbol{\mu}_P$. Thus, the numerator U is heavily affected by the parameter samples within $\boldsymbol{\Theta}$ which are by chance near to $\boldsymbol{\vartheta}^*$ (in the sense of q). As the posterior sample might not be particularly dense around $\boldsymbol{\vartheta}^*$, especially if the relative distance between evaluation point and posterior mode $(1/\sigma)\|\boldsymbol{\vartheta}^* - \boldsymbol{\mu}_P\|$ is large, this leads to U being of high variance since it is only affected by few points in this case. Unfortunately, further analytical computation of $\text{Var}(U)$ remains difficult in this case.

Moreover, the problem of denseness is even more prominent in high dimensional settings as the distribution of the corresponding transition proposal density may be heavily skewed if s_d is badly chosen, i.e. too small, as seen in Figure 3.2. In this case, U may be heavily overestimated if the evaluation point $\boldsymbol{\vartheta}^*$, even if chosen within a high posterior region, is by chance very close to one of the MCMC sample components $\boldsymbol{\vartheta}_j \in \boldsymbol{\Theta}$. Besides choosing s_d appropriately, an additional approach to avoid this issue is to compute the estimator $\hat{\pi}(\boldsymbol{\vartheta}^*|D)$ and the resulting marginal likelihood estimate for a large set of high density points and to derive an averaged estimate based on this set.

3.3.3 New algorithms for marginal likelihood estimation

The algorithm for estimation of the marginal likelihood of the data within a Bayesian model as presented by Chib and Jeliazkov (2001) is suitable approach utilizing a MCMC-

generated sample of the posterior distribution. However, especially in the setting of a high-dimensional parameter space which is a major focus of the present chapter, we pointed out two aspects in which the proposed algorithm can be further improved. Considering these issues we will propose modified versions of the estimation algorithm within this part of this section.

Firstly, the proposal distribution used within the algorithm does not have to coincide with that from a MCMC algorithm, which was potentially used for generating the posterior sample. The only necessary condition is that it must fulfill the detailed-balance condition (3.1), which holds for a large class of possible proposal kernels and corresponding acceptance functions. Thus, a proposal distribution that minimizes variance of the marginal likelihood estimator based on summary statistics of the underlying sample should be preferred over an MCMC-optimal proposal distribution which may lead to bias and large variance.

Secondly, the marginal likelihood estimator as proposed by Chib and Jeliazkov (2001) is evaluated at one point within the posterior domain, preferably some high density point. However, the choice of evaluation point affects the eventual estimate due to the probabilistic nature of the underlying posterior sample. Therefore, it might yield more accurate results to evaluate the estimator for a set of randomly chosen points and use the resulting average as the final estimate.

Accounting for these aspects we suggest a set of modified algorithms, which are particularly suitable for high-dimensional models in which the above mentioned points are especially important. Of course, the algorithms can also by applied to models of lower dimensions. Like the algorithm by Chib and Jeliazkov (2001) the proposed algorithms solely utilize a given unnormalized posterior density function $f(D\,|\cdot)\pi(\cdot)$, consisting of a likelihood function f and a prior density π, as well as a sample $\boldsymbol{\Theta} = \{\boldsymbol{\vartheta_j}\}_{j=1,\dots,J}$ from the corresponding posterior distribution.

Modified algorithm with rescaled random-walk proposal

We first suggest a novel algorithm, which is a slight adaption of the algorithm originally proposed by Chib and Jeliazkov (2001). This new algorithm – presented as Algorithm 3 on page 74 – also utilizes a state-dependent proposal density $q(\boldsymbol{\vartheta}, \cdot)$. However, here q is a multivariate normal density utilizing the unscaled empirical covariance of the underlying posterior sample and thus generally does not coincide with the proposal from a preceding MCMC algorithm, which commonly uses a scaled version of the posterior covariance, if it is available in the first place (Gelman et al., 1996).

As a novel feature the algorithm is not evaluated for a single point, but for a set of K points $\{\boldsymbol{\vartheta_k^*}\}$ instead. The eventual marginal likelihood estimator is then computed as the average of the K single estimates on the log scale to be more resilient with respect to outliers. Note that the main part of additional computational work is done in 1(b) and 1(d) where the model has to be evaluated for the new parameter vectors $\boldsymbol{\vartheta_k^*}$ and $\boldsymbol{\breve{\vartheta}}_n^{(k)}$, which yields $K \cdot (N + 1)$ additional likelihood evaluations, which is the costly part when dealing

---

Algorithm 3: Marginal likelihood estimation with rescaled random-walk proposal

---

Input: $\pi(\cdot), f(\boldsymbol{D}|\cdot)$: prior density and likelihood function

Input: $\boldsymbol{\Theta} = (\boldsymbol{\vartheta}_j)_{j=1,\dots,J}, (\pi(\boldsymbol{\vartheta}_j), f(\boldsymbol{D}|\boldsymbol{\vartheta}_j))_{j=1,\dots,J}$: sample from
$\pi(\boldsymbol{\vartheta}|\boldsymbol{D}) \propto \pi(\boldsymbol{\vartheta}) f(\boldsymbol{D}|\boldsymbol{\vartheta})$ together with corresponding prior probabilities
and likelihoods

Input: K: number of evaluation points

Input: N: number of proposal samples

Output: $\log \hat{f}(D)$: estimated log marginal likelihood

1. **for** $k = 1$ **to** K **do**

    (a) Draw $\boldsymbol{\vartheta}_k^*$ from $\mathcal{N}\left(\overline{\boldsymbol{\Theta}}, \widehat{\mathrm{Cov}}(\boldsymbol{\Theta})\right)$

    (b) Compute
    $$U_k = J^{-1} \sum_{j=1}^{J} A(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}_k^*) \, q(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}_k^*)$$
    with
    $$A(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) = \min\left\{1, \frac{\pi(\boldsymbol{\vartheta}_2) f(\boldsymbol{D}|\boldsymbol{\vartheta}_2)}{\pi(\boldsymbol{\vartheta}_1) f(\boldsymbol{D}|\boldsymbol{\vartheta}_1)}\right\},$$
    $$q(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*) = \phi_{\boldsymbol{\vartheta}, \widehat{\mathrm{Cov}}(\boldsymbol{\Theta})}(\boldsymbol{\vartheta}^*)$$

    (c) **for** $n = 1$ **to** N **do** Draw $\breve{\boldsymbol{\vartheta}}_n^{(k)}$ from $\mathcal{N}\left(\boldsymbol{\vartheta}_k^*, \widehat{\mathrm{Cov}}(\boldsymbol{\Theta})\right)$, i.e. from $q(\boldsymbol{\vartheta}_k^*, \cdot)$;

    (d) Compute
    $$L_k = N^{-1} \sum_{n=1}^{N} A\left(\boldsymbol{\vartheta}_k^*, \breve{\boldsymbol{\vartheta}}_n^{(k)}\right)$$

    (e) Compute
    $$\log \hat{f}_k(D) = \log f(D|\boldsymbol{\vartheta}_k^*) + \log \pi(\boldsymbol{\vartheta}_k^*) - (\log U_k - \log L_k)$$

    end

2. Compute
$$\log \hat{f}(D) = K^{-1} \sum_{k=1}^{K} \log \hat{f}_k(D)$$

with ODE models. Thus, for an efficient algorithm we want the factor $K \cdot (N+1)$ to be as low as possible subject to an overall accurate estimator $\hat{f}(D)$.

It should also be remarked, that the novel algorithm does not at all require the posterior sample to be computed by MCMC methods. As mentioned earlier, the only condition is that the proposal kernel q defined in 1(b) fulfills the detailed balance condition (3.1), which holds for any absolutely continuous posterior distribution $\pi(\boldsymbol{\vartheta} | \boldsymbol{D})$ due to the multivariate normality of q. Thus, the algorithm could also be applied to any available posterior samples if the corresponding unnormalized posterior density $f(D | \boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta})$ and its values for the sample $\boldsymbol{\Theta}$ are also available.

Modified algorithm with state-independent proposal

The second new algorithm which we propose makes further use of the flexibility in the choice of q. Here, the proposal density $q(\boldsymbol{\vartheta}, \cdot)$ is chosen such that it does not depend on the past parameter $\boldsymbol{\vartheta}$, i.e.

$$q(\boldsymbol{\vartheta}, \cdot) = q(\cdot)$$

which yields the following estimator for the pointwise posterior density

$$\hat{\pi}(\boldsymbol{\vartheta}^* | D) = \frac{J^{-1} \sum_{j=1}^{J} A(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}^*) q(\boldsymbol{\vartheta}^*)}{N^{-1} \sum_{n=1}^{N} A(\boldsymbol{\vartheta}^*, \breve{\boldsymbol{\vartheta}}_n)},$$

where as above $\{\boldsymbol{\vartheta}_j\}_{j=1,\dots,J}$ and $\left\{\breve{\boldsymbol{\vartheta}}_j\right\}_{n=1,\dots,N}$ denote samples from the posterior and $q(\cdot)$, respectively. However, due to the proposal q being asymmetric in this case the acceptance function A of the corresponding Metropolis-Hastings algorithm (Algorithm 1) is then given by

$$A(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) = \min\left\{1, \frac{\pi(\boldsymbol{\vartheta}_2) f(\boldsymbol{D} | \boldsymbol{\vartheta}_2) q(\boldsymbol{\vartheta}_1)}{\pi(\boldsymbol{\vartheta}_1) f(\boldsymbol{D} | \boldsymbol{\vartheta}_1) q(\boldsymbol{\vartheta}_2)}\right\}.$$

such that the estimator for the pointwise posterior density can be written as

$$\hat{\pi}(\boldsymbol{\vartheta}^* | D) = q(\boldsymbol{\vartheta}^*) \frac{J^{-1} \sum_{j=1}^{J} \min\left\{1, \frac{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D}) q(\boldsymbol{\vartheta}_j)}{\pi(\boldsymbol{\vartheta}_j | \boldsymbol{D}) q(\boldsymbol{\vartheta}^*)}\right\}}{N^{-1} \sum_{n=1}^{N} \min\left\{1, \frac{\pi(\breve{\boldsymbol{\vartheta}}_n | \boldsymbol{D}) q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D}) q(\breve{\boldsymbol{\vartheta}}_n)}\right\}}. \tag{3.17}$$

Thus, by applying an state-independent proposal density q one obtains Algorithm 4 (page 76) for estimation of the marginal model likelihood. In contrast to Algorithm 3, here we have to specify a proposal density q beforehand. However, we will see that this can also be done based on the information contained in the posterior sample. Furthermore, since the proposal q is independent from the evaluation point $\boldsymbol{\vartheta}_k^*$ it is not required to generate a new sample from $q(\cdot)$ for each of the K evaluation points, but instead one can use one sample for all K evaluation points (see point 2(d)). Thus, the overall estimator requires only $K+N$ additional posterior evaluations whereas Algorithm 3 based on a state-dependent proposal needs $K(N+1)$ additional posterior evaluations.

Algorithm 4: Marginal likelihood estimation with state-independent proposal

---

Input: $\pi(\boldsymbol{\vartheta}), f(\boldsymbol{D}\,|\boldsymbol{\vartheta})$: prior density and likelihood function

Input: $\boldsymbol{\Theta} = \{\boldsymbol{\vartheta}_j\}_{j=1,\dots,J}, \{\pi(\boldsymbol{\vartheta}_j), f(\boldsymbol{D}\,|\boldsymbol{\vartheta}_j)\}_{j=1,\dots,J}$: sample from
$\pi(\cdot\,|\boldsymbol{D}) \propto \pi(\boldsymbol{\vartheta})f(\boldsymbol{D}\,|\boldsymbol{\vartheta})$ together with corresponding prior probabilities and
likelihoods

Input: $q(\cdot)$: proposal density function (sampling from q must be possible)

Input: K: number of evaluation points

Input: N: number of proposal samples

Output: $\log \hat{f}(D)$: log marginal likelihood

1. **for** $n = 1$ **to** N **do**

   (a) Draw $\breve{\boldsymbol{\vartheta}}_n$ from $q(\bullet)$

   (b) Compute and store $\pi(\breve{\boldsymbol{\vartheta}}_n)$, $f\left(\boldsymbol{D}\,\middle|\breve{\boldsymbol{\vartheta}}_n\right)$ and $q(\breve{\boldsymbol{\vartheta}}_n)$

   end

2. **for** $k = 1$ **to** K **do**

   (a) Draw $\boldsymbol{\vartheta}_k^*$ from $q(\cdot)$

   (b) Compute and store $\pi(\boldsymbol{\vartheta}_k^*)$, $f(\boldsymbol{D}\,|\boldsymbol{\vartheta}_k^*)$ and $q(\boldsymbol{\vartheta}_k^*)$

   (c) Compute

   $$U_k = J^{-1} \sum_{j=1}^{J} A\left(\boldsymbol{\vartheta}_j, \boldsymbol{\vartheta}_k^*\right)$$

   with

   $$A\left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2\right) = \min\left\{1, \frac{\pi(\boldsymbol{\vartheta}_2)f(\boldsymbol{D}\,|\boldsymbol{\vartheta}_2)\cdot q(\boldsymbol{\vartheta}_1)}{\pi(\boldsymbol{\vartheta}_1)f(\boldsymbol{D}\,|\boldsymbol{\vartheta}_1)\cdot q(\boldsymbol{\vartheta}_2)}\right\}$$

   (d) Compute

   $$L_k = N^{-1} \sum_{n=1}^{N} A\left(\boldsymbol{\vartheta}_k^*, \breve{\boldsymbol{\vartheta}}_n\right)$$

   end

3. Compute

   $$\log \hat{f}_k(D) = \log f(D\,|\boldsymbol{\vartheta}_k^*) + \log \pi\left(\boldsymbol{\vartheta}_k^*\right) - \left(\log q(\boldsymbol{\vartheta}_k^*) + \log U_k - \log L_k\right)$$

4. Compute

   $$\log \hat{f}(D) = K^{-1} \sum_{k=1}^{K} \log \hat{f}_k(D)$$

Choice of proposal density

An open question refers to the choice of proposal function q used within Algorithm 4. Assuming, we knew the normalized posterior density $\pi\left(\cdot\,|\boldsymbol{D}\right)$ and we chose $q(\cdot)=\pi\left(\cdot\,|\boldsymbol{D}\right)$, then the acceptance function is constant $A\equiv 1$ and we obtain

$$\hat{\pi}(\boldsymbol{\vartheta}^*\,|D)=q\left(\boldsymbol{\vartheta}^*\right)=\pi\left(\boldsymbol{\vartheta}^*\,|\boldsymbol{D}\right).$$

Moreover, as the pointwise posterior density estimator is continuous with respect to the maximum log difference of the posterior and proposal distribution it follows that the estimators variance becomes small if we chose q such that it approximately mimics the normalized posterior (in the sense of the maximum log difference). This continuity argument for our proposed estimator given through Equation (3.17) and Algorithm 4 is provided by the following proposition.

Proposition 2. *Let $\pi\left(\cdot\,|\boldsymbol{D}\right)$ denote a normalized posterior density and $q(\cdot)$ a density function such that*

$$\|\log\pi\left(\cdot\,|\boldsymbol{D}\right)-\log q(\cdot)\|_\infty := \sup_{\boldsymbol{\vartheta}\in\mathbb{R}^{\mathrm{d}}}|\log\pi\left(\boldsymbol{\vartheta}\,|\boldsymbol{D}\right)-\log q(\boldsymbol{\vartheta})|<\infty.$$

Then for the pointwise posterior density estimator $\hat{\pi}(\boldsymbol{\vartheta}^*\,|D)$ given by (3.17) and for all points $\boldsymbol{\vartheta}^*$ it holds

$$|\log\hat{\pi}(\boldsymbol{\vartheta}^*\,|D)-\log\pi\left(\boldsymbol{\vartheta}^*\,|\boldsymbol{D}\right)|\leq 3\|\log\pi\left(\cdot\,|\boldsymbol{D}\right)-\log q(\cdot)\|_\infty.$$

Proof.

$$|\log\hat{\pi}(\boldsymbol{\vartheta}^*\,|D)-\log\pi\left(\boldsymbol{\vartheta}^*\,|\boldsymbol{D}\right)|=\left|\log\frac{\hat{\pi}(\boldsymbol{\vartheta}^*\,|D)}{\pi\left(\boldsymbol{\vartheta}^*\,|\boldsymbol{D}\right)}\right|$$

$$\leq\left|\log\frac{q(\boldsymbol{\vartheta}^*)}{\pi\left(\boldsymbol{\vartheta}^*\,|\boldsymbol{D}\right)}\right|+\left|\log\frac{J^{-1}\sum_{j=1}^J\min\left\{1,\frac{\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})q(\boldsymbol{\vartheta}_j)}{\pi(\boldsymbol{\vartheta}_j|\boldsymbol{D})q(\boldsymbol{\vartheta}^*)}\right\}}{N^{-1}\sum_{n=1}^N\min\left\{1,\frac{\pi(\breve{\boldsymbol{\vartheta}}_n|\boldsymbol{D})q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^*|\boldsymbol{D})q(\breve{\boldsymbol{\vartheta}}_n)}\right\}}\right|$$

Assuming

$$J^{-1}\sum_{j=1}^J\min\left\{1,\frac{\pi\left(\boldsymbol{\vartheta}^*\,|\boldsymbol{D}\right)q(\boldsymbol{\vartheta}_j)}{\pi\left(\boldsymbol{\vartheta}_j\,|\boldsymbol{D}\right)q(\boldsymbol{\vartheta}^*)}\right\}>N^{-1}\sum_{n=1}^N\min\left\{1,\frac{\pi\left(\breve{\boldsymbol{\vartheta}}_n\,|\boldsymbol{D}\right)q(\boldsymbol{\vartheta}^*)}{\pi\left(\boldsymbol{\vartheta}^*\,|\boldsymbol{D}\right)q(\breve{\boldsymbol{\vartheta}}_n)}\right\},$$

it holds

$$|\log \hat{\pi}(\boldsymbol{\vartheta}^* | D) - \log \pi(\boldsymbol{\vartheta}^* | \boldsymbol{D})|$$

$$\leq \left| \log \frac{q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D})} \right| + \log \frac{J^{-1} \sum_{j=1}^{J} \min\left\{ 1, \frac{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D}) q(\boldsymbol{\vartheta}_j)}{\pi(\boldsymbol{\vartheta}_j | \boldsymbol{D}) q(\boldsymbol{\vartheta}^*)} \right\}}{N^{-1} \sum_{n=1}^{N} \min\left\{ 1, \frac{\pi(\breve{\boldsymbol{\vartheta}}_n | \boldsymbol{D}) q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D}) q(\breve{\boldsymbol{\vartheta}}_n)} \right\}}$$

$$\leq \left| \log \frac{q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D})} \right| - \log \left\{ N^{-1} \sum_{n=1}^{N} \min\left\{ 1, \frac{\pi(\breve{\boldsymbol{\vartheta}}_n | \boldsymbol{D}) q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D}) q(\breve{\boldsymbol{\vartheta}}_n)} \right\} \right\}$$

$$\leq \left| \log \frac{q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D})} \right| - \log \left\{ N^{-1} \sum_{n=1}^{N} \min\left\{ 1, \exp\left(\log \frac{\pi(\breve{\boldsymbol{\vartheta}}_n | \boldsymbol{D})}{q(\breve{\boldsymbol{\vartheta}}_n)} + \log \frac{q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D})} \right) \right\} \right\}$$

$$\leq \left| \log \frac{q(\boldsymbol{\vartheta}^*)}{\pi(\boldsymbol{\vartheta}^* | \boldsymbol{D})} \right| - \log \left\{ N^{-1} \sum_{n=1}^{N} \min\left\{ 1, \exp\left(-2 \| \log \pi(\cdot | \boldsymbol{D}) - \log q(\cdot) \|_\infty \right) \right\} \right\}$$

$$\leq 3 \| \log \pi(\cdot | \boldsymbol{D}) - \log q(\cdot) \|_\infty$$

For the opposite case the result follows analogously. $\qquad\square$

From Proposition 2 immediately follows that the variance of the pointwise posterior density estimator is bounded

$$\text{Var}\left[\log \hat{\pi}(\boldsymbol{\vartheta}^* | D) \right] \leq \left(6 \| \log \pi(\cdot | \boldsymbol{D}) - \log q(\cdot) \|_\infty \right)^2 \quad \text{for all } \boldsymbol{\vartheta}^*$$

as well as the mean squared error

$$\mathbb{E}\left[\left(\log \hat{\pi}(\boldsymbol{\vartheta}^* | D) - \log \pi(\boldsymbol{\vartheta}^* | \boldsymbol{D}) \right)^2 \right] \leq \left(6 \| \log \pi(\cdot | \boldsymbol{D}) - \log q(\cdot) \|_\infty \right)^2 \quad \text{for all } \boldsymbol{\vartheta}^*.$$

Unfortunately, securing that $\| \log \pi(\cdot | \boldsymbol{D}) - \log q(\cdot) \|_\infty$ is indeed small for a specific choice of q is not feasible in a real setting as $\pi(\cdot | \boldsymbol{D})$ is only known through a sample and thus full information on, e.g. its tails or maximum values, is not available. Therefore, Proposition 2 should be seen as a theoretical motivation for defining q to be a good approximation of $\pi(\cdot | \boldsymbol{D})$.

Thus, the proposal density q has to be specified solely based on the available information on the posterior contained in the sample $\boldsymbol{\Theta} = \{ \boldsymbol{\vartheta}_j \}_{j=1,\dots,J}$. Thereby, our goal is to define q such that it approximately reproduces the unknown posterior density.

The first approach exploits the fact that the posterior is asymptotically normal for an increasing number of data points (see Section 2.1.1). Thus, by defining q as a suitable normal distribution

$$q_{\text{N}}(\boldsymbol{\vartheta}) = \phi_{\overline{\boldsymbol{\Theta}}, \widehat{\text{Cov}}(\boldsymbol{\Theta})}(\boldsymbol{\vartheta})$$

based on the empirical sample mean $\overline{\boldsymbol{\Theta}}$ and covariance $\widehat{\text{Cov}}(\boldsymbol{\Theta})$, q should provide a reasonable approximate of the unknown posterior density in many cases.

An alternative approach is to define a kernel density estimate for q based on the posterior sample $\mathbf{\Theta}$ (Scott, 1992), i.e.

$$q_{\mathrm{KD}}(\boldsymbol{\vartheta}) = J^{-1} \sum_{j=1}^{J} K_H \left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_j \right)$$

where K_H denotes the kernel function with bandwidth H, which is a measure for the covariance of a single kernel. As one is required to evaluate and sample from the resulting proposal density q_{KD} the kernel function K_H should be chosen accordingly. For this purpose, we define K_H to be a centred multivariate normal density with covariance matrix H. As a rule of thumb (Scott, 1992), a reasonable choice for H is given by

$$\hat{H} = J^{\frac{-2}{d+4}} \widehat{\mathrm{Cov}} \left(\mathbf{\Theta} \right).$$

Choosing a normal kernel function guarantees that the posterior $\pi \left(\cdot \, | \boldsymbol{D} \right)$ is absolutely continuous with respect to q_{KD} such that the detailed-balance condition (3.1) is fulfilled.

As any analytical properties always depend on the actual shape of the underlying posterior and the chosen proposal distribution q we will examine the performance of the presented algorithms in a simulation setting subject to different assumptions regarding the posterior function and compare the accuracy of the corresponding estimators to that of the original marginal likelihood estimator presented in Chib and Jeliazkov (2001).

3.3.4 Simulation study: Marginal likelihood estimation

Aim of this simulation-based study is to investigate the performance of the marginal likelihood estimation algorithms presented above. Thereby, we want to examine different algorithm calibrations and posterior distributions. The algorithm specifications are hereby the choice of proposal density q to be used, the number K of estimator evaluations, and the number N of newly proposed parameter vectors within each evaluation.

Simulation setup

The algorithms will be tested within three distinct settings of estimating the marginal likelihood of an underlying Bayesian model. These models each consist of a prior density $\pi(\boldsymbol{\vartheta})$ and likelihood function $f(\boldsymbol{\vartheta})$ which are defined on the parameter space $\mathbb{R}^d$. Note, that for the purpose of this simulation we assume the likelihood function to be directly available without simulating any data. The three models are chosen such that typical model characteristics of the resulting posterior function can be addressed within this study, these are either approximate normality, skewness, or mulitmodality of the posterior. Thus, we are interested in how the considered estimation algorithms perform under those circumstances.

In contrast to the simulation-based study from Section 2.3, here we will not consider a simple SIR-model to examine the algorithms, as the considered likelihood functions will

be evaluated a few million times for each tested model and algorithm, which makes utilizing as ODE-based likelihood infeasible for our purposes. However, the performance of the respective algorithms are lastly only dependent on the shape of the posterior function, regardless of where it comes from. When later discussing our results in Section 3.3.4, we will address the implications for ODE-based models.

Setup 1

Within the first setting we assume the prior and likelihood function to be centred multivariate Gaussian densities with diagonal covariance matrices σI_d and νI_d, respectively.

$$\pi(\boldsymbol{\vartheta}) = \phi_{0,\sigma I_d}(\boldsymbol{\vartheta}), \quad f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) = \phi_{\boldsymbol{\vartheta},\nu I_d}(0).$$

The covariance matrices being diagonal indicate the d components of the parameter vector $\boldsymbol{\vartheta}$ to be a priori and also a posteriori independent. This model and in particular this likelihood function could be interpreted as a Bayesian estimation approach regarding the mean of a d-dimensional random vector with known diagonal covariance matrix Σ_D, of which N_D independent realization are supposed to be available as data $\boldsymbol{D} = \{D_n\}_{n=1,\ldots,N_D}$ having an empirical mean of zero. The covariance matrix within the Gaussian likelihood would then result from the single data point observational error covariance Σ_D and the number of observations in the data due to

$$\nu I_d = \frac{\Sigma_D}{N_D}.$$

The posterior density, which again will be denoted by $\pi(\boldsymbol{\vartheta} \,|\, \boldsymbol{D})$, is the probability density function that fulfils

$$\pi(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}) \propto f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) \, \pi(\boldsymbol{\vartheta}).$$

For this simple model the log marginal likelihood LML which will be subject of estimation can be calculated analytically by

$$\text{LML} = \log \int_{\mathbb{R}^d} f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) \, \pi(\boldsymbol{\vartheta}) \, d\boldsymbol{\vartheta} = -\frac{d}{2} \log\left(2\pi \cdot (\sigma + \nu)\right)$$

The assumption of a Gaussian posterior is motivated by the asymptotic normality of the posterior distribution. Thus, this model yields a reasonable approximation for settings where many data are available. For that reason we also set the parameters $\sigma = 1$ and $\nu = 0.01$, such that the likelihood has a bigger impact on the posterior which corresponds to a data heavy scenario. Regarding the model space dimension d we will explore four different scenarios of $d \in \{1, 5, 10, 30\}$ to examine the effect of dimensionality. These four dimensions will also be explored for the following two models.

Setup 2

The second investigated setting corresponds to the Bayesian estimation of parameters of a

multivariate exponential distribution. Here, the prior density and likelihood function are defined by

$$\pi(\boldsymbol{\vartheta}) = \prod_{i=1}^{d} \sigma e^{-\sigma \vartheta_i}, \quad f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) = \prod_{i=1}^{d} \vartheta_i^{n} e^{-\vartheta_i n \bar{x}_i}.$$

This likelihood function coincides with having n observations $\boldsymbol{D} = \{X_k\}_{k=1,\ldots,n}$ with empirical mean $\bar{\boldsymbol{x}}$ from a d-dimensional exponential distribution with independent components and parameter $\boldsymbol{\vartheta}$. This model is worth to be investigated as the resulting posterior firstly has a restricted domain as all parameter components ϑ_i have to be positive and secondly the posterior is skewed around its mean. To increase these effects we chose the parameters $\sigma = 1$, $n = 3$ and $\bar{x}_i = 1$ for all $i = 1, \ldots, d$ such that the resulting posterior is far from asymptotic normality. Again in this model the log marginal likelihood can be calculated by

$$\text{LML} = \sum_{i=1}^{d} \log \frac{n! \sigma}{(\sigma + n\bar{x}_i)^{n+1}}.$$

Setup 3
Within the third and last setting we aim to investigate the effects of a multimodal posterior. Thereby, we define the prior density as in the first case

$$\pi(\boldsymbol{\vartheta}) = \phi_{0, \sigma I_d}(\boldsymbol{\vartheta}).$$

For the likelihood function we assume the shape of a mixed normal distribution, i.e.

$$f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}) = 0.4 \, \phi_{\boldsymbol{\vartheta}, \nu I_d}(\mu) + 0.6 \, \phi_{\boldsymbol{\vartheta}, \nu I_d}(-\mu),$$

where

$$\mu = (1, 0, \ldots, 0)$$

is the d-dimensional zero vector with the first component set to one. Here, the marginal posterior distribution of the first component has two modes, while the marginal posteriors of all other components are centred normal distributions as in our first setting. For this model the log marginal likelihood is given by

$$\text{LML} = -\frac{d}{2} \log(2\pi \cdot (\sigma + \nu)) - (2(\sigma + \nu))^{-1}.$$

Also in this case we set the parameters $\sigma = 1$ and $\nu = 0.01$ implying a high impact of the likelihood and therefore the multi-modality.

Posterior sampling
Altogether, the three models combined with the four considered space dimensions $d \in \{1, 5, 10, 30\}$ yield twelve distinct models to investigate. Note that within this simulation study we are interested in the performance of the marginal likelihood estimators in the above presented settings. As all the estimation algorithms depend only on the prior and

likelihood function as well as a sample from the posterior, no data $\boldsymbol{D}$ will be simulated for the purpose of this study. In fact, the virtual data is implicitly given by the likelihood functions as mentioned above. For each model we then generated one posterior sample using MCMC methods subject to the unnormalized posterior density $f(\boldsymbol{D}|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta})$. Using this sample we computed the probability distribution of the marginal likelihood estimators described in Section 3.3.3 by evaluating each estimator $M = 1000$ times. Again note, that all presented estimators are random variables. Based on the empirical distribution we can derive the performance of each estimator, e.g. by looking at the mean squared error.

Each sample $\boldsymbol{\Theta}$ corresponding to one specific posterior distribution

$$\pi\left(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}\right) \propto f\left(\boldsymbol{D}\,|\,\boldsymbol{\vartheta}\right)\pi(\boldsymbol{\vartheta})$$

was computed using a Metropolis-Hasting algorithm, i.e. Algorithm 1 as described in Section 3.2.3. The proposal density q was defined as a centred Gaussian distribution with covariance matrix

$$\Sigma_{\mathrm{MCMC}} = \frac{2.4^2}{d}\,\mathrm{Cov}_{\pi(\cdot|\boldsymbol{D})},$$

where the scaling factor $s_d = 2.4^2/d$ was chosen according to Gelman et al. (1996) and $\mathrm{Cov}_{\pi(\cdot|\boldsymbol{D})}$ refers to the covariance matrix of the defined posterior distribution, which in all presented settings can be calculated analytically. The initial vector of the chain $\boldsymbol{\Theta}$ was set to zero in the normal and mixed normal setting (Setups 1 and 3) whereas it was set to one (for each component) in the exponential setting (Setup 2). The total chain length was set to $1,000,000$, while the eventual sample of size $J = 100,000$ was generated by thinning the full chain.

Investigated algorithm configurations

For each of the three model settings and four space dimensions we estimated the log marginal likelihood LML using different configurations of the two algorithms described in Section 3.3.3. Additionally, we compare these estimates to that from the original estimator proposed by Chib and Jeliazkov (2001).

For algorithm 3 we examined three different settings for the number of evaluation points K and number of newly proposed parameter vectors N around each evaluation point. The tested configurations were

$$(K, N) \in \{(1, 10000), (10, 1000), (100, 100)\}.$$

Since the number of additional posterior evaluations is given by $K(N + 1)$, we chose the configurations such that the product KN remained constant.

The estimator proposed by Chib and Jeliazkov (2001) was configured by setting $K = 1$ and $N = 10000$. Their algorithm basically coincides with a version of Algorithm 3 using a different proposal function q within Step 1(b), namely that which was used within the

MCMC run. For this purpose, we set

$$q\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^*\right) = \phi_{\vartheta, \widehat{\Sigma}_{\mathrm{MCMC}}}\left(\boldsymbol{\vartheta}^*\right) \qquad \text{with} \quad \widehat{\Sigma}_{\mathrm{MCMC}} = \frac{2.4^2}{d}\widehat{\mathrm{Cov}}\left(\boldsymbol{\Theta}\right)$$

using the sample covariance, since the true posterior covariance is generally not known. Again, this represents the canonical approach of using the same proposal kernel for both MCMC sample computation and marginal likelihood estimation.

For Algorithm 4, for which a state-independent proposal must be specified in advance, we examined the two proposal densities q_{N} and q_{KD} suggested in Section 3.3.3, which were based on asymptotic normality or kernel density estimation, respectively. Considering the further configuration we examined

$$(K, N) \in \{(1, 9999), (1000, 9000), (2000, 8000)\},$$

where the total number of additional posterior evaluations of Algorithm 4 is $K + N$. As all estimation algorithms examined here require the same number of posterior evaluations, comparability of their results with respect to computational effort can be assured.

Altogether, this yields ten distinct estimation algorithms to be tested, i.e. three for Algorithm 3 (A), three for Algorithm 4 using either q_{KD} (B) or q_{N} (C) and one for the original algorithm from Chib and Jeliazkov (2001) (D). For each of the twelve models and each of the ten algorithms the log marginal likelihood estimator $\widehat{\mathrm{LML}}_\bullet^{(K,N)}$ was computed $M = 1000$ times ($\bullet \in \{\mathrm{A,B,C,D}\}$). Based on these samples $\{\widehat{\mathrm{LML}}_{\bullet,i}^{(K,N)}\}_{i=1,\dots,M}$ we calculated the mean and quantiles of the estimators empirical distributions. Additionally we calculated the mean square error $\mathrm{MSE}_\bullet^{(K,N)}$ as a measure of accuracy of each estimator

$$\mathrm{MSE}_\bullet^{(K,N)} = \frac{1}{M}\sum_{i=1}^{M}\left(\widehat{\mathrm{LML}}_{\bullet,i}^{(K,N)} - \mathrm{LML}\right)^2.$$

Results

The empirical distribution of the considered marginal likelihood estimators are displayed in Figures 3.4 (Setup 1, page 87), 3.5 (Setup 2, page 88), 3.6 (Setup 3, page 89) for the models assuming a normal, exponential and mixed posterior, respectively. Also provided are the corresponding mean squared errors for each algorithm and model setting. To better classify the mean errors, recall that the marginal likelihood estimators were calculated on the log scale.

Considering the results from the normal posterior setup (Setup 1, Figure 3.4) the first thing to notice is that regardless of space dimension all newly proposed estimation algorithms from Section 3.3.3 outperform the original algorithm from Chib and Jeliazkov (2001), which is displayed in Column (D) on the left. Especially in higher dimension ($d = 30$) the bias of the original algorithm becomes apparent as already discussed in Section 3.3.2. The results from the slightly modified algorithm (Column A) show that by

rescaling the proposal density, the estimator's performance can be improved. Even more so, if the estimator is evaluated at multiple points ($K > 1$) at the cost of accuracy of each evaluation, i.e. decreasing N. Considering Algorithm 4 utilizing a state-independent proposal density (Columns (B),(C)), the accuracy can be further improved. In particular the algorithm based on a multivariate normal proposal density (C) achieves high accuracy, which comes certainly due this density providing very good approximation to the underlying model posterior which itself is a multivariate normal. It should be remarked that the algorithm utilizing the kernel density estimate (B) of the posterior function performs poor for higher dimensions ($d \geq 10$), presumably since suitable smooth density estimates are difficult to construct in higher dimensions (Liu et al., 2007).

Regarding the second model using an exponential posterior (Setup 2, Figure 3.5) similar results can be observed. The comparative performance of the algorithms in lower dimensions ($d \leq 5$) remains roughly the same, although the strong dominance of Algorithm (C) is not present in this case whereas the kernel density estimate (B) works very well in the one-dimensional model. For increasing dimension, Algorithms (A) and (C) work best, while Algorithm (B) performs again poorly for the same reasons as before. The malfunction of Algorithm (B) for $(K, N) = (100, 100)$ comes due to the algorithm frequently proposing points outside the posterior density's domain ($[0, \infty]^d$), which leads to missing results.

Within the third model, based on a mixed normal posterior, all algorithms have problems to assess the true marginal likelihood (Setup 3, Figure 3.6). The positive exception is the kernel density based Algorithm (B) which works rather accurate, but solely in lower dimensions as for the other model setups, whereas Algorithms (A) and (C) display bias and high variance for $d \leq 5$. While the accuracy of all algorithms decreases with increasing dimension (see MSE) it should be noted that this effect is very weak for Algorithm (C), where the mean square errors only slightly increased, e.g. 0.055 for $d = 30$ versus 0.035 for $d = 1$ based on $(K, N) = (2000, 8000)$. This can also be observed for the other model posteriors. Within every model setup it is apparent that all algorithms perform better when the number of evaluation points is increased, although there appears to be some saturation effect for the state-independent proposal algorithms ((B) and (C)) where an increase of $K = 1000$ to $K = 2000$ yields no effect.

Discussion of the simulation results

The algorithms proposed in 3.3.3 showed improved performance in estimating the marginal model likelihood when compared to the original approach by Chib and Jeliazkov (2001). Especially in high-dimensional models the original Algorithm (D) works poorly due to heavy bias and large variance. Therefore, the optimal transition proposal density in the sense of MCMC-sampling may not be the optimal choice when it comes to marginal likelihood estimation based on the detailed-balance approach using Estimator (3.3).

Rescaling the proposal covariance matrix and repeated estimation for different evaluation points, i.e. applying Algorithm 3, considerably improved the accuracy of the estimation for all considered posterior functions (see Algorithm (A) for $K > 1$). However, applying

a state-independent proposal density, i.e. applying Algorithm 4 (page 76), could further improve the accuracy, depending on the exact choice of density q. Within low dimensions, using a kernel density estimate q_{KD} (Algorithm (B)) provides a flexible approach yielding robust results regardless of the shape of the underlying posterior, whereas for higher dimensions a kernel density estimate is not recommendable. In the case of an uni-modal posterior (Setups 1 and 3), using a normal density for point proposal q_{N} (Algorithm (C)) yields very good results, even for higher dimensions. Also for these algorithm, repeated evaluation of the estimator at the cost of individual accuracy, provides better results. Note, that all considered algorithms required the same number of model posterior evaluations and therefore also approximately the same computation time.

Therefore, as a default Algorithm (C) with $K >> 1$ provides a safe choice for robust marginal likelihood estimation. The only exception is the case of a low-dimensional multi-modal posterior density (Setup 2), in which algorithm (B) yields better results due to its flexibility in mimicking the posterior density. As easy way to check on multi-modality in a low-dimensional model setting is to look at histograms of the single parameter distributions using the available posterior sample $\boldsymbol{\Theta}$. Furthermore, it should be remarked that a single evaluation of the kernel density estimate $q_{\mathrm{KD}}(\boldsymbol{\vartheta})$ for a specific point $\boldsymbol{\vartheta}$ is computationally much more expensive than the evaluation of q_{N}, whereas simulating from q_{KD} only requires one additional sampling step to choose the multivariate kernel to eventually sample from. Thus, choosing a kernel density estimate over a Gaussian proposal function is only reasonable in cases where evaluation of the likelihood constitutes the major computational work (as in e.g. ODE-based models), such that the additional work for computing q_{KD} is only marginal in comparison.

In the application of modelling infectious disease transmission, utilizing algorithm (C) based on a state-independent Gaussian proposal distribution should be strongly recommended in most applications for two reasons. Firstly, in many cases the assumption of an approximately normal posterior is justified due to the usually many incidence data available in this context, e.g. as in our main application in Chapter 5. On that matter, in case of doubt approximate normality can be checked by plotting histograms of the posterior sample in order to confirm the assumption. Secondly, applying high-dimensional models incorporating and estimating many parameters is a common approach in infectious disease epidemiology. Both these aspects indicate that our newly proposed estimation Algorithm (C) should be preferred in this context whereas reasons to apply alternative methods (multi-modality, heavy skewness) can be easily identified.

3.4   Discussion

In this section we adapted methodology to generate samples from posterior distributions and estimate the corresponding marginal likelihood of high-dimensional complex models, which are very common in the context of dynamic transmission of infectious diseases. The proposed algorithms address the challenges which arise when considering a posterior

function that requires large computational effort to evaluate and which approximate shape is hardly known in advance.

For posterior sampling we presented an adaptive Metropolis-Hastings approach by Haario et al. (2001) which incorporates newly gained information on the posterior distribution into the algorithm on the fly. This secures an efficient movement of the sample chain within the parameter space. Hereby, the adaptive approach works best if the posterior is approximately normal or at least uni-modal, which holds true if the amount of data is sufficiently large (Gelman et al., 2013). However, given a multi-modal or otherwise contorted posterior the adaptive algorithm might not be the most efficient choice.

Regarding marginal likelihood calculation we proposed new estimation algorithms inspired by Chib and Jeliazkov (2001) that make use of a previously generated posterior sample to reduce the number of additionally required model evaluations. Using analytical arguments and by simulation, we could show that the our newly suggested algorithms provide improved performance compared to the original version, especially in high-dimensional settings. We also provided guidance on which algorithms to use depending on the shape of the underlying posterior density, which can be assessed by checking corresponding histograms of the available sample.

The here presented methods are suitable approaches when dealing with time series models where the expectation structure is given through the solution of an ODE system. For alternative approaches in modelling infectious disease incidence data other methods to compute or sample the posterior may by certainly preferred. On that regard, if the data is modelled through a latent Gaussian model, a certain class of structured additive regression models, a modern approach is compute the posterior density by using integrated nested Laplace approximations (Rue et al., 2009). This approach circumvents costly MCMC method by directly computing an analytical approximation utilizing the hierarchical model structure and the Gaussian form of the prior, which yields much faster results.

In the case modelling the underlying expectation structure as a stochastic process instead of a deterministic ODE system the ABC-methods suggested by McKinley (2009) provide an efficient tool to compute approximate posterior distributions. Hereby, the approximate likelihood is not computed directly, which would involve many hidden parameters in that case, but it is approximated by simulating data from the model and calculate their deviation to the observed data. Thus, the model dimension can be reduced to only consider the parameters defining the stochastic process and further observational distributions as it is the case for our class of ODE models.
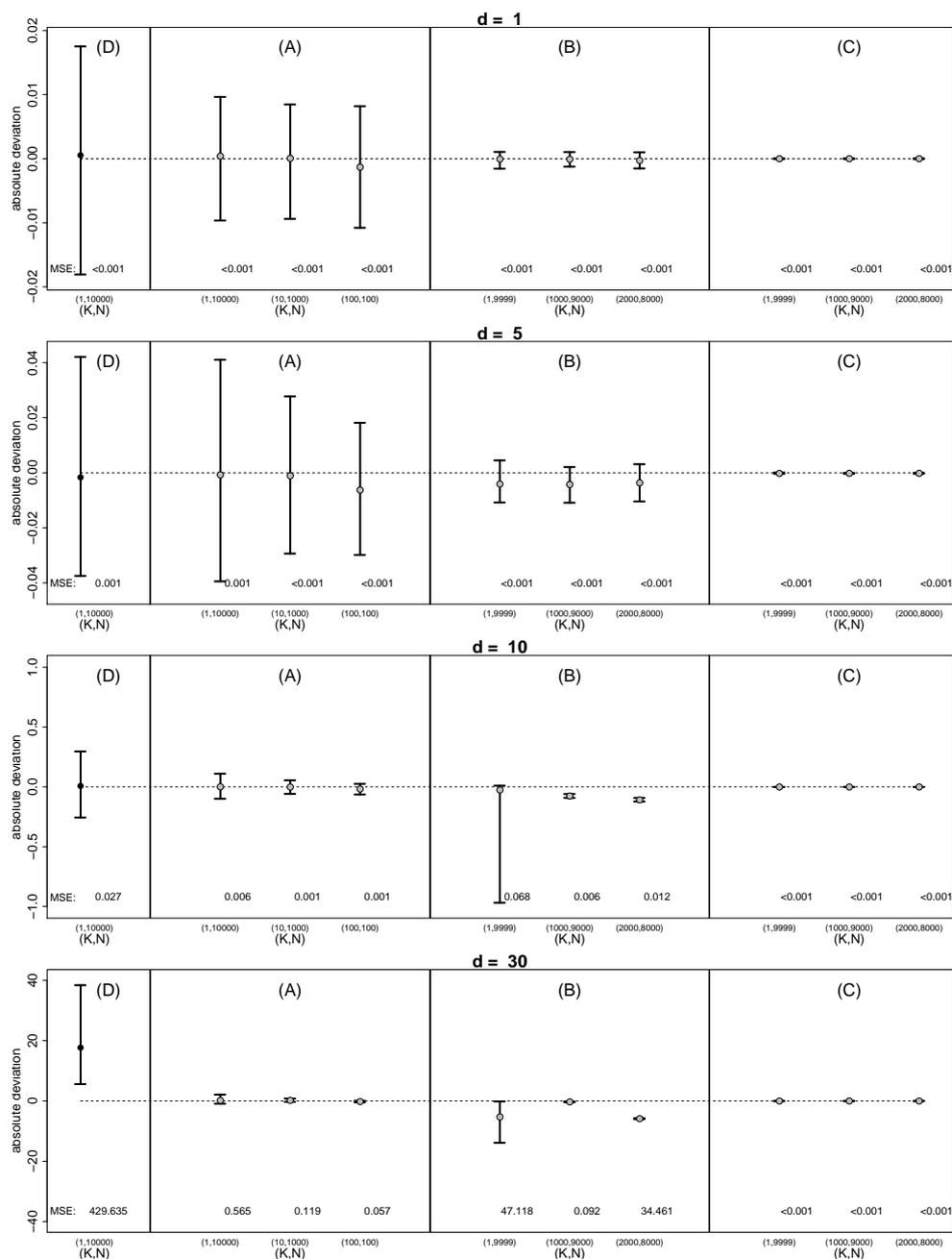
Figure 3.4: Sample mean and empirical 0.025 and 0.975 quantiles of the examined estimators $\widehat{\mathrm{LML}}_{\bullet}^{(K,N)}$ assuming a normal posterior (first model setup) within different underlying space dimensions $d \in \{1, 5, 10, 30\}$. The dotted line marks the true marginal likelihood subject of estimation. The investigated algorithms {D,A,B,C}) are given at the top, while the respective estimator configurations (K, N) are displayed on the x-axis. Also provided are the corresponding mean square error MSE of each configuration. Note the different scaling of the y-axes.
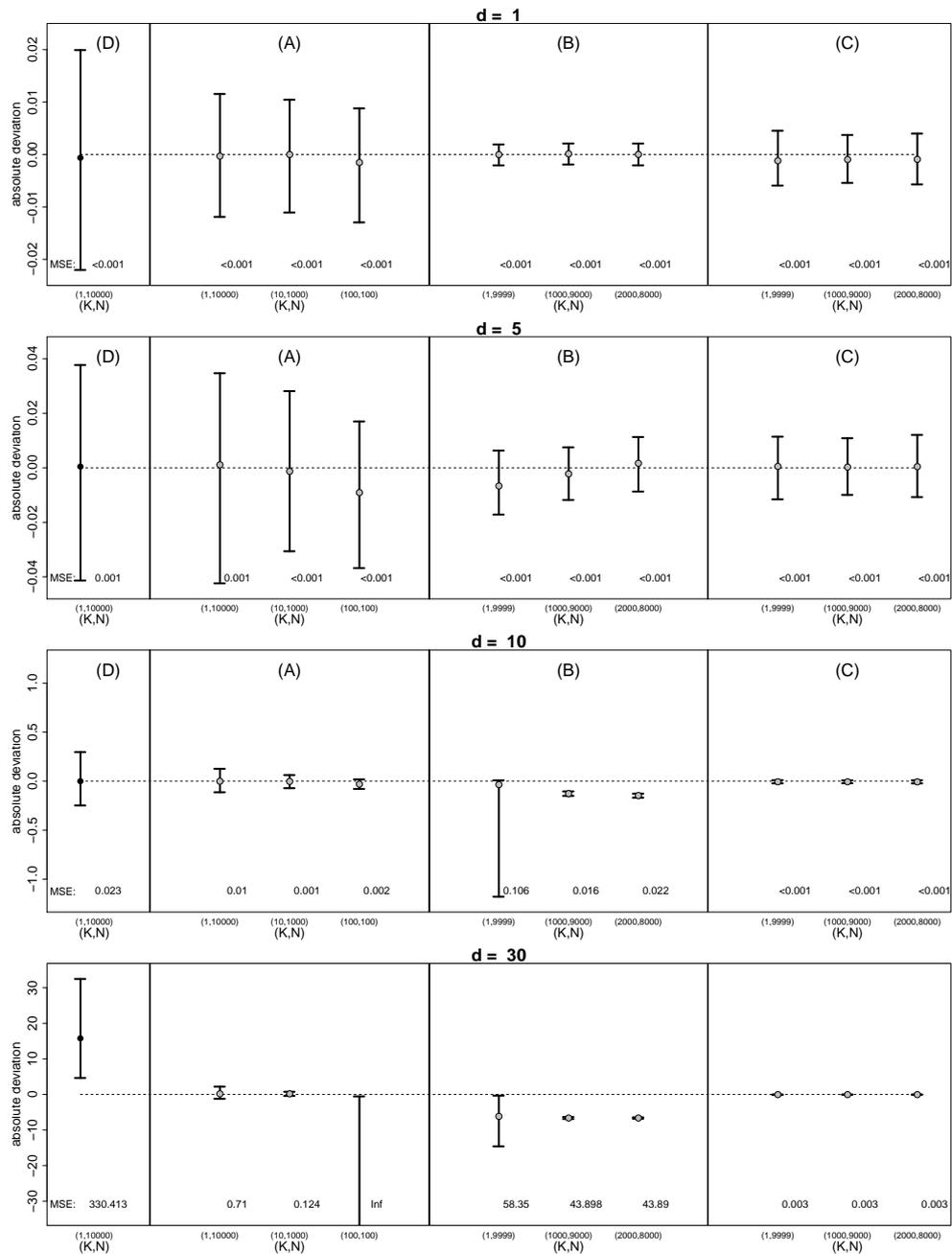
Figure 3.5: Sample mean and empirical 0.025 and 0.975 quantiles of the examined estimators $\widehat{\mathrm{LML}}_{\bullet}^{(K,N)}$ assuming an exponential posterior (second model setup) within different underlying space dimensions $d \in \{1, 5, 10, 30\}$. The dotted line marks the true marginal likelihood subject of estimation. The investigated algorithms {D,A,B,C}) are given at the top, while the respective estimator configurations (K, N) are displayed on the x-axis. Also provided are the corresponding mean square error MSE of each configuration. Note the different scaling of the y-axes.
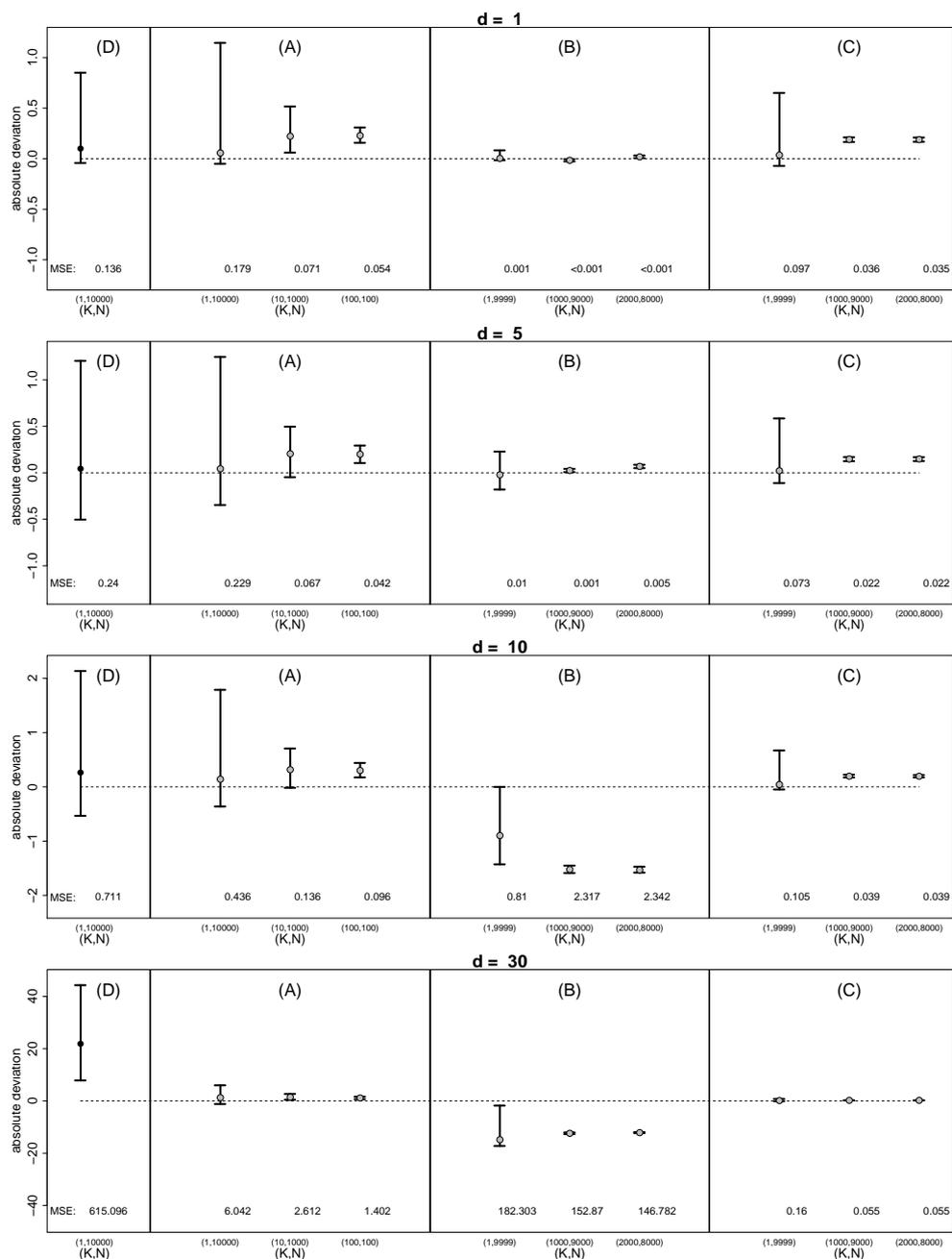
Figure 3.6: Sample mean and empirical 0.025 and 0.975 quantiles of the examined estimators $\widehat{\mathrm{LML}}_{\bullet}^{(K,N)}$ assuming a mixed normal posterior (third model setup) within different underlying space dimensions $d \in \{1, 5, 10, 30\}$. The dotted line marks the true marginal likelihood subject of estimation. The investigated algorithms $\{D,A,B,C\}$) are given at the top, while the respective estimator configurations (K, N) are displayed on the x-axis. Also provided are the corresponding mean square error MSE of each configuration. Note the different scaling of the y-axes.

Chapter 4

Bayesian parameter inference for dynamic infectious disease modelling: Rotavirus in Germany

> The content of this chapter is largely based on the main article published in Weidemann et al. (2014a).

In applied infectious diseases epidemiology there has been a substantial increase in the utilization of mathematical modelling, especially to support decision making in public health policy (O'Neill, 2010). While the focus within model building most often lies on the correct representation of epidemiological transmission dynamics, an accurate assessment of parameter inference for such dynamic models is in many cases neglected.

Currently, understanding the transmission dynamics of rotavirus in Germany is of particular importance in order to investigate the possible impact of implementing a routine childhood immunization against rotavirus, which was recently recommended by the Standing Committee on Vaccination in Germany (Ständige Impfkommission, 2013). Therefore, within this chapter we aim to not only develop a mathematical model based on ordinary differential equations for the dynamic transmission and stochastic observation of rotavirus infections in Germany, but also to adequately address uncertainty in such an approach. The most prominent sources of uncertainty thereby arise by from the choice of transmission model, its respective unknown parameters, and the resulting predictive distribution. We therefore applied a Bayesian framework for parameter inference and model evaluation with a special focus on residual autocorrelation in the data as presented in Chapter 2.2. Note that here we aim to model the rotavirus transmission dynamics not accounting for any vaccination mechanisms in order to establish a first assessment of the necessary transmission and epidemiological aspects. Considerations regarding the impact of a routine vaccination program will be covered within Chapter 5.

To study the underlying transmission dynamics of rotavirus and to analyse the resulting

potential impact of a vaccination program, SIR-type modelling has already been applied to the populations of the United States (Pitzer et al., 2009), England and Wales (Atkins et al., 2012; Atchison et al., 2010), Western Europe (Van Effelterre et al., 2010), Australia (Shim and Banks, 2006) and Kyrgyzstan (de Blasio et al., 2010). A modelling approach based on partial differential equations was used to investigate the vaccination impact on incidence and health care cost in the United States and Mexico (Shim and Castillo-Chavez, 2009; Shim and Galvani, 2009), which allowed to model the age distribution of underlying population as a continuous variable. Mathematical modelling has also been used to investigate the role of single demographic or epidemiological aspects, such as birth rates (Pitzer et al., 2011) or multiple strain dynamics (Pitzer et al., 2011).

In order to conduct inference for such models, various approaches have been applied in the past. Brookhart et al. (2002) used profile likelihood estimation to address the problem of possible overparametrization. Their approach decomposes the model parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$ into a vector of profiled parameters $\boldsymbol{\vartheta}_1$ and a vector $\boldsymbol{\vartheta}_2$ to be estimated. Profiled maximum likelihood estimators according to the likelihood function $f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta})$ were then calculated by

$$\boldsymbol{\vartheta}^* = (\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2^*(\boldsymbol{\vartheta}_1)), \quad \text{with} \quad \boldsymbol{\vartheta}_2^*(\boldsymbol{\vartheta}_1) = \arg\max_{\boldsymbol{\vartheta}_2} f(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$$

for different fixed choices of the profiled parameter vector $\boldsymbol{\vartheta}_1$, which ideally consists of parameters which are highly collinear to other parameters in $\boldsymbol{\vartheta}_2$. The profiled estimation thus circumvents problems caused by inconclusive MLEs. Whitaker and Farrington (2004) used a likelihood based approach to estimate force of infection parameters and critical vaccination coverage thresholds for mumps and rubella from serological survey data.

Bayesian approaches utilizing MCMC procedures to sample from the posterior distribution were applied by Elderd et al. (2006); Birrell et al. (2011); Dorigatti et al. (2012) to estimate parameters of ODE-based transmission models. In fact, these were all applications of the Bayesian melding approach as pointed out by Coelho et al. (2011). Bayesian melding (Poole and Raftery (2000)) is a Bayesian estimation framework for dynamic models in which the output $\boldsymbol{X}$ is a deterministic function of some model parameters $\boldsymbol{X} = M(\boldsymbol{\vartheta})$, such that data $\boldsymbol{D} = (\boldsymbol{D}_{\boldsymbol{\vartheta}}, \boldsymbol{D}_{\boldsymbol{X}})$ on both the output and the input parameters yields information on the joint distribution of $(\boldsymbol{\vartheta}, \boldsymbol{X})$ subject to the restriction imposed by the deterministic model function M.

Parameters of stochastic transmission models have been estimated via plug-and-play approaches based on either frequentist (He et al. (2010)) or Bayesian methods (Toni et al. (2009)). Such procedures search for parameter configurations which are able to simulate model output that is similar to the observed data, which in Bayesian statistics is also known as Approximate Bayesian Computing (ABC) (McKinley, 2009). For more details on the non-Bayesian analogue see, e.g., Ionides et al. (2011).

However, in applied infectious disease epidemiology, e.g. when modelling rotavirus transmission, many models demand a high complexity level. In those cases it has become a habit to base many parameter estimates on values from previous epidemiological studies

or on expert opinion alone, e.g. (Atchison et al., 2010; Pitzer et al., 2009; de Blasio et al., 2010). This means to fix the majority of parameters in advance and to perform data-driven model calibration for only a few model parameters – disregarding the full epidemiological information contained in the available incidence data. Moreover, in the setting of SIR-type and other ODE-based models, this procedure may yield misleading results due to the high parameter sensitivity of such models combined with possible parameter collinearity, which may lead to bias in the estimation of other parameters of interest.

Within this chapter we want to present a more comprehensive view on uncertainty compared to previous ODE-based approaches in epidemic modelling by using a Bayesian approach. We formulate a complex age-structured model capturing the underlying disease transmission dynamics via a system of ordinary differential equations, where we considered in detail all relevant aspects regarding the transmission of the virus, e.g. infectiousness, seasonality, and waning immunity. Additionally, a stochastic observational component describes the reporting process based on the previously computed transmission dynamics and occurred cases. In doing so we also provide a more mechanistic framework for describing routinely collected surveillance data compared to more time series orientated works (Held et al., 2005; Paul et al., 2008). Within a Bayesian framework we performed model inference combining available incidence data with knowledge from focused epidemiological studies. To additionally address uncertainty concerning model selection we applied Bayesian model averaging techniques, which allowed us to compute posterior estimates and incidence predictions unconditioned on a particular model, but on a set of models instead. Moreover, by adjusting the likelihood impact based on the cumulative autocorrelation of the observation residuals as outlined in Chapter 2, we are able to obtain a more realistic variance estimation accounting for potential mismatches between our model and the real underlying processes.

The chapter is organized as follows. In Section 4.1 the German rotavirus incidence data used to motivate our work is presented. Furthermore, the dynamic transmission model together with the stochastic observational component will be explained. The Bayesian two-step estimation and model averaging procedure, yielding results for the posterior distribution of the model parameters will be presented in Section 4.2. In Section 4.3 we apply our methods to the rotavirus incidence data and provide an interpretation of the epidemiological and statistical insights. The gained knowledge on rotavirus transmission, possible extensions of our model, and further analyses are discussed in Section 4.4.

4.1 Data, dynamic transmission and stochastic observation

This section explains the rotavirus incidence data and our proposed statistical model consisting of dynamic transmission of the virus and the stochastic case observation. It is important to distinguish the function of the two model components: the deterministic

transmission model describes the disease's dynamic prevalence within the population at any time point and determines the mechanisms of infection spreading, while the stochastic observational model explains how underlying transmission dynamics lead to occurrence and detection of new cases.

As a side note, our her employed model was build to describe the rotavirus dynamics in Germany. That means, that our developed transmission model can not necessarily be applied to any other infectious disease data or underlying population. However, going into the details of our model one can easily determine which aspects of model building may be adapted to other pathogens sharing the same transmission characteristics as rotavirus.

4.1.1 Epidemiological data basis

Since 2001, acute rotavirus infection is a notifiable disease in Germany (Koch and Wiese-Posselt, 2011). Laboratory-confirmed cases are routinely reported to the local health offices and forwarded electronically via the state health authority to the Robert Koch Institute (RKI). Each data set includes information on age, sex, federal state of residency, onset of symptoms, hospitalization and fatal outcome.

We examined data on the weekly number of rotavirus infections from 2001 onwards. After the licensure of rotavirus vaccines in 2006, the estimated rotavirus vaccination coverage reached 20% in 2008, averaged over all federal states (Dudareva et al., 2012). Therefore, we expected a significant impact on the rotavirus incidence in 2009 due to the increased coverage such that data from 2009 and the following years was excluded from the here presented analysis.

The data was separated by the two regions of eastern (EFS) and western (WFS) federal states of Germany (the state of Berlin is counted as a western federal state). It is a common phenomenon among notifiable gastroenteritis in Germany, that the reported incidence in the EFS is significantly higher than in the WFS (Rosner et al., 2010). These differences are not necessarily a result of a higher disease burden in the EFS, but are assumed to arise from different healthcare seeking behaviour of parents with sick children and different diagnostic test ordering behaviour of doctors.

From 2001 till 2008, a total of 441,508 rotavirus cases were reported to the RKI (see Figure 4.1). This translates into an average annual incidence of 1,097 and 20 cases per 100,000 population in the age groups <5 and ≥5 years of age, respectively, as well as 47 and 175 cases per 100,000 in the WFS and EFS, respectively. The data suggests a strong seasonal pattern of rotavirus incidence with its maximum during March with an average peak of more than 3,000 weekly cases and its minimum in August with just above 200 cases per week. One can observe a distinct increase in the peak number of reported cases with beginning of 2005. We expect the 2004 change in reimbursement for a hospitalized rotavirus case to be a major reason for this increase (where a gastroenteritis caused by a specific pathogen was reimbursed higher than without a specific pathogen being identified) (Koch and Wiese-Posselt, 2011).

Additional demographic data on monthly birth rates, annual age-stratified mortality
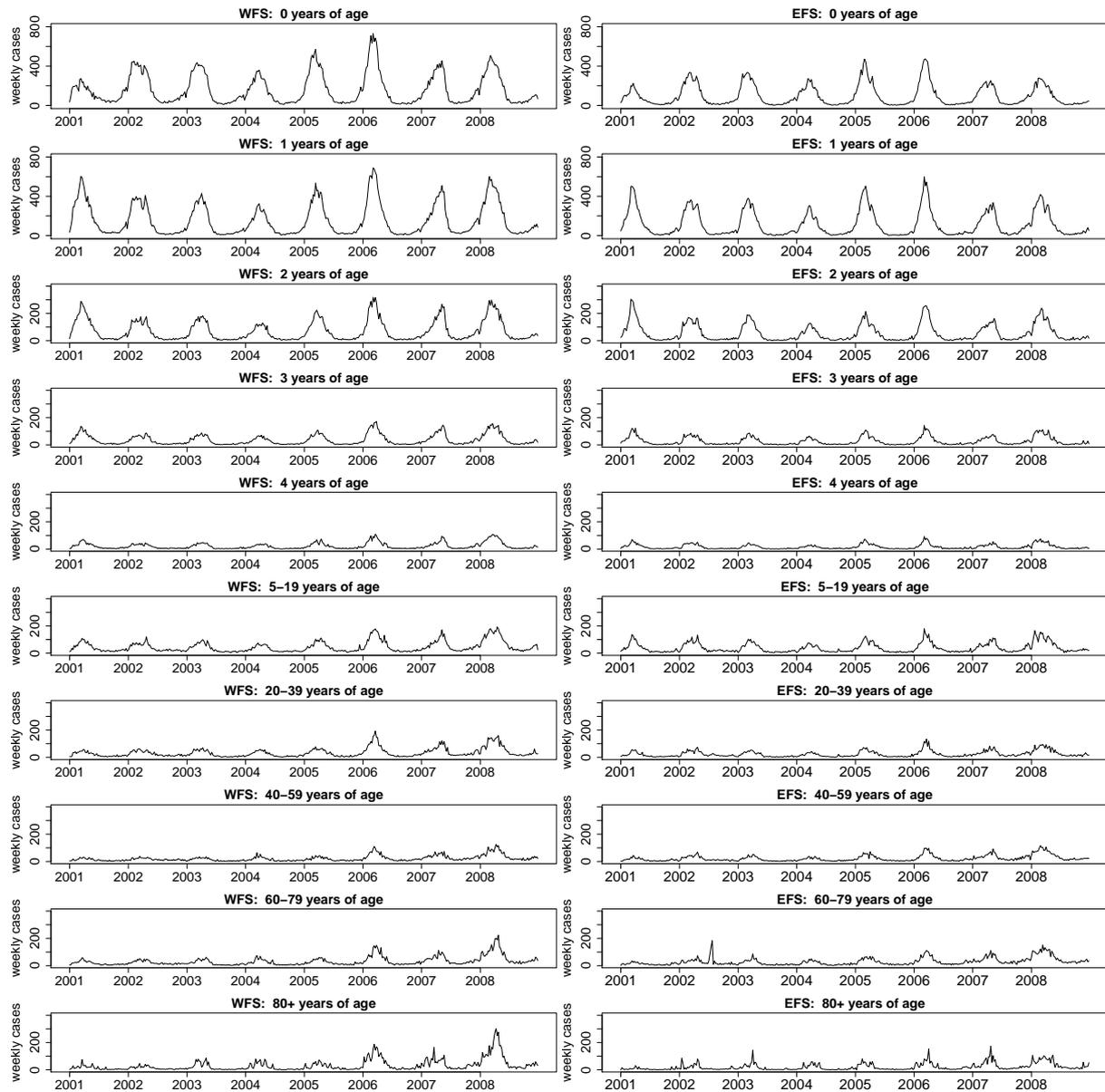
Figure 4.1: Case notification data on the weekly number of reported rotavirus cases stratified by the regions of western (left) and eastern (right) federal states and ten age groups: 0, 1, 2, 3, 4, 5-19, 20-30, 40-59, 60 - 79, 80+ years of age. Note the different scaling of the y-axes.

rates, age-stratified migration data from 1990-2008, and age-stratified population counts in the EFS and WFS for the years of 2001-2008 was obtained from the GENESIS database at the Federal Statistical Office (Federal Bureau of Statistics, 2013). Since the rotavirus-associated case fatality in Germany is below 0.01% we did not account for disease induced

mortality.

4.1.2   The dynamic transmission model

We constructed a detailed susceptible-infectious-susceptible (SIS)-type model based on ODEs to mimic the underlying rotavirus transmission dynamics in Germany. Thereby, we were mainly inspired by the models presented in (Pitzer et al., 2009; Atchison et al., 2010). Since our long term goal is to utilize the model to investigate the potential impact of a routine vaccination against rotavirus, we aimed at covering all relevant epidemiological and demographic aspects affecting the transmission within the underlying population. Thus, the resulting model might look overly complicated considering the available incidence data. However, this complexity is necessary in order to implement vaccination mechanisms into the model at a later stage.

In contrast to other infectious diseases, e.g. measles or varicella, one rotavirus infection does not provide lifelong immunity against the disease. However, after recovery some immunity against subsequent infection is acquired. Thus, to take the heterogeneity in the risk for succeeding rotavirus infections in the individual into account, we differentiated between several stages of susceptibility. Therefore, the uninfected population was split into three states of susceptibility S_k $(k = 1, 2, 3)$, corresponding to the number of infections already suffered (none, one, at least two). The states S_2 and S_3 yield a specific relative risk of infection α_k $(k = 2, 3)$ compared to the first susceptibility state S_1. Moreover, each state S_i corresponds to a specific risk of developing symptoms $\theta_k \in [0, 1]$ $(k = 1, 2, 3)$. This decomposition into multiple susceptibility levels constitutes the main aspect to be considered when building a model to capture the rotavirus epidemiology. A structural overview of our final model is given in Figure 4.2.

The model decomposes the whole population into several compartments or states, where each state variable counts the number of people contained in the corresponding compartment. Starting point of the employed model is the state M containing all newly born infants protected by maternal antibodies, which protect against acquiring infection in the first months of life. With loss of this maternal protection these infants move to the first susceptibility state S_1. Following each susceptibility state S_k is a pair of states corresponding to either symptomatic or asymptomatic infection I_k, A_k $(k = 1, 2, 3)$. In case of infection the probability θ_k of developing symptoms and moving to I_k depends on the corresponding susceptibility level k. All infected individuals, regardless of whether they developed symptoms, go through a period of asymptomatic infection (A_k) before recovery. Thereby, recovery from symptoms happens at a rate μ, whereas recovery from asymptomatic infection happens at a rate ω. After infection individuals move to the next susceptibility state S_{k+1}. The only exception is after recovery from the third infection (I_3, A_3), where individuals move back to the third susceptibility state S_3. As stated above, the state variables $M(t), S_k(t), I_k(t)$ and $A_k(t)$ represent the absolute numbers of individuals belonging to these states at time t.
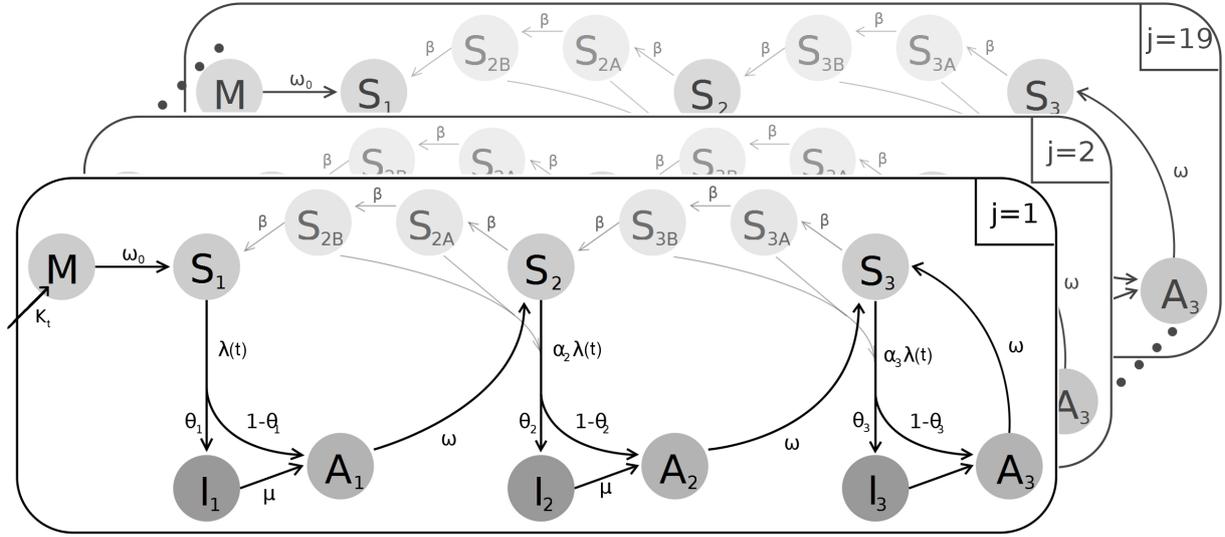
Figure 4.2: Layer structure of the age-stratified compartmental transmission model, in which each age group is represented by one layer.

There exists evidence for immunity waning in the absence of boosters through subsequent rotavirus infections (Fischer et al., 2002; Moulton et al., 1998). To incorporate this waning protection, we assumed that individuals move from state S_k to S_{k-1} ($k = 2, 3$) at a constant rate β. For a more realistic representation we introduced two intermediate states of susceptibility S_{kA} and S_{kB} ($k = 2, 3$), in order to model a temporal delay. Thus, the time to loss of immunity after infection corresponds to a $\Gamma(3, \beta)$-distribution instead of an $\exp(\beta)$-distribution. Similar mechanisms allowing a possible loss of acquired immunity were also implemented in the models by Atchison et al. (2010) and de Blasio et al. (2010).

Rotavirus incidence is the highest among children under 5 years of age with its peak at one year of age as the data suggest. To account for this heterogeneity, the data was stratified according to age by distinguishing between $n_D = 10$ age groups: one group for each of the first five years of age, one group from 5 to 19 years of age, and four subsequent age groups with a range of 20 years each (see Figure 4.1 for the respective data time series). In order to later be able to include fixed vaccination ages into the model at two, four, or six months of age and for a more realistic ageing process within the first years, the model population was decomposed even further. Thus, the first year of age was split into six equally broad age groups, while each group for the second till fifth year of age was divided into two age groups, which yields a total of $n_A = 19$ model age groups and 266 age-specific model states. To denote a specific age group of a model state, the state's name is marked with an upper index (j), ($j = 1, \ldots, n_A$). Ageing from one group to the next happens at a rate δ_j determined by the length of the age groups.

Operating with case count data requires a realistic modelling of the underlying population. In order to model the dynamic evolution of the population size as accurately as

possible, we fixed time- and age-specific death rates $\gamma_t^{(j)}$ as well as migration rates $m_t^{(j)}$ for each age group ($j = 1, \ldots, n_A$) and also a time-specific birth rate κ_t, each from 1990 till 2008. These rates were based on the corresponding data from the German Federal Statistical Office. Note that the resulting population size is not constant over time, since the applied death and birth rates may not coincide.

4.1.3 Disease transmission

In transmission modelling the rate at which susceptibles become infected is typically determined by the force of infection, which provides the ongoing, although changing, risk of acquiring infection within a given population. In our case the force of infection is a composition of several components affecting the transmission of rotavirus, i.e. infectiousness, susceptibility, contact pattern, and seasonality. Regarding infectiousness we assumed that both symptomatically and asymptomatically infected individuals contribute to the force of infection, but with symptomatically infected individuals having a different relative intensity $p > 0$. Thus, symptomatically infected people do not automatically have a higher infectiousess than asymptomatic people, since a lower infectiousness might be explained by a reduced contact frequency. The age-specific force of infection $\lambda_{\text{inf}}^{(j)}$ induced by the j-th age group then takes the form

$$\lambda_{\text{inf}}^{(j)}(t) = \sum_{k=1}^{3} \left(p I_k^{(j)}(t) + A_k^{(j)}(t) \right).$$

The contact pattern within transmission is modelled through a WAIFW-matrix (who acquires infection from whom) $C \in \mathbb{R}_{\geq 0}^{n_A \times n_A}$. The component $c_{i,j}$ of the matrix refers to the average number of weekly contacts of an individual from age group i with the j-th age group, which would lead to transmission of the virus. Thus, the parameter $c_{i,j}$ also contains information on the relative susceptibility and infectiousness of the respective age groups.

The seasonality component mimics the varying transmission rates over the course of a year due to climatic or other environmental factors, which is assumed to be equal for all age groups. Therefore, we assumed a positive periodic form by using

$$\lambda_{\text{seas}}(t) = \exp \left\{ a_1 \cos 2\pi \left(\frac{t}{52} - b_1 \right) + a_2 \cos 2\pi \left(\frac{t}{26} - b_2 \right) \right\}.$$

In contrast to previous modelling approaches (Pitzer et al., 2009; Atchison et al., 2010) we chose such an exponential form to allow the peak in transmission to be more seasonally localised rather than harmonic. For the same reason, we utilized an additional half annual wave for more flexibility in the seasonal variation. Note that, although λ_{seas} is called the seasonal term, also $\boldsymbol{\lambda}_{\text{inf}} = (\lambda_{\text{inf}}^{(1)}, \ldots, \lambda_{\text{inf}}^{(n_A)})$ varies over time, since it depends on the time-varying variables $I_k^{(j)}(t)$ and $A_k^{(j)}(t)$. As this is only an indirect time-dependence, it seems more suitable to refer to $\boldsymbol{\lambda}_{\text{inf}}(t)$ as a state-dependent and to $\lambda_{\text{seas}}(t)$ as a purely time-dependent function.

Combining these parts we obtain the force of infection, i.e.

$$\boldsymbol{\lambda}(t) = \frac{1}{N(t)} \lambda_{seas}(t) \boldsymbol{C} \boldsymbol{\lambda}_{\text{inf}}(t),$$

with $N(t)$ denoting the total population size at time t, which is just the sum of all considered states. Note that $\boldsymbol{\lambda}(t)$ is a time-dependent n_A-dimensional vector providing the force of infection for each of the $n_A = 19$ age groups. The force of infection affecting only to the j-th age class is thus given by

$$\lambda^{(j)}(t) = \frac{1}{N(t)} \lambda_{seas}(t) \sum_{i=1}^{n_A} c_{i,j} \lambda_{\text{inf}}^{(i)}(t).$$

For individuals in the states S_2 and S_3, who already suffered from a previous infection, the force of infection $\boldsymbol{\lambda}(t)$ is multiplied by the corresponding relative risks α_k ($k = 2, 3$) to represent the higher immunity level gained through prior infection. Finally, the risk θ_i of developing symptoms during an infection also depends on the prior state of susceptibility $S_i, i = 1, 2, 3$.

Again note, that the force of infection defines only the rate at which susceptibles move from their respective states S_i to the subsequent disease states I_i and A_i. However, all other movements between the compartments are determined through fixed rates, which are neither time- nor state-dependent. The only other varying rates in the model are the demographic process defining the in- and outflow of the underlying population, i.e. births κ, deaths γ and migration m.

Altogether, the overall disease spread within the population is represented through 266 possible states and the corresponding transmission dynamics are governed by a system of ordinary differential equations given through (4.1). The ODE system holds for all age groups $j = 1, \ldots, n_A$. Thereby $\mathbb{I}_{\{A\}}(x)$ refers to the indicator function which is 1 for $x \in A$ and 0 otherwise. $N^{(j)}$ denotes the total size of age group j. For an overview of all utilized parameters and their interpretation see Table 4.1 (page 105).

4.1.4   Stochastic observation of new cases

The ODE system (4.1) models the absolute number of individuals in each state over time, which, among else, yields the varying number of currently infected individuals in each age group, i.e. the disease prevalence. Through an observational component we now link this model output with the actual data, which consist of weekly numbers of newly reported cases, i.e. the disease incidence, and are subject to stochasticity. Similar models for stochastic case reporting were also introduced in (Birrell et al., 2011; Dorigatti et al., 2012).

First note, that the output of the ODE system is deterministic given a specific parameter vector. Since the true disease spreading is certainly a random process, we can actually not expect the different epidemic seasons to proceed equally as they appear to be more or

$$\frac{dM^{(j)}}{dt} = \mathbb{I}_{\{1\}}(j)\kappa_t + \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}M^{(j-1)} - \left(\omega_0 + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)M^{(j)}$$

$$\frac{dS_1^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}S_1^{(j-1)} + \omega_0 M^{(j)} + \beta S_{2B}^{(j)} - \left(\lambda^{(j)}(t) + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_1^{(j)}$$

$$\frac{dI_1^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}I_1^{(j-1)} + \theta_1\lambda^{(j)}(t)S_1^{(j)} - \left(\mu + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)I_1^{(j)}$$

$$\frac{dA_1^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}A_1^{(j-1)} + (1-\theta_1)\lambda^{(j)}(t)S_1^{(j)} - \left(\omega + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)A_1^{(j)}$$

$$\frac{dS_2^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}S_2^{(j-1)} + \omega A_1^{(j)} + \beta S_{3B}^{(j)} - \left(\alpha_2\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_2^{(j)}$$

$$\frac{dS_{2A}^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}S_{2A}^{(j-1)} + \beta S_2^{(j)} - \left(\alpha_2\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{2A}^{(j)}$$

$$\frac{dS_{2B}^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}S_{2B}^{(j-1)} + \beta S_{2A}^{(j)} - \left(\alpha_2\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{2B}^{(j)}$$

$$\frac{dI_2^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}I_2^{(j-1)} + \theta_2\alpha_2\lambda^{(j)}(t)\left(S_2^{(j)} + S_{2A}^{(j)} + S_{2B}^{(j)}\right) - \left(\mu + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)I_2^{(j)}$$

$$\frac{dA_2^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}A_2^{(j-1)} + (1-\theta_2)\alpha_2\lambda^{(j)}(t)\left(S_2^{(j)} + S_{2A}^{(j)} + S_{2B}^{(j)}\right) - \left(\omega + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)A_2^{(j)}$$

$$\frac{dS_3^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}S_3^{(j-1)} + \omega\left(A_2^{(j)} + A_3^{(j)}\right) - \left(\alpha_3\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_3^{(j)}$$

$$\frac{dS_{3A}^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}S_{3A}^{(j-1)} + \beta S_3^{(j)} - \left(\alpha_3\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{3A}^{(j)}$$

$$\frac{dS_{3B}^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}S_{3B}^{(j-1)} + \beta S_{3A}^{(j)} - \left(\alpha_3\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{3B}^{(j)}$$

$$\frac{dI_3^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}I_3^{(j-1)} + \theta_3\alpha_3\lambda^{(j)}(t)\left(S_3^{(j)} + S_{3A}^{(j)} + S_{3B}^{(j)}\right) - \left(\mu + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)I_3^{(j)}$$

$$\frac{dA_3^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}A_3^{(j-1)} + (1-\theta_3)\alpha_3\lambda^{(j)}(t)\left(S_3^{(j)} + S_{3A}^{(j)} + S_{3B}^{(j)}\right) - \left(\omega + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)A_3^{(j)}$$

$$(4.1)$$

less pronounced in some years. Thus, the output of the ODE system, which suggests very similar disease patterns for every year, should be interpreted as the expected mean prevalence and thus provides 'only' the expectation structure for the observed data at each time point.

To obtain the expected number of newly infected cases from the model, we have to consider the derivatives of the states $I_i^{(j)}$ given in the ODE system (4.1). The instantaneous rate of new symptomatic infections $H^{(j)}(t)$ in age group j is given by the corresponding

derivatives in (4.1), i.e.

$$H^{(j)}(t) = \theta_1 \lambda^{(j)}(t) S_1^{(j)}(t) + \alpha_2 \theta_2 \lambda^{(j)}(t) \left(S_2^{(j)}(t) + S_{2A}^{(j)}(t) + S_{2B}^{(j)}(t) \right)$$
$$+ \alpha_3 \theta_3 \lambda^{(j)}(t) \left(S_3^{(j)}(t) + S_{3A}^{(j)}(t) + S_{3B}^{(j)}(t) \right).$$

Therefore, the expected number of newly symptomatically infected individuals $Y^{(j)}(t)$ in age group j in week t is

$$Y^{(j)}(t) = \int\limits_{t}^{t+1} H^{(j)}(u)\, du.$$

We differentiated between cases from the EFS (e) and WFS (w), i.e. the expected number of new infections in each age group is divided regionally

$$Y^{(j)}(t) = Y^{(j,w)}(t) + Y^{(j,e)}(t).$$

There is no evidence for differences in the underlying transmission dynamics between the two regions, e.g. different contact behaviour or hygienic measures. Thus, we assumed the actual underlying incidences (per 100,000 people) to be equal in both regions for each age group. This implies that the ratio of expected numbers of new cases $Y^{(j,i)}(t)$ ($i = e, w$) in age group j in the ESF and WFS is equal to the ratio of populations $\text{Pop}^{(j,i)}(t)$ ($i = e, w$) of age group j in these regions, i.e.

$$Y^{(j,e)}(t) = w^{(j,e)}(t) Y^{(j)}(t), \quad Y^{(j,w)}(t) = \left(1 - w^{(j,e)}(t) \right) Y^{(j)}(t), \tag{4.2}$$

with

$$w^{(j,e)}(t) = \frac{\text{Pop}^{(j,e)}(t)}{\text{Pop}^{(j,e)}(t) + \text{Pop}^{(j,w)}(t)}.$$

However, there is evidence that the differences in reported disease incidence arise from different habits of doctors (related to the ordering of diagnostic tests for patients with diarrhoea) and also different healthcare seeking behavior of parents with sick children in the two regions (Rosner et al., 2010). Thus, we introduced a time- and region-specific detection rate $h^{(i)}(t)$ ($i = e, w$), which gives the probability of a symptomatic case to be reported. To avoid overparameterisation and since no auxiliary information is available, we assume this detection rate to be age-independent, although an age-dependent detection rate might be more realistic. With the change in reimbursement happening at the end of 2004, the detection rate $h^{(i)}(t)$ was assumed to be constant in each region for the time before 2005 and after the first half year of 2005 with linear growth in between. Hence, the time-dependent region-specific detection rates $h^{(i)}(t)$ were defined as

$$h^{(i)}(t) = \begin{cases} h^{(i)} & \text{if } t \le t_{2005} & \text{(until 2005)} \\ h^{(i)} + (q^{(i)} h^{(i)} - h^{(i)}) \frac{t - t_{2005}}{26} & \text{if } t_{2005} < t \le t_{2005} + 26 & \text{(first half year 2005)} \\ q^{(i)} h^{(i)} & \text{if } t > t_{2005} + 26 & \text{(after first half year 2005)} \end{cases}$$

with $h^{(i)}$ and $q^{(i)}$ $(i = e, w)$ representing the detection rates before the year of 2005 and their relative increase afterwards, respectively, and t_{2005} denoting the first week of year 2005.

Finally, the weekly number of reported cases $X^{(j,i)}(t)$ in age group j and region i was assumed to be negative binomially distributed

$$X^{(j,i)}(t) \sim \text{NegBin}\left(h^{(i)}(t) \cdot Y^{(j,i)}(t), d\right)$$

with expectation $h^{(i)}(t) \cdot Y^{(j,i)}(t)$ origining from the ODE model and dispersion coefficient d. A negative binomial distribution was chosen over a Poisson distribution as it allows for a higher variance to compensate for stochasticity in disease transmission which could not be accounted for by the deterministic model, and since the overdispersion was also suggested by the data. To keep the number of parameters low the same dispersion parameter d was applied to all age groups and both regions.

4.2 Bayesian inference and model averaging

In applied infectious disease epidemiology, calibration of dynamic transmission models has been commonly based on results from external epidemiological and clinical studies or on expert opinion in the past. This means, that most parameters were fixed at previously estimated values whereas only a few uncertain model parameters were left to be estimated through data. However, transmission models based on ODEs are known to behave very sensitively to their input parameters, which is why a data driven inference for (almost) all parameters seems more appropriate for these models, instead of solely using estimates from detached previous results. A Bayesian framework incorporating this external knowledge as prior information seems especially suited for this task.

Furthermore, we want to address the impact of residual autocorrelation, which is an inherent issue in infectious disease modelling, due to the ODE system providing the expected incidence progress for each season while the observed epidemic course may considerable deviate from this expectation for some seasons. Thus, ignoring this systematic deviation from the mean, i.e. autocorrelation of residudals, may lead to bias in parameter inference, especially underestimation of the parameter variance. To account for this bias, we applied our approach based on readjusting the likelihood function given the residual autocorrelation in the data as presented in Section 2.2. This procedure is particularly suitable for our complex model, since it does not require the estimation of additional parameters and therefore does not further increase the model's complexity.

4.2.1 Prior elicitation

Altogether, a 17-dimensional parameter vector $\boldsymbol{\vartheta}$ regarding the model from Section 4.1.2 is to be estimated, i.e.

$$\boldsymbol{\vartheta} = \Big(\mu, \omega, \omega_0, \beta \in \mathbb{R}_{>0}, \mathrm{p} \in \mathbb{R}_{>0}, \mathrm{a} \in \mathbb{R}^2, \mathrm{b} \in (0,1)^2, \mathrm{c} \in \mathbb{R}^3_{>0},$$

$$h^{(e)}, h^{(w)} \in (0,1), q^{(e)}, q^{(w)} \in \mathbb{R}_{>0}, \mathrm{d} \in \mathbb{R}_{>0}\Big).$$

See Table 4.1 for an explanation of the single parameter components. The vector $\boldsymbol{\vartheta}$ includes many parameters with an intuitive epidemiological interpretation, for instance the recovery rates from symptomatic or asymptomatic infection μ and ω, respectively. In order to construct prior distributions, the epidemiological literature was searched for estimated values and corresponding uncertainty, e.g. given through 95% confidence intervals. For some parameters, in particular those concerning transmission, i.e. the components of the contact matrix $\boldsymbol{C}$ or the relative infectiousness with symptomatic infection p, reliable evidence was not available in the literature. Hence, we assumed vague priors yielding a wide range for these parameters. On their respective natural space most parameters are restricted to the positive real axis or to the interval $(0,1)$. To take these parameter restrictions into account, we first applied an appropriate log- or logit-transformation if necessary. This allows us to quantify priors on the whole real axis such that we are able to use the same distributional family for each parameter, regardless of its natural space. We chose to construct the individual prior densities on the unrestricted transformed parameter space via a skew normal distribution ($\mathcal{SN}$) (Azzalini, 1985), since it offers three distribution parameters to be calibrated as opposed to, e.g., the beta or gamma distribution, making it a suitable choice in the case of prior information available as point estimators and a confidence interval. The $\mathcal{SN}$-density to the parameters m, σ and s is given by

$$\psi_{m,\sigma,s}(x) = \frac{2}{\sigma}\phi\left(\frac{x-m}{\sigma}\right)\Phi\left(s\left(\frac{x-m}{\sigma}\right)\right),$$

where ϕ and Φ refer to the standard normal distribution's density and distribution function, respectively.

We transformed the estimated values and borders of the reported 95% confidence interval of each parameter via the log or logit transformation and interpreted the resulting values as the 0.025, 0.5 and 0.975-quantile of the corresponding prior distribution on the transformed parameter space. This interpretation allowed us to make the log or logit transformation in the first place, since quantiles are invariant under monotone transformation whereas expected values would be not. Then for each model parameter the $\mathcal{SN}$-distribution parameters m, σ and s were computed such that the corresponding skew normal distribution fits the three desired quantiles on the transformed space. To do so, we numerically mimimized the function

$$\mathrm{Dev}(m,\sigma,s) = \left|\int_0^{q_{0.025}}\psi_{m,\sigma,s}(x)dx - 0.025\right| + \left|\int_0^{q_{0.5}}\psi_{m,\sigma,s}(x)dx - 0.5\right| + \left|\int_0^{q_{0.975}}\psi_{m,\sigma,s}(x)dx - 0.975\right|,$$

where $q_{0.025}$, $q_{0.5}$ and $q_{0.975}$ refer to the respective quantiles, which were used for fitting the $\mathcal{SN}$-distribution. The parameters (m, σ, s) yielding the desired quantiles were then used for defining the prior density of the corresponding model parameter.

To reobtain the according prior distribution on the natural parameter space, we applied density transformation which led to the following density functions on the corresponding restricted spaces

$$f_X(x) = \frac{\psi_{m,\sigma,s}\left(\text{logit}\left(x\right)\right)}{x(1-x)}, \quad \text{if } \text{logit}\left(X\right) \sim \mathcal{SN}(m, \sigma, s)$$

or

$$f_X(x) = \frac{\psi_{m,\sigma,s}\left(\log(x)\right)}{x}, \quad \text{if } \log(X) \sim \mathcal{SN}(m, \sigma, s).$$

For most parameter configurations these density functions yield shapes comparable to those of a beta or gamma distribution on their respective scale, but offer more flexibility due to the three parameters available.

The joint prior distribution π of all parameters was then constructed as the product of the individual prior densities assuming independence between the parameters. See Table 4.1 for the prior distribution of each parameter and the corresponding literature evidence to inform the prior quantiles.

In order to avoid overparameterisation of the model, we exclusively fixed the parameters on infection induced immunity, which we found to have the highest collinearity to other model parameters. Based on epidemiological studies (Fischer et al., 2002; Velásquez et al., 1996; Mrukowicz et al., 1999) the relative risk for a subsequent infection is defined at a factor of $\alpha_2 = 0.6$ and $\alpha_3 = 0.4$ after one or two infections, respectively. The probability of developing symptoms is $\theta_1 = 0.5$, $\theta_2 = 0.25$, and $\theta_3 = 0$ with the first, second, and third infection, respectively.

4.2.2 Approximate posterior distribution

Initially, the likelihood of a single parameter vector $\boldsymbol{\vartheta}$ was computed by assuming that each data point $D_t^{(j,i)}$, i.e. number of reported cases corresponding to age group j, region i and week t, is independently negative binomially distributed with expectation $h_{\boldsymbol{\vartheta}}^{(i)}(t)Y_{\boldsymbol{\vartheta}}^{(j,i)}(t)$ and dispersion parameter $d_{\boldsymbol{\vartheta}}$ as described in Section 4.1.4. Hereby, subscript $\boldsymbol{\vartheta}$ indicates the dependence of the corresponding variables on the model input parameter. Thus, the loglikelihood corresponding to the underlying independent negative binomial distribution of the observed case numbers $\boldsymbol{D} = \left(D_t^{(j,i)}\right)$ is given by

$$\text{LL}\left(\boldsymbol{D} \,|\, \boldsymbol{\vartheta}\right) = \sum_{t=1}^{T}\sum_{j=1}^{n_D}\sum_{i=e,w} \log\left\{p_{\text{NegBin}}\left(D_t^{(j,i)} \,\Big|\, h_{\boldsymbol{\vartheta}}^{(i)}(t)\bar{Y}_{\boldsymbol{\vartheta}}^{(j,i)}(t), d_{\boldsymbol{\vartheta}}\right)\right\},$$

where $p_{\text{NegBin}}\left(\cdot \,|\, \mu, d\right)$ is the probability mass function of the negative binomial distribution with expectation μ and dispersion d. Furthermore, $\bar{Y}$ denotes the predicted incidences

Table 4.1: Prior distributions of all epidemiological model parameters are given as skew normal distribution on an appropriate transformed scale. Literature evidence where available is given in the last column. For completeness, parameters not subject of estimation and corresponding literature evidence are also provided.

Parameter	Interpretation	Prior distribution	Source
α_1, α_2	Reduced susceptibility after previous infections	fixed at $(0.6; 0.4)$	Fischer et al.; Velásquez et al.; Mrukowicz et al.
$\theta_1, \theta_2, \theta_3$	Probability of developing symptoms	fixed at $(0.5; 0.25; 0)$	Fischer et al.; Velásquez et al.; Mrukowicz et al.
μ	Rate of recovery from symptomatic infection (1/duration in weeks)	$\log \mu \sim \mathcal{SN}(0.69, 0.31, 83)$	Anderson and Weber
ω	Rate of recovery from asymptomatic infection (1/duration in weeks)	$\log \omega \sim \mathcal{SN}(0.08, 0.34, -1.38)$	Pickering et al.; Wilde et al.
ω_0	Rate of waning maternal protection (1/duration in weeks)	$\log \omega_0 \sim \mathcal{SN}(-2.11, 0.35, 0)$	Clark et al.; Clemens et al.; Pickering et al.
p	Relative infectiousness of symptomatically infected individuals	$\operatorname{logit} p \sim \mathcal{SN}(0, 1.12, 0)$	(Assumption)
β	Rate of waning natural immunity	$\log \beta \sim \mathcal{SN}(-3.30, 0.43, -0.13)$	Anderson and Weber; Fischer et al.; Ward et al.
c_1, c_2, c_3	Contact rates	$\log c_i \sim \mathcal{SN}(0, 0.82, 0)$	(Assumption)
a_1, a_2	Amplitude of seasonality	$a_i \sim \mathcal{SN}(0, 1.02, 0)$	(Assumption)
b_1, b_2	Phase shift of seasonality	$\operatorname{logit} b_i \sim \mathcal{SN}(0, 1.87, 0)$	(Assumption)
$h^{(e)}, h^{(w)}$	Case detection rates in eastern and western federal states before 2005	$\operatorname{logit} h^{(i)} \sim \mathcal{SN}(-1.87, 0.47, -2.94)$	Soriano-Gabarró et al.
$q^{(e)}, q^{(w)}$	Relative increase of detection rate in eastern and western federal states	$\log q^{(i)} \sim \mathcal{SN}(0, 0.36, 0)$	(Assumption)
d	Dispersion parameter of the observations' negative binomial distribution	$\log d \sim \mathcal{SN}(-0.01, 0.82, 0)$	(Assumption)
κ_t	Birth rates	fixed	Federal Bureau of Statistics
δ	Ageing rates	fixed	(Federal Bureau of Statistics, 2013)
γ_t	Death rates	fixed	(Federal Bureau of Statistics, 2013)
m_t	Migration rates	fixed	(Federal Bureau of Statistics, 2013)

coming from the ODE model, where $\bar{Y}$ may be an aggregation of some of the n_A age strata defined in the model, such that they correspond to the n_D age groups available in the data, if necessary (in particular the lower age groups had a finer decomposition within the model).

As a first step we then computed the unnormalized posterior logdensity $\log \pi \left(\boldsymbol{\vartheta} \,|\, \boldsymbol{D} \right)$ of a given parameter vector $\boldsymbol{\vartheta}$ as the sum of the prior logdensity $\log \pi(\boldsymbol{\vartheta})$ and the above loglikelihood $\mathrm{LL}\left(\boldsymbol{D} \,|\, \boldsymbol{\vartheta} \right)$, i.e.

$$\log \pi \left(\boldsymbol{\vartheta} \,|\, \boldsymbol{D} \right) = \mathrm{LL}\left(\boldsymbol{D} \,|\, \boldsymbol{\vartheta} \right) + \log \pi \left(\boldsymbol{\vartheta} \right), \tag{4.3}$$

which is the classical definition of the posterior distribution in Bayesian inference once you account for the normalization constant (Gelman et al., 2013). Unfortunately, the computational effort required for computing the normalized posterior distribution in this complex model setting is quite large. Typically this computation would be done by Markov Chain Monte Carlo techniques (see Section 3.2), which require a few hundred thousand evaluations of the posterior to guarantee convergence of the sample distribution because of the large number of parameters to be estimated within this multivariable setting. Since a single evaluation of our model and the likelihood is computationally quite expensive, we instead relied on the posterior being asymptotically normal around the posterior mode as discussed in Section 2.1.1. Due to the size of our dataset consisting of more than 8,000 observations the assumption of a nearly asymptotic posterior appears to be fair, since the asymptotic properties already hold for smaller sample sizes as shown in the simulation study in Section 2.3. Therefore, based on the unnormalized posterior logdensity (4.3) we numerically computed the posterior mode using an alternating combination of gradient and Nelder-Mead simplex methods (Press et al., 2007). With every start of the Nelder-Mead method the algorithm allowed the optimum to escape from a potential local maximum while intermediate gradient steps improved the overall performance of the algorithm, which eventually yields the posterior mode $\boldsymbol{\vartheta}^*$ For more details see Section 6.1 explaining the implementation of our methods.

However, considering the potentially high residual autocorrelation within the data, we were concerned about overemphasising the strength of the data subject to the independence assumption given the model. The remaining autocorrelation becomes particularly apparent when looking at the residuals from such modelling, i.e. the deviation of the data to the expected case numbers coming from the ODE model, which still exhibit strong autocorrelation. Holding on to the assumption of independent observations would result in an overrating of the likelihood and consequently in an underestimation of the posterior variance as discussed in Section 2.2.

Therefore, in a second step we adjusted the likelihood for the actual information content of the data expressed by the autocorrelation of the residuals. This was done by estimating the cumulative autocorrelation $\widehat{\mathrm{CA}}^{(j,i)}(\boldsymbol{\vartheta}^*)$ for each given time series, i.e. each age group and both regions, based on the optimal calibration of the model after the first step using the posterior mode $\boldsymbol{\vartheta}^*$. The cumulative autocorrelation may be interpreted as a measure

for the effective sample size as defined in (Thiébaux and Zwiers, 1984). Assuming that the loglikelihood $\text{LL}\left(\boldsymbol{D}\,|\,\boldsymbol{\vartheta}\right)$ of a parameter vector $\boldsymbol{\vartheta}$ is not based on T independent observations for each time series, but on a number of independent observations given by $\widehat{\text{CA}}(\boldsymbol{\vartheta}^*)^{-1}T$, we discounted the likelihood by that number as suggested similarly in McMillan et al. (2010). This might seem to be an ad-hoc adjustment method, but it has a theoretical motivation provided in Section 2.2.1 and allows a valid inference without having to incorporate dependent observations directly within the model, which presents a much more complicated task.

The actual procedure for estimating the cumulative autocorrelation $\widehat{\text{CA}}^{(j,i)}(\boldsymbol{\vartheta}^*)$ for a certain age group j and region i, given the first step posterior mode $\boldsymbol{\vartheta}^*$, is explained in detail in Section 4.2.2 below.

Altogether, this leads to an adjusted step-2 loglikelihood given by

$$\text{LL}_{\text{CA}}\left(\boldsymbol{D}\,|\,\boldsymbol{\vartheta}\right) = \sum_{t=1}^{T}\sum_{j=1}^{n_D}\sum_{i=e,w} \widehat{\text{CA}}^{(j,i)}(\boldsymbol{\vartheta}^*)^{-1} \log\left\{ p_{\text{NegBin}}\left(D_t^{(j,i)}\,\Big|\,h_{\boldsymbol{\vartheta}}^{(i)}(t)\bar{Y}_{\boldsymbol{\vartheta}}^{(j,i)}(t), d_{\boldsymbol{\vartheta}}\right)\right\}.$$

During this second step, the final unnormalized posterior logdensity $\log \pi_{\text{CA}}\left(\cdot\,|\,\boldsymbol{D}\right)$ incorporating the loglikelihood LL_{CA} is subject of optimization, i.e.

$$\log \pi_{\text{CA}}\left(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}\right) = \text{LL}_{\text{CA}}\left(\boldsymbol{D}\,|\,\boldsymbol{\vartheta}\right) + \log \pi\left(\boldsymbol{\vartheta}\right), \tag{4.4}$$

leading to a second step posterior mode $\boldsymbol{\vartheta}_{\text{CA}}^*$. Computation of the second step posterior mode was done analogously to the first step. Based on the asymptotic properties (see Section 2.1.1) our final approximate posterior distribution is defined as normally distributed with expectation $\boldsymbol{\vartheta}_{\text{CA}}^*$ and covariance matrix $-H^{-1}$, where H is the Hessian matrix of $\log \pi_{\text{CA}}\left(\cdot\,|\,\boldsymbol{D}\right)$ evaluated at $\boldsymbol{\vartheta}_{\text{CA}}^*$, i.e.

$$\boldsymbol{\vartheta}\,|\,\boldsymbol{D} \sim \mathcal{N}\left(\boldsymbol{\vartheta}_{\text{CA}}^*, -\left(\left.\frac{\partial^2}{\partial\boldsymbol{\vartheta}^2}\log \pi_{\text{CA}}\left(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}\right)\right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_{\text{CA}}^*}\right)^{-1}\right). \tag{4.5}$$

Note, that as one necessary convergence criteria of the optimization process we required the Hessian H at the final optimum to be negative definite such that the resulting posterior covariance is well defined. Negative definiteness of the Hessian at the posterior mode can be secured due to the mode being a local maximum of the twice-differentiable log posterior density.

Estimating the cumulative autocorrelation of the rotavirus incidence data

The cumulative autocorrelation is a measure for the number of effectively independent data points within a time series similar to the effective sample size presented in (Thiébaux and Zwiers, 1984). In our inference framework we use the CA as an adjustment factor for the

loglikelihood function to account for the reduced information content within dependent samples. For the rational behind adjusting the likelihood based on the CA see also 2.2 providing an analytical assessment and simulation based results. The CA of a time series $\boldsymbol{X} = (X_t)_{t=1,\dots,T}$ is defined by

$$\text{CA} = \left(\sum_{\tau=-T}^{+T} \hat{\rho}(\tau) \right), \tag{4.6}$$

where $\hat{\rho}(\tau)$ are estimates of the symmetric autocorrelation of lag τ within the time series $\boldsymbol{X}$.

In our application there are $2 \cdot n_D$ time series to analyse, resulting from the n_D age groups and the two regions supported by the data. For each of the time series we carried out an Anscombe transformation of the residuals (Hardin and Hilbe, 2007) with the aim of making the residuals to be distributed "as normal as possible". The reasoning behind this transformation is that all dependencies within Gaussian random vectors are contained in their correlation coefficient. Thus, all monotonic dependencies are transformed into linear dependencies, which can then be measured through the correlation. The series of Anscombe residuals $\left(r_t^A \right)_{t=1,\dots,T}$ for a time series $\boldsymbol{X} = (X_t)_{t=1,\dots,T}$ is defined as

$$r_t^A = \frac{A(x_t) - A(\hat{\mu}_t)}{V(\hat{\mu}_t)^{\frac{1}{6}}},$$

with $(\hat{\mu}_t)_{t=1,\dots,T}$ being the series of estimated expectations at each time point and the function A being

$$A(x) = \int_{-\infty}^{x} V(\mu)^{-\frac{1}{3}} d\mu, \tag{4.7}$$

where $V(\mu)$ denotes the expectation dependent variance function. In the case of a negative binomial response $f(\cdot \,|\, \mu, d)$ this function takes the form $V(\mu) = \mu + \mu^2/d$. Hence, (4.7) leads to

$$r_t^A = \frac{\frac{3}{2} X_t^{\frac{2}{3}} \,_2F_1\left(\frac{1}{3}, \frac{2}{3}, \frac{5}{3}, -\frac{X_t}{\hat{d}}\right) - \frac{3}{2} \hat{\mu}_t^{\frac{2}{3}} \,_2F_1\left(\frac{1}{3}, \frac{2}{3}, \frac{5}{3}, -\frac{\hat{\mu}_t}{\hat{d}}\right)}{\left(\hat{\mu}_t + \frac{\hat{\mu}_t^2}{\hat{d}}\right)^{\frac{1}{6}}}.$$

Here $_2F_1$ denotes the Gaussian hypergeometric function (Abramowitz and Stegun, 1964). We applied this transformation to each of the 20 time series $(D_t^{(j,i)})_{t=1,\dots,T}$ with the corresponding expectations

$$\hat{\mu}_t^{(j,i)} = h_{\boldsymbol{\vartheta}^*}^{(i)}(t) \cdot Y_{\boldsymbol{\vartheta}^*}^{(j,i)}(t)$$

computed through our model and dispersion $\hat{d} = d_{\vartheta^*}$ given by the posterior mode ϑ^* from the first optimization step.

To obtain an estimation for the autocorrelations $\rho(\tau)$ within the newly constructed time series $\left(r_t^A \right)_{t=1,\dots,T}^{(j,i)}$ we fitted an ARMA(p,q)-process to each time series separately.

The orders p^* and q^* of the process were chosen through the Akaike information criteria (see Section 2.1.2) and may differ for each time series. From the coefficient estimates for the ARMA(p^*, q^*) process we computed the autocorrelation function $\rho(\tau)$ as described in Brockwell and Davis (1991, Ch.3.2). We preferred this parametric ARMA(p, q)-based estimation for the ACF $\hat{\rho}(\tau)$ over the direct empirical ACF estimates, i.e.

$$\hat{\rho}(\tau) = \frac{(T-1)}{(T-\tau)} \frac{\sum_{t=1}^{T-\tau} r_t^A \cdot r_{t+\tau}^A}{\sum_{t=1}^{T} (r_t^A)^2},$$

as the latter appeared to suffer from high variance, especially for high-lag autocorrelations.

Using the respective estimated autocorrelation functions $\hat{\rho}$ from the fitted ARMA(p, q)-model for each time series $(D_t^{(j,i)})_{t=1,\ldots,T}$ ($j = 1, \ldots, 10$, $i = e, w$) we computed the estimated cumulative autocorrelation $\widehat{\text{CA}}^{(j,i)}(\boldsymbol{\vartheta}^*)$ based on the posterior mode $\boldsymbol{\vartheta}^*$ computed in the first step according to the definition (4.6).

4.2.3 Model averaging based on posterior distributions

In addition to uncertainty in parameter estimation, we also wanted to address uncertainty in model selection by applying Bayesian model averaging techniques. The set of different models results from the choice between 6 different contact structures, represented through different contact matrices $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_6 \in R_{>0}^{19 \times 19}$, and 3 possible parameter spaces: $\Theta \in \left\{ \Theta_{(\mu, \omega)}, \Theta_{(\mu)}, \Theta_{()} \right\}$, where the lower index indicates the parameters which were fixed, i.e. not subject of estimation. The contact matrices we investigated are an extended pool of contact structures already studied in (Pitzer et al., 2009; Atchison et al., 2010; de Blasio et al., 2010). A simplified representation of the six different types of contact matrix is shown below, where each matrix is given as a composition of submatrices regarding either the first 14, the subsequent 3, or the last 2 rows and columns, which correspond to the age groups young (0-4 years), middle (5-59 years), and old (60+ years), respectively. All components within one submatrix were assumed to be equal to one parameter c_i. Additionally, certain submatrices were modelled to share the same component value c_i, according to the contact patterns shown below. Thus, each matrix calibration except for $\boldsymbol{C}_1$ depends on 3 parameters $c_1, c_2, c_3 > 0$.

$$\boldsymbol{C}_1 = \begin{pmatrix} c_1 & c_1 & c_1 \\ c_1 & c_1 & c_1 \\ c_1 & c_1 & c_1 \end{pmatrix}, \quad \boldsymbol{C}_2 = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{pmatrix}, \quad \boldsymbol{C}_3 = \begin{pmatrix} c_1 & c_1 & c_1 \\ c_2 & c_2 & c_2 \\ c_3 & c_3 & c_3 \end{pmatrix},$$

$$\boldsymbol{C}_4 = \begin{pmatrix} c_1 & c_1 & c_3 \\ c_1 & c_2 & c_3 \\ c_3 & c_3 & c_3 \end{pmatrix}, \quad \boldsymbol{C}_5 = \begin{pmatrix} c_1 & c_2 & c_3 \\ c_2 & c_2 & c_3 \\ c_3 & c_3 & c_3 \end{pmatrix}, \quad \boldsymbol{C}_6 = \begin{pmatrix} c_1 & c_3 & c_3 \\ c_3 & c_2 & c_3 \\ c_3 & c_3 & c_3 \end{pmatrix}. \tag{4.8}$$

Regarding the parameter space Θ, we decided to examine the impact of fixing parameters beforehand, especially regarding the effects on model fit and the estimation of other relevant model parameters. For that investigation we chose to analyse parameters determining the duration of illness or rather infectiousness, since the corresponding evidence has a high degree of uncertainty and we expected those parameters to significantly correlate with other model parameters. In the cases of parameters being fixed, they were fixed at their respective point estimates as shown in Table 4.1, i.e. $\mu = 7/3$ and $\omega = 7/8$.

Given an ensemble of models $\mathcal{M} = \{M_i\}_{i \in I}$ the Bayes factor $B_{i,0}$ of a specific model M_i was computed by

$$B_{i,0} = \frac{f^{(\mathcal{M})}(\boldsymbol{D} \,|\, M_i)}{f^{(\mathcal{M})}(\boldsymbol{D} \,|\, M_0)},$$

with M_0 being some reference model and $f^{(\mathcal{M})}(\boldsymbol{D} \,|\, M_i)$ being the marginal likelihood of model M_i (see Section 2.1.2). As the posterior in our case is approximated by a normal distribution, the marginal likelihood can be computed exactly via

$$
\begin{aligned}
f^{(\mathcal{M})}(\boldsymbol{D} \,|\, M_i) &= \int_{\Theta} \exp\left\{ \mathrm{LL}_{\mathrm{CA}}(\boldsymbol{D} \,|\, M_i, \boldsymbol{\vartheta}) \right\} \pi(\boldsymbol{\vartheta} \,|\, M_i) \, d\boldsymbol{\vartheta} \\
&= (2\pi)^{\frac{d_i}{2}} \det(-H_i)^{-\frac{1}{2}} \exp\left\{ \mathrm{LL}_{\mathrm{CA}}\left(\boldsymbol{D} \,\middle|\, M_i, \boldsymbol{\vartheta}^*_{\mathrm{CA},i}\right) \right\} \pi\left(\boldsymbol{\vartheta}^*_{\mathrm{CA},i} \,\middle|\, M_i\right),
\end{aligned}
\tag{4.9}
$$

where d_i is the dimension of the model specific input parameter $\boldsymbol{\vartheta}_i$ and H_i is the Hessian matrix of the model's adjusted log posterior $\log \pi_{\mathrm{CA}}(\cdot \,|\, M_i, \boldsymbol{D})$ evaluated at its respective mode $\boldsymbol{\vartheta}^*_{\mathrm{CA},i}$ as above. In the case of a non-normal posterior the above term corresponds to the marginal likelihood approximation based on a Taylor expansion as suggested in (Raftery, 1996).

To guarantee comparability of the models we had to secure that the final posterior distribution of each model was adjusted by the same cumulative autocorrelation factors. Otherwise, high autocorrelations measured in the first step would lead to higher posterior mode values in the second step optimization, and thus to a higher marginal likelihood. This seems counterintuitive since high residual autocorrelations in the data are typically an indicator for a bad model fit. Adjusting all models via the same cumulative autocorrelation factors $\{\mathrm{CA}^{(i,j)}\}$ avoids this malfunction. Therefore, to choose one set of cumulative autocorrelation estimators $\{\widehat{\mathrm{CA}}^{(i,j)}\}$, we computed for each time series $\boldsymbol{D}^{(i,j)}$ the mean cumulative autocorrelation $\{\overline{\mathrm{CA}}^{(i,j)}\}$, which was obtained by averaging the model specific cumulative autocorrelations $\mathrm{CA}^{(i,j)}(M_i)$ over all $m = 18$ models, i.e.

$$\overline{\mathrm{CA}}^{(i,j)} = \frac{1}{|\mathcal{M}|} \sum_{M_i \in \mathcal{M}} \mathrm{CA}^{(i,j)}(M_i).$$

In the second step of our optimization procedure we then proceeded with those mean cumulative autocorrelations in order to adjust the likelihood of all considered models instead of using the respective model specific estimates $\{\mathrm{CA}^{(i,j)}(M_i)\}$. Since the variation of

the estimated cumulative autocorrelations $\mathrm{CA}^{(i,j)}$ for a specific time series (i,j) given the $m = 18$ models is expected to be small, taking the mean is a convenient choice.

Finally, for model averaging we computed the individual model weights w_i from the Bayes factors assuming equal prior model probabilities resulting in

$$
w_i = \frac{f^{(\mathcal{M})}\left(\boldsymbol{D}\,|M_i\right)}{\displaystyle\sum_{M_j \in \mathcal{M}} f^{(\mathcal{M})}\left(\boldsymbol{D}\,|M_j\right)}.
$$

Note the possibility of using only a subset of models defined above for the averaging procedure. Two classes of subsets are of particular interest: those models, that share a common contact structure $\boldsymbol{C}$, and those models, that share the same parameter space Θ, which means that the parameters μ and ω are either fixed or subject of estimation for each considered model. We will refer to those model subsets as horizontal and vertical averaging, respectively.

4.2.4 Computing the averaged model predictions

Knowing the weights $\{w_i\}$ corresponding to an ensemble of models $\mathcal{M} = \{M_i\}_{i \in I}$, we computed the averaged predictive distribution for the age- and region-stratified incidence $X_t^{(i,j)}$ and its expectation $h_t^{(i)} \cdot Y_t^{(i,j)}$ for any time t by sampling a model and then sampling from its respective posterior distribution. To obtain the desired quantities, we applied the following algorithm, Algorithm 5, to our considered ensemble of models.

The resulting sample consisting of K arrays for each of

$$
\boldsymbol{X} = \left(X_t^{(i,j)}\right) \quad \text{and} \quad \mathbb{E}\left[\boldsymbol{X}\right] = \left(\mathbb{E}\left[X_t^{(i,j)}\right]\right) = \left(h_t^{(i)} \cdot Y_t^{(i,j)}\right)
$$

has the desired distribution given by:

$$
\begin{aligned}
\mathbb{P}\left(X_t^{(i,j)} = x\right) &= \sum_{M_i \in \mathcal{M}} w_i \cdot \mathbb{P}\left(X_t^{(i,j)} = x \,\middle|\, M_i\right) \\
\text{and} \quad \varphi\left(\mathbb{E}\left[X_t^{(i,j)}\right] = x\right) &= \sum_{M_i \in \mathcal{M}} w_i \cdot \varphi\left(\mathbb{E}\left[X_t^{(i,j)}\right] = x \,\middle|\, M_k\right),
\end{aligned} \tag{4.10}
$$

where $\mathbb{P}$ and φ denote the probability mass function of $\boldsymbol{X}$ and density function of $\mathbb{E}\left[\boldsymbol{X}\right]$, respectively.

4.2.5 The averaged posterior distribution

Given a set of weighted models $\{M_i, w_i\}_{i \in I}$ the averaged posterior distribution $\bar{\pi}$ can be computed as a weighted sum of the corresponding cumulative distribution functions:

$$
\bar{\pi}\left(\boldsymbol{\vartheta} \le \boldsymbol{\vartheta}^{(0)}\,\middle|\, \boldsymbol{D}\right) = \sum_{i \in I} w_i \cdot \pi_{\mathrm{CA}}^{M_i}\left(\boldsymbol{\vartheta} \le \boldsymbol{\vartheta}^{(0)}\,\middle|\, \boldsymbol{D}\right). \tag{4.11}
$$

Algorithm 5: Predictive incidence sampling from an prespecified ensemble of models

---

Input: $\mathcal{M} = \{M_i\}_{i \in I}$: an ensemble of models to simulate case number predictions from the respective posterior

Input: $\{w_i\}_{i \in I}$: a set of model weight corresponding to $\mathcal{M}$

Output: $\left\{\boldsymbol{h}_{(k)}^{(i)} \cdot \boldsymbol{Y}_{(k)}^{(i,j)}\right\}_{k=1,\dots,K} \left\{\boldsymbol{X}_{(k)}^{(i,j)}\right\}_{k=1,\dots,K}$: predictive samples for the expected and reported number of cases for each age group and region

- **for** $k = 1$ **to** K **do**

  1. Draw a model $M^{(k)}$ from $\{M_i\}_{i \in I}$ according to the probabilities $\{w_i\}_{i \in I}$

  2. Draw a parameter vector $\boldsymbol{\vartheta}^{(k)}$ according to the posterior distribution $\pi_{\mathrm{CA}}^{M^{(k)}}(\cdot \,|\, \boldsymbol{D})$ corresponding to model $M^{(k)}$

  3. Compute the expected incidences $h_t^{(i)} \cdot Y_t^{(i,j)}$ resulting from the respective ODE system of model $M^{(k)}$ subject to parameter $\boldsymbol{\vartheta}^{(k)}$ and draw a sample for the observed incidence
     $X_t^{(i,j)} \sim \mathrm{NegBin}\left(h_t^{(i)} \cdot Y_t^{(j,i)}, d_{\boldsymbol{\vartheta}^{(k)}}\right)$ for $t = 1,\dots,T$, $i = e, w$
     and $j = 1,\dots,n_D$

  end

Hereby, the set multidimensional set $\left\{\boldsymbol{\vartheta} \leq \boldsymbol{\vartheta}^{(0)}\right\} \subseteq \Theta$ is defined by

$$\left\{\boldsymbol{\vartheta} \leq \boldsymbol{\vartheta}^{(0)}\right\} = \left\{\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_d) \,\middle|\, \forall i = 1, \ldots, d : \vartheta_i \leq \vartheta_i^{(0)}\right\}.$$

By interpreting fixed parameters as a probability distribution with its entire mass concentrated in one point we can thus compute an averaged cumulative posterior distribution function, even if not all considered models share the same parameter space such as in the case of vertical averaging, where an averaging of the posterior densities would not be possible.

In the case of horizontal averaging, since all models share a common parameter space, we get an averaged probability density computed equivalently:

$$\bar{\pi}\left(\boldsymbol{\vartheta} \,\middle|\, \boldsymbol{D}\right) = \sum_{i \in I} w_i \cdot \pi^{M_i}\left(\boldsymbol{\vartheta} \,\middle|\, \boldsymbol{D}\right),$$

where π^{M_i} denotes the posterior density corresponding to model M_i. Such a representation is also possible for the marginal posterior densities of individual parameters ϑ_i in $\boldsymbol{\vartheta}$, if the parameter was subject of estimation within all models.

4.3   Application to German rotavirus incidence data

This section contains the results of modelling the rotavirus case notification data from Germany presented in Section 4.1. We computed the marginal likelihood corresponding to the first and second step log posterior density as described in Section 4.2 for all 18 models originating from the six different contact patterns $\boldsymbol{C}_i$ ($i = 1, \ldots, 6$) and the three possible parameter configurations $\Theta \in \left\{\Theta_{(\mu,\omega)}, \Theta_{(\mu)}, \Theta_{()}\right\}$ (the index indicates the parameters which were fixed in advance). Based on the marginal likelihoods we applied Bayesian averaging to a selected set of models assuming equal prior probabilities, as described in the following.

In our base case analysis we selected all 18 models to be subject of averaging. The corresponding model weights with respect to the adjusted posterior densities are displayed in the left panel of Table 4.2. When comparing the posterior model probabilities, it becomes clear that those models using contact pattern $\boldsymbol{C}_6$ reached the highest marginal likelihoods. This coincides with the results from the European POLYMOD study on social mixing patterns (Mossong et al., 2008), since contact matrix $\boldsymbol{C}_6$ yields the closest fit to the German contact pattern derived in this study. The contact and parameter combination $\left(\boldsymbol{C}_6, \Theta_{()}\right)$ resulted in a marginal likelihood with a log difference of more than 4 compared with any other model, yielding a model probability of $w\left(\boldsymbol{C}_6, \Theta_{()}\right) = 0.977$ and hence dominating all other models. The model weights computed in this base case analysis were used later for computing predictive distributions of the German rotavirus incidence.

In a second analysis we applied horizontal averaging, i.e. we averaged among the three sets of models sharing the same parameter configuration given through either $\Theta_{(\mu,\omega)}$, $\Theta_{(\mu)}$, or $\Theta_{()}$. Each of these sets contained six models with pairwise different contact patterns.

Table 4.2: Model weights resulting from the model specific marginal likelihoods. The left panel shows weights resulting from all 18 constructed models being subject of averaging, where the right panel shows the vertical weights resulting from averaging among models sharing the same contact structure C_i.

total weights	C_1	C_2	C_3	C_4	C_5	C_6
$\Theta_{(\mu,\omega)}$	0.000	0.000	0.000	0.000	0.000	0.023
$\Theta_{(\mu)}$	0.000	0.000	0.000	0.000	0.000	0.000
$\Theta_{()}$	0.000	0.000	0.000	0.000	0.000	0.977

vertical weights	C_1	C_2	C_3	C_4	C_5	C_6
$\Theta_{(\mu,\omega)}$	0.034	0.000	0.999	0.275	0.888	0.023
$\Theta_{(\mu)}$	0.966	0.000	0.000	0.177	0.111	0.000
$\Theta_{()}$	0.000	1.000	0.001	0.548	0.001	0.977

All three sets of models yielded the same result of contact structure C_6 giving the best model fit. This could also be derived from the model probabilities computed in our first analysis by calculating the probabilities $w(C_j | \Theta_\bullet)$ conditioned on a specific parameter space $\Theta_\bullet \in \left\{ \Theta_{(\mu,\omega)}, \Theta_{(\mu)}, \Theta_{()} \right\}$.

In an analogous analysis we calculated the vertical model weights, i.e. we averaged within each of the six sets of models that share the same contact structure but differ with respect to their parameter configuration $\Theta \in \left\{ \Theta_{(\mu,\omega)}, \Theta_{(\mu)}, \Theta_{()} \right\}$. These vertical weights cannot easily be obtained by looking at the total weights displayed in the left panel of Table 4.2, due to the limited number of digits used. However, the recalculated vertical weights for each contact pattern $C_i, i = 1, \ldots, 6$ are given in the right panel of Table 4.2. Surprisingly, there is generally no clearly best parameter space in $\left\{ \Theta_{(\mu,\omega)}, \Theta_{(\mu)}, \Theta_{()} \right\}$ as the parameter space yielding the highest model probability varies for different underlying contact patterns. This might be due to the fixed prior estimates for the parameters μ and ω being already near their respective posterior mode estimates (when included in the estimation process) for some contact matrices. In these cases the resulting marginal likelihood for the restricted parameter space $\Theta_{(\mu,\omega)}$ is higher compared to that from the more flexible model using the augmented space $\Theta_{()}$.

4.3.1 Incidence predictions

To obtain predictions for the German reported rotavirus incidence for each age group and region we computed a total of $K = 5,000$ samples from the 18 models according to Algorithm 5 using the base case model weights shown in the left panel of Table 4.2 and each model's individual posterior distribution. Using this sample we calculated the expected number of reported cases $\mathbb{E}[X_t^{(j,i)}]$ for each age group $j = 1, \ldots, n_D$, region $i = e, w$ and week $t = 1, \ldots, T$ as well as the 95% prediction interval for the expectation $\mathbb{E}[X_t^{(j,i)}] = h_t^{(i)} \cdot Y_t^{(j,i)}$ with respect to the uncertainty regarding model choice and the respective input parameter. We also computed the equi-tailed 95% prediction intervals for the observations $X_t^{(j,i)} \sim \text{NegBin}(h_t^{(i)} \cdot Y_t^{(j,i)}, d)$ including the additional uncertainty regarding the stochastic observation of events. These results are plotted in Figure 4.3 for the three age groups 0-4, 5-59, and 60-99 years of age for both, EFS and WFS.

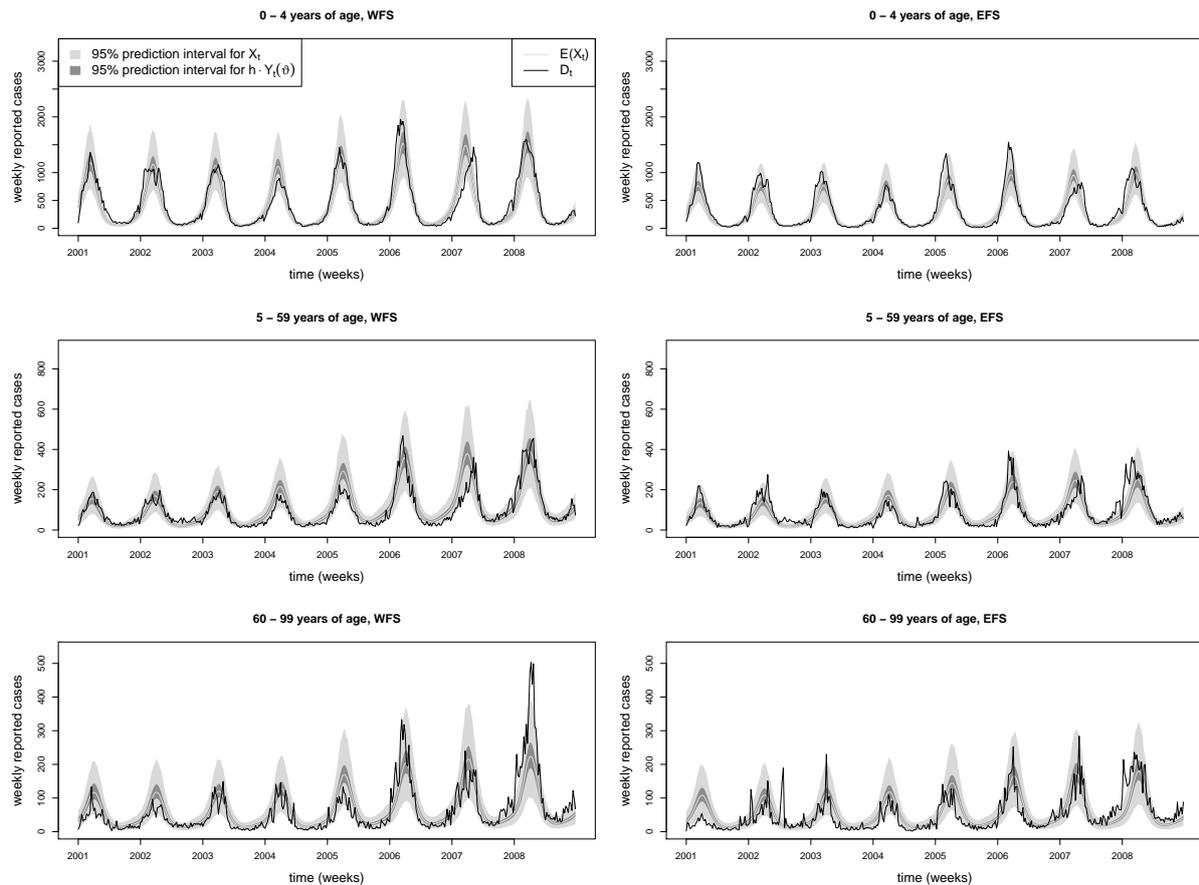We see that the vast majority of observed data points are included within the 95%

Figure 4.3: Weekly number of reported rotavirus cases D_t among the three aggregated age groups of 0-4, 5-59 and 60-99 years of age in the western (left) and eastern (right) federal states. Also shown are the aggregated model averaged predictive distributions of X_t and $\mathbb{E}[X_t]$. Note the different scaling of the y-axis for each age group.

prediction intervals, with no clear sign of over or underestimation by the model predicted expectation in any group. Considering the age group 60-99, which exhibits a higher degree of fluctuation due to a smaller base incidence, the model is still able to catch the average incidence. The predicted seasonality matches the observed rotavirus seasonality very well, with the only exception of the 2007 season. The reasons for the observed peak incidence delay in 2007 are still unknown to the epidemiologists – possibilities are the prevalence of a specific serotype, or special environmental conditions in that season. Although our model was primarily developed to mimic the long term dynamics of the expected German rotavirus incidence, it also captures the point wise observational uncertainty very well for any age group and region, which makes it very suitable for future predictions.

4.3.2 Epidemiological insights
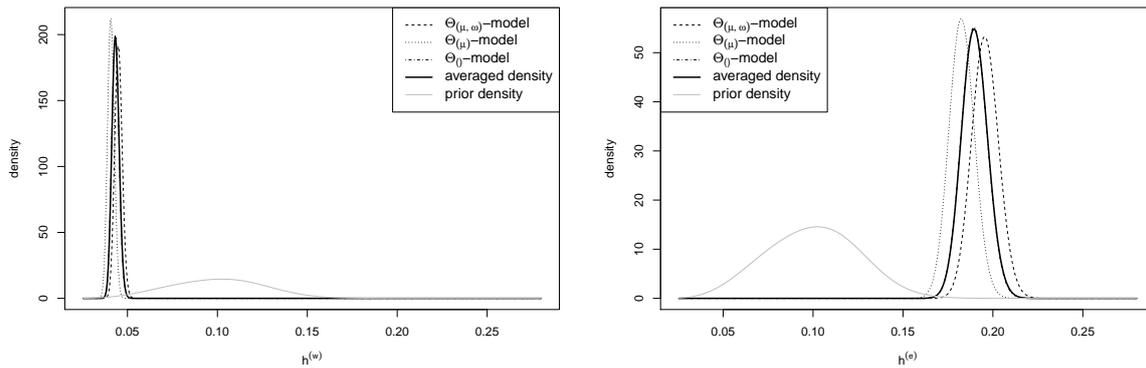
Figure 4.4: Single and vertically averaged posterior distribution of detection rates $h^{(w)}$ (left) and $h^{(e)}$ (right) in the western and eastern federal states, respectively, according to the models using contact pattern $\boldsymbol{C}_6$

Beside the ability to compute a predictive distribution for the future rotavirus incidence, we also obtained posterior distributions for each model parameter. For epidemiological quantities, which directly find a representation as a parameter in the model, we can interpret the corresponding posterior distribution as updated knowledge when combining prior knowledge with the available time series data. Especially for those parameters, that come with a high degree of prior uncertainty due to difficulties in assessing them in studies in real life, the estimates are of particular interest. Examples of such parameters are the detection ratio or the relative infectiousness of symptomatically infected individuals, which includes aspects such as higher excretion of virus but also the lower number of potentially infectious contacts due to bed rest. From the weighted posterior distributions we can compute point estimates by taking, e.g., the posterior median as well as 95% equi-tailed credibility intervals (CI).

Especially the difference in underdetection between the EFS and WFS is of interest, because significant incidence differences between EFS and WFS have been found among other notifiable diseases in Germany as well (Rosner et al., 2010). However, despite the lack of detailed prior information on the reporting behaviour in Germany, we were able to obtain sharp estimates for the parameters $h^{(w)}$ and $h^{(e)}$. For this analysis we considered only the three models using contact structure $\boldsymbol{C}_6$ and computed their marginal posterior distributions for the parameters $h^{(w)}$ and $h^{(e)}$. In Figure 4.4, we see that the posterior median for the detection ratio in the WFS $h^{(w)}$ was computed at 4.5%, (95% CI 4.1-4.9%), 4.1% (3.7-4.5%), and 4.3% (4.0-4.7%) for the model using parameter configuration $\Theta_{(\mu,\omega)}, \Theta_{(\mu)}$ and $\Theta_{()}$, respectively. The averaged posterior density using the weights from the vertical averaging regarding to contact pattern $\boldsymbol{C}_6$ shows a median estimate of 4.3% (3.9-4.7%) for parameter $h^{(w)}$.

Regarding the estimation of the detection ratio in the EFS, the same models computed posterior medians of 19.6% (18.1-21.1%), 18.3% (16.9-19.7%), and 19.0% (17.6-20.4%). The averaged posterior distribution for this parameter suggests a median estimate of 19.0% (17.6-20.5%). These results not only imply that the detection ratio in the EFS is more than 4 times higher compared to the WFS, but also yield concrete estimates for the disease's underdetection which is otherwise difficult to assess. Note, that these ratios describe only the period from 2001 till 2004. Our model allowed a temporal break of these parameters at the end of 2004, such that the detection ratios from 2004 onwards are estimated at 6.3% (5.7-6.9%) and 24.1% (22.3-25.9%) in the WFS and EFS, respectively. This reflects the potential effect due to the change in reimbursement such that laboratory confirmation was requested for more cases after 2004. Note that the relative increase was higher in the WFS, which may result from the higher degree of underdetection in that region beforehand.
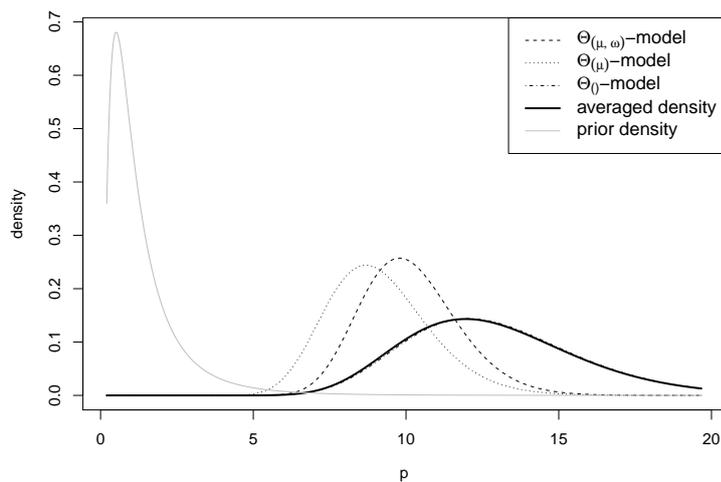

Figure 4.5: Single and vertically averaged posterior distribution for the relative infectiousness of symptomatically infected individuals p according to the models using contact pattern $\boldsymbol{C}_6$.

We conducted the same analysis for the parameter p, which represents the relative infectiousness of symptomatically infected individuals, using again the three models with con-

tact pattern C_6 (Figure 4.5). According to the choice of parameter space from $\Theta_{(\mu,\omega)}, \Theta_{(\mu)}$ and $\Theta_{()}$ we obtained posterior medians of 10.0 (7.4-13.6), 9.0 (6.3-12.9), and 12.7 (8.2-19.6), respectively. Calculating the averaged posterior distribution using the same weights as above, we obtained an averaged posterior median of 12.6 with credibility interval (8.1-19.6). This estimate suggests, that a symptomatically infected person causes more than 10 times as many secondary rotavirus infections as an infected person without symptoms. This difference may be easily explained by the increased excretion of virus and only few viruses being necessary for infection. However, it was not entirely clear whether these aspects would outweigh the dampening effects, such as lesser physical contacts and a potentially higher attention for hygienic counter measures, that might actually decrease the number of secondary cases induced by a symptomatic infection. Therefore, with the goal of taking targeted intervention measures to reduce disease burden, it is important to know which group has the highest impact on the force of infection, where this new insight provides an explicit answer to that problem.

Another interesting aspect of the employed model is the interplay of immunity gain through infection versus immunity loss over time. Considering the high incidence among children and the rising incidence among elderly, it appears that the typical age-specific immunity states within the model change over life course (see Figure 4.2). While children rapidly move through the first two infectivity states I_k and A_k ($k = 1, 2$), once they arrive at S_3 they frequently go through asymptomatic infection A_3 maintaining their high immunity level. With high age the contact rate decreases and so does the force of infection such that a loss of immunity ($S_3 \to S_2$) becomes more likely. Therefore, according to the model, infections among the elderly happen less frequently than in the middle age group but with a higher chance of proceeding symptomatically which explains the higher reported incidences in the elderly.

Posteriors of further model parameters

Posterior distributions from vertical averaging using contact pattern C_6 for all other parameters to be estimated in each model are shown in Figures 4.6 and 4.7. Note the differences in the estimated contact parameters c_1, c_2, and c_3 defining the transmission rates among young children (0-4 years), among adults (5-59 years), and among elderly and between the groups, respectively (see contactmatrix C_6 in Equation (4.8)). The transmission rates among children and adults (c_1, c_2) are estimated to be considerably higher than among elderly (c_3). Although these rates also contain information on age-specific susceptibility and infectiousness such large relative differences seem implausible. However, a more strict prior for these parameters might be able to avert this issue.

The averaged posterior distribution for the partially fixed parameters μ and ω are given in Figure 4.7. It can be observed that the averaged cdfs of these parameters have a jump in those spots in which the parameter were fixed according to the parameter spaces $\Theta_{(\mu,\omega)}$ and $\Theta_{(\mu)}$. Thereby, the jump size is equal to the posterior model probabilities $w(C_6, \Theta_{(\mu,\omega)})$ and $w(C_6, \Theta_{(\mu)})$ of the corresponding models based on these restricted parameter spaces. How-
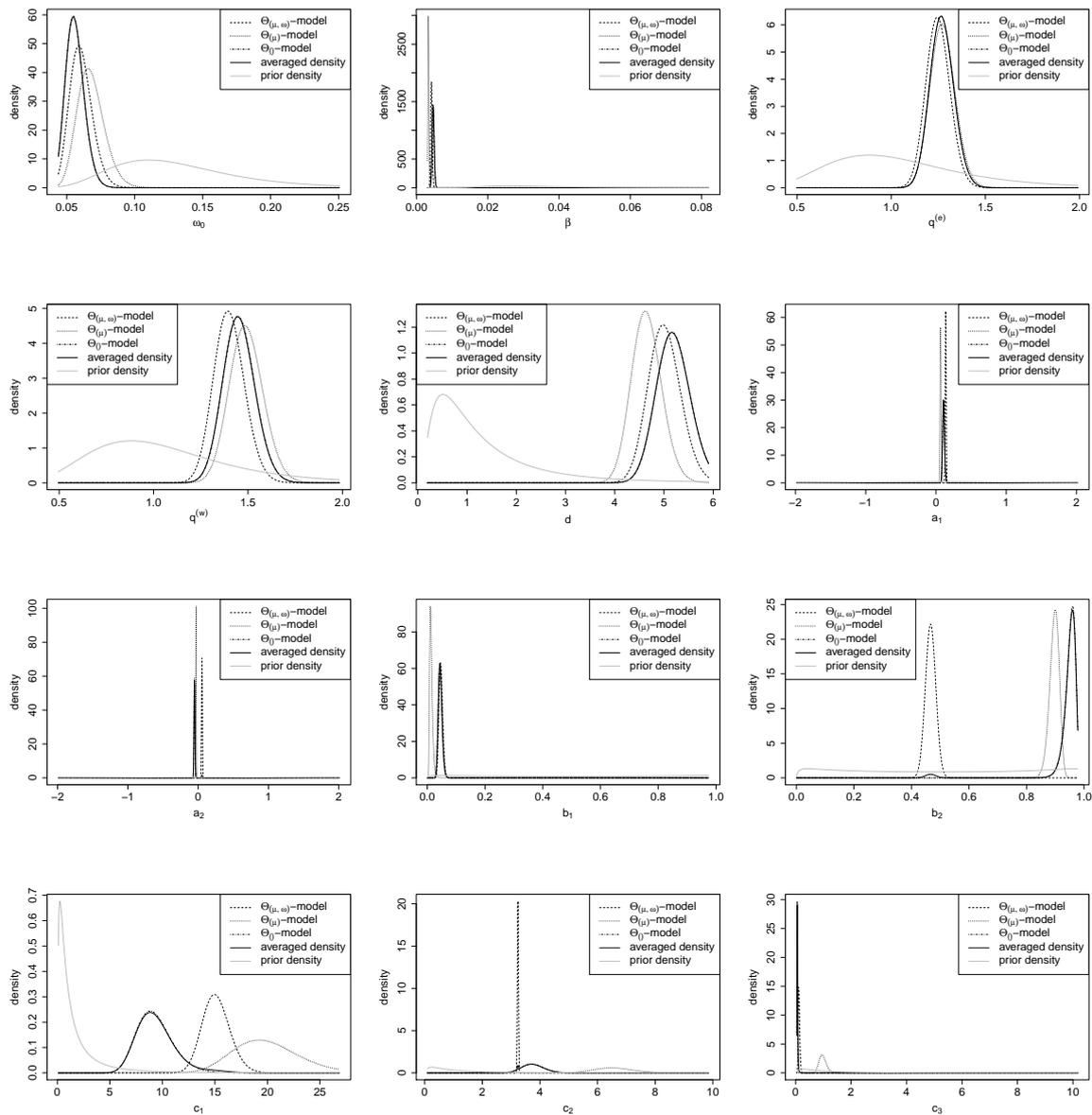
Figure 4.6: Averaged posterior densities of all parameters to be estimated according to the models using contact pattern C_6.

ever, note that these averaged cumulative distribution functions are primarily displayed to illustrate the methodological possibilities as the main purpose of our analysis was to investigate the effects of fixing parameters, not to obtain an averaged posterior among models with partially fixed parameters. One could have obtained the same posterior considering one single model with a mixed prior for the parameters (μ, ω) consisting of degenerated distributions and the respective skew-normal distributions defined above – each weighted equally.
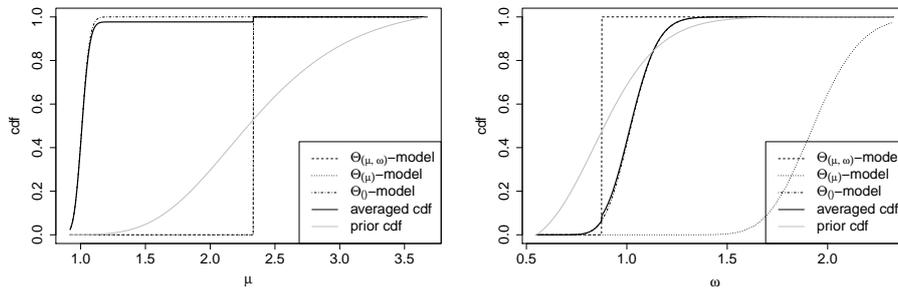
Figure 4.7: Averaged posterior cumulative distribution function for the parameters μ and ω according to the models using contact pattern $\boldsymbol{C}_6$.

4.3.3 Comparison to different transmission models

We also re-implemented the transmission models presented in (Pitzer et al., 2009; Atchison et al., 2010) and applied them to the German rotavirus data. Both model structures were taken as described by the ODE systems in the respective works for the United States (Pitzer et al., 2009) and England and Wales (Atchison et al., 2010) and supplemented with adopted versions of our observational component and our definition of the force of infection including the proposed contact pattern $\boldsymbol{C}_6$. Using the same inference approach, both models yielded significantly lower marginal likelihoods. These reference models had in particular problems to mimic the incidences in the age groups 60+, with especially the model from (Pitzer et al., 2009) heavily underestimating the corresponding number of reported cases, presumably due to a missing component regarding potential immunity loss.

4.3.4 Statistical insights

The main intention of our Bayesian approach was to account for uncertainty originating from both parameter quantification and model selection. Since the model-specific posterior distributions were calculated separately for each considered model, we were able to examine the impact of different model assumptions on the posterior of the parameter vector $\boldsymbol{\vartheta}$. This

relation is of particular interest considering the consequences these estimates might have on health policy making or subsequent epidemiological studies.

Parameter spaces and collinearities

Looking at the median estimates for the relative infectiousness of symptomatically infected individuals p as shown in Figure 4.5, it becomes clear that certain parameter estimates react very sensitively to changes in model structure and parameter space. Of particular interest is the impact of fixing parameter values in advance as often done in the literature (Pitzer et al., 2009; Atchison et al., 2010; de Blasio et al., 2010). In our case, the models using configuration $\left(\boldsymbol{C}_6, \Theta_{(\mu,\omega)}\right)$ and $\left(\boldsymbol{C}_6, \Theta_{(\mu)}\right)$, where one or both of the parameters $\mu = 7/3$ and $\omega = 7/8$ were fixed, the posterior median of variable p was 10.0 and 9.0, respectively. In contrast the model using the full space $\left(\boldsymbol{C}_6, Z_{()}\right)$ yielded a median estimate for p at 12.6.

Adding only ω to the parameter vector ϑ did yield a slight decrease in the estimated value for p, while the posterior median for ω itself at 1.93 was higher than its corresponding fixed value. Adding also μ, we obtained notably different estimates. In the corresponding model $\left(\boldsymbol{C}_6, \Theta_{()}\right)$, the posterior medians for ω and μ were given at 1.01 and 1.02, significantly lower than the estimates from model $\left(\boldsymbol{C}_6, \Theta_{(\mu)}\right)$, while the median for p at 12.6 increased. With a closer look at the corresponding posterior correlations, it seems that the model prefers parameter vectors yielding the same μ/ω ratio and compensates variations in these mean durations of infection by adjusting the contact rates c_1, c_2, c_3 and p. In order to prevent large variations of some parameter's estimates when adding other correlated parameters to the vector which is subject of estimation, one approach might be to define stricter priors instead of fixing a parameter entirely. In that regard, we have already seen that the prior distribution for the contact parameters appears to be too vague, especially in the light of the large amounts of data which dominate the posterior through the likelihood.

The above insights are also useful when thinking about potential extensions of the model structure. One modification could be the introduction of an infectious period preceding the symptomatic phase, which was left out in the original model because epidemiological studies (Anderson and Weber, 2004; Pickering et al., 1988) have shown this period to be rather short. However, while the estimated durations of symptomatic and asymptomatic phase might change under such a model variation, the relative infectiousness p is likely to be robust. This can be seen from the duration parameters μ and ω varying significantly among the compared models, while the estimate for parameter p remains relatively stable, since it is primarily affected by the mean infectiousness of a childhood case (many symptomatic infections) versus an adult case (almost no symptomatic infections), but only partially by the duration of symptomatic and asymptomatic period.

Two step optimization

Also of interest is the impact of the second step optimization on the posterior, in which we adjusted for the estimated cumulative autocorrelation in the data as given in equation (4.4). Using the modified likelihood LL_{CA} had significant effects on the final estimation results, as we can see by, e.g., analysing the posterior estimation for parameter $h^{(e)}$.

As shown in Figure 4.4 in model $\left(\boldsymbol{C}_6, \Theta_{(\mu,\omega)}\right)$ the second step posterior median for $h^{(e)}$ was 19.0% with 95% CI (17.6-20.5%) whereas after the first optimization step using only the posterior density as defined in equation (4.3) the median was estimated at 19.6% (19.3-19.9%). By reducing the impact of the likelihood by a factor determined by the cumulative autocorrelation the posterior density flattened around the posterior mode. As a consequence the posterior median estimates moved closer to the prior medians, but more importantly we obtained a larger posterior variance and hence also wider credibility regions.

Furthermore, by downscaling the likelihood the marginal likelihoods of each of the selected models, as computed via equation (4.9), moved closer together. This can also be observed in the case of vertical averaging with respect to contact structure $\boldsymbol{C}_5$, where the largest difference of the marginal likelihoods was 7.2 on the log scale after the second step procedure while the minimal difference after only one estimation step was 25.4, which would lead to nearly degenerated model weights.

Normal approximation of the posterior

The accuracy of the normal approximation of the posterior distribution was already investigated in Section 2.3.4, although for a much simpler epidemic model. Here, we aim to check whether the normal approximation holds also for the higher dimensional setting of the rotavirus transmission model.

As it is not feasible to compare the whole 17-dimensional posterior against its approximation, we instead checked the match of the respective conditional log posteriors with respect to one or two parameter components, conditioned on all other parameter components being equal to their posterior mean. The conditional log posterior density of the parameter component ϑ_i is given by

$$\log \pi_{\text{CA}}\left(\vartheta_i \,\middle|\, \boldsymbol{\vartheta}_{-i} = \boldsymbol{\vartheta}^*_{-i}, \boldsymbol{D}\right) = \log \pi_{\text{CA}}\left(\vartheta_i, \boldsymbol{\vartheta}^*_{-i} \,\middle|\, \boldsymbol{D}\right) - \log \pi_{\text{CA}}\left(\boldsymbol{\vartheta}^* \,\middle|\, \boldsymbol{D}\right),$$

where $\boldsymbol{\vartheta}_{-i}$ denotes the vector of components in $\boldsymbol{\vartheta}$ which are not ϑ_i, and $\boldsymbol{\vartheta}^*$ denotes the second step posterior mode $\boldsymbol{\vartheta}^*_{\text{CA}}$. Hereby, the term $-\log \pi_{\text{CA}}\left(\boldsymbol{\vartheta}^* \,\middle|\, \boldsymbol{D}\right)$ functions as an additive constant such that the conditional log posterior's maximum is at zero (this does not imply that $\log \pi_{\text{CA}}\left(\boldsymbol{\vartheta}^* \,\middle|\, \boldsymbol{D}\right)$ is the normalizing constant of the log posterior). Analogously the conditional log posterior according to the normal approximation as defined in (4.5) is given by

$$\log \pi_{\mathcal{N}}\left(\vartheta_i \,\middle|\, \boldsymbol{\vartheta}_{-i} = \boldsymbol{\vartheta}^*_{-i}, \boldsymbol{D}\right) = \log \phi_{\boldsymbol{\vartheta}^*, \boldsymbol{\Sigma}}\left(\vartheta_i, \boldsymbol{\vartheta}^*_{-i}\right) - \log \phi_{\boldsymbol{\vartheta}^*, \boldsymbol{\Sigma}}\left(\boldsymbol{\vartheta}^*\right)$$
$$= -\frac{1}{2}\left(\left(\vartheta_i, \boldsymbol{\vartheta}^*_{-i}\right) - \boldsymbol{\vartheta}^*\right)^T \boldsymbol{\Sigma}^{-1}\left(\left(\vartheta_i, \boldsymbol{\vartheta}^*_{-i}\right) - \boldsymbol{\vartheta}^*\right),$$

where $\boldsymbol{\Sigma}$ is the approximated posterior covariance matrix based on the Fisher information (see Equation (4.5)) and $(\vartheta_i, \boldsymbol{\vartheta}^*_{-i})$ denotes $\boldsymbol{\vartheta}^*$ with its i-th component being replaced by ϑ_i. Thus, the conditional log densities each represent a slice of the corresponding unconditioned log posterior density. To check whether the two conditional densities are approximately equal, we evaluated them over a range determined by the 95% credibility region (CrR) of the unconditioned posterior. Based on the approximated posterior the random variable

$$\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*\right)$$

is χ^2_d distributed with 17 degrees of freedom. Thus, the 95% CrR is given by

$$\mathrm{CrR}_{95\%} = \left\{ \boldsymbol{\vartheta} \in \Theta \,\middle|\, \log \phi_{\boldsymbol{\vartheta}^*, \boldsymbol{\Sigma}}\left(\boldsymbol{\vartheta}\right) - \log \phi_{\boldsymbol{\vartheta}^*, \boldsymbol{\Sigma}}\left(\boldsymbol{\vartheta}^*\right) > -\frac{q_{0.95}(\chi^2_{17})}{2} \approx -13.79 \right\}.$$

We calculated the conditional posteriors for the parameters μ and ω according to model $\left(\boldsymbol{C}_6, \Theta_{()}\right)$. Note that the densities were computed on the respective log transformed space. The conditional log densities according to the true posterior and its normal approximation together with the 95% CrR thresholds are given in Figure 4.8 (top row). Additionally, we computed the two-dimensional conditional log posterior for the parameter vector (μ, ω) and plotted the contour lines where the true posterior and the normal approximation cross the thresholds corresponding to the 5%-, 25%-, 50%-, 75%-, and 95% CrRs (bottom figure).

One can see that the normal approximation matches the actual log posterior density quite well such that all qualitative statements about the parameter estimates, e.g. those from Section 4.3.2, remain true. However, the match is certainly not perfect as the parameter values at which the densities cross the critical 95% CrR threshold slightly deviate. Moreover, the approximation is well capable to capture the parameter correlations as displayed by the 2-dimensional distribution in Figure 4.8. It should be noted that the relative deviation of posterior approximation increases with growing distance to the posterior mode. This can be easily explained since the normal approximation is based on a second order Taylor approximation (Abramowitz and Stegun, 1964) of the log posterior around its mode and therefore, becomes increasingly inaccurate within distant regions.

Convergence of the optimization procedure

One of the main reasons for computing the normal approximation of the posterior was to save many costly model evaluation which would be otherwise required if, e.g., the posterior had been approximated by an MCMC-sample (see Section 3.2). However, computing the posterior mode using the Nelder-Mead algorithm as described in Section 4.2 already required up to 100,000 posterior evaluations for some considered models, which yielded a computation time of a few days for one model. Moreover, it is not possible to shorten the procedure at the cost of approximation accuracy as the approximation approach requires the Hessian at the computed mode to be negative definite, which is not guaranteed if the optimization is stopped too early.
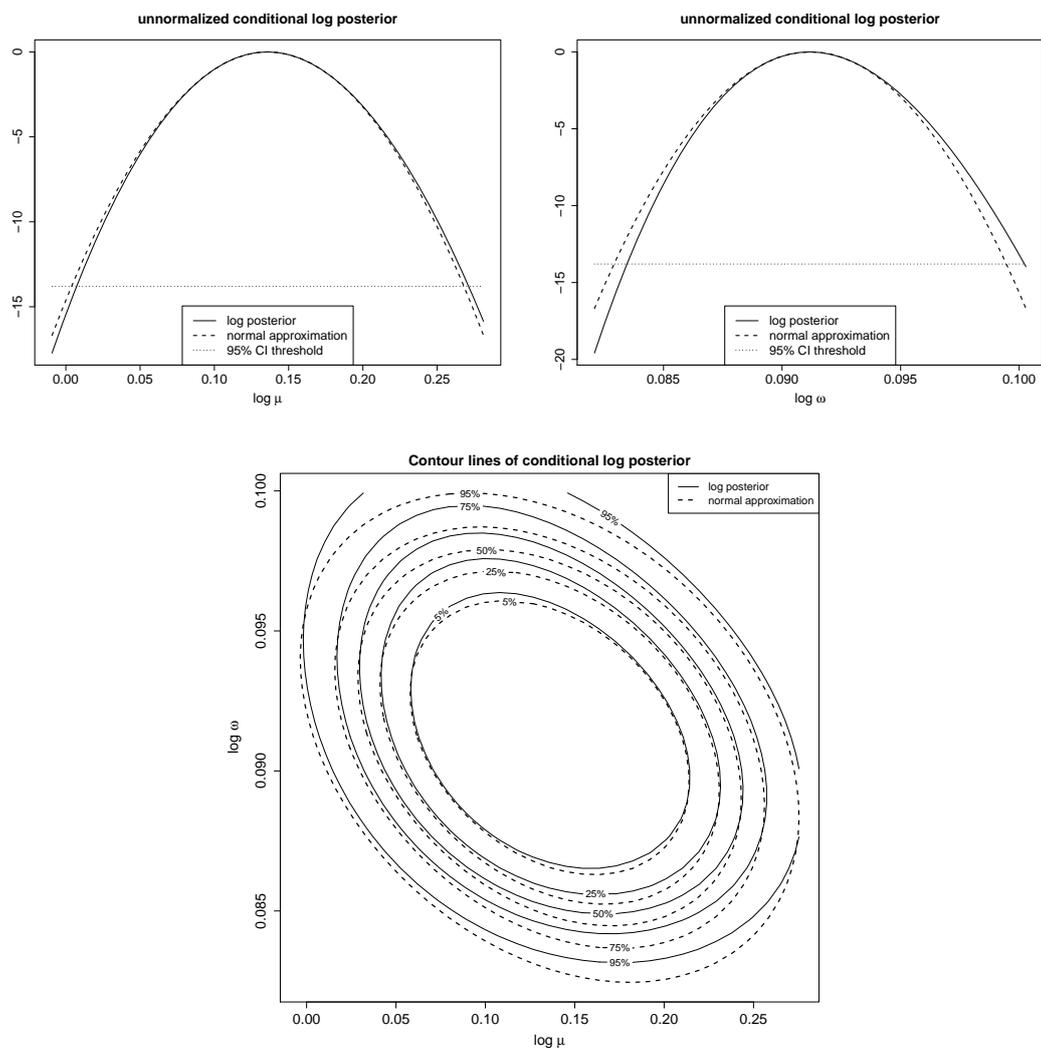
Figure 4.8: Comparison of the conditional log posterior density and its corresponding normal approximation for the single parameters μ, ω and the joint parameter vector. Also displayed are the approximate 95% credibility region borders corresponding to the full posterior distribution.

4.4 Discussion

In this chapter, we presented a realistic model for the transmission and reporting of rotavirus in Germany representing all epidemiologically relevant aspects. Such a detailed mechanistic model is a necessary precondition for investigating potential intervention measures such as a nationwide routine vaccination. Although our model building was heavily focused on describing the rotavirus dynamics in Germany, one can easily point out those model components which can be adapted to other pathogens sharing similar transmission characteristics.

We used a Bayesian framework to put focus on the analysis of uncertainty regarding model parameters and model selection. To address the strong remaining autocorrelation in the residuals of our ODE-based modelling, we applied a pragmatic and robust correction method by rescaling the likelihood function. The Bayesian approach allowed us to gain knowledge on the underlying epidemiology of disease transmission as was done similarly in (Birrell et al., 2011; Dorigatti et al., 2012). Among epidemiological insights, we obtained estimates for the region-specific detection rates in Germany, but also found the force of infection to be primarily driven by symptomatically as opposed to asymptomatically infected individuals.

As the underlying transmission process is exceedingly complex, it is not feasible to construct a mathematical model embodying the "true dynamics". Hence, the observed residual autocorrelation is an indicator of this imperfection. Thus, acknowledging this model dependence is important when interpreting the corresponding inference results. To lessen the reliance on certain model assumptions, besides accounting for the inherent residual autocorrelation, we additionally applied methods to integrate several model structures into the estimation process, using Bayesian averaging tools to generate synthesised results. Altogether, our approach is novel in the field of disease transmission modelling based on ODEs as uncertainty regarding both, model structure and parameter, is treated in a more comprehensive way and we are able to incorporate this into the future analysis of intervention strategies.

Considering the strong dependencies of the parameter estimates with respect to the underlying model structure, but also the multivariate correlation among the input parameters, we found that one should be very careful with fixing crucial input parameters as it is sometimes done in other studies. Since the fixing of certain model parameters affects the inference of other input parameter, the resulting transmission dynamics governed by the final model calibration might be biased. Due to an erroneous predictive behaviour of the model, this might have serious consequences for public health decisions depending on such modelling results. It is important to note, that such effects cannot be revealed by conducting univariate sensitivity analysis on parameters in question, as it is often done to quantify predictive uncertainty in the context of infectious disease modelling (Bilcke et al., 2011).

We also applied adapted versions of the models which were developed previously (Pitzer

et al., 2009; Atchison et al., 2010) to the German rotavirus incidence data. Based on the marginal likelihood our model yielded a significantly better fit, especially regarding the elder age groups. However, those reference models were aimed at the rotavirus epidemiology in the United States and England and Wales and also primarily at the target group of children younger than 5 years of age, which requires a different focus in model building. This makes a comparison of the models difficult but again shows how dependent such models are on the underlying data and processes used for developing them.

Chapter 5

Modelling the epidemiological impact of rotavirus vaccination in Germany

> The content of this chapter is largely based on the article and its supplementary material published in Weidemann et al. (2014b).

Two rotavirus vaccines, RotaTeq® (Merck & Co) and Rotarix® (GlaxoSmithKline), were licensed for use in Europe in 2006. In July 2013, the German Standing Committee on Vaccination (STIKO) decided to adopt rotavirus vaccination into the national vaccination schedule for children (Koch et al., 2013). In Europe, routine rotavirus vaccination has also been introduced in Austria, Belgium, the United Kingdom, Finland, and Luxembourg as of today (Vesikari, 2008).

Predicting the epidemiological impact of a vaccination program constitutes a challenging task due to the complex transmission processes driving the disease spread within a given population. Mathematical modelling of these transmission dynamics while also accounting for vaccination processes provides an analytical tool to assess the potential effects of routine vaccination. However, a careful treatment of uncertainty arising from the choice of model and its respective parameters is necessary when adequately predicting the resulting disease incidence.

The aim of this chapter is to extend our model proposed in Chapter 4 by introducing vaccination mechanisms into the transmission dynamics. The augmented model is then used to i) estimate the rotavirus vaccine effectiveness (VE) based on notification data, and ii) to predict the epidemiological impact of rotavirus vaccination in Germany.

Before adoption into the national vaccination schedule, rotavirus vaccination was already recommended in some German federal states but not reimbursed by all insurance companies (Dudareva et al., 2012). Still, especially in the five eastern federal states (EFS) both rotavirus vaccines have been widely used since 2008 with approximately equal share (Dudareva et al., 2012). As a consequence, a significant rotavirus incidence decrease in vaccinated age-groups was already observed in the communicable disease reporting system

when comparing seasons before and after rotavirus vaccine introduction (Dudareva et al., 2012). By again applying a Bayesian inference framework the notification data enable the estimation of the direct rotavirus vaccine effectiveness via a transmission model. Such a model-induced VE estimate can yield additional information for later incidence predictions, since efficacy estimates from clinical trials alone do not necessarily apply under field conditions.

In the subsequent step the estimated parameter distributions are utilized to sample the predictive distribution of the rotavirus incidence following introduction of routine vaccination, with a special focus on addressing uncertainty arising from the stochastic modelling, parameter estimation, and demographic development. Such an assessment of epidemiological impact is necessary to support informed policy-making regarding the recommendation of new vaccinations. Considering the introduction of rotavirus vaccination, dynamic epidemic models for rotavirus transmission and vaccination impact were also developed for the United States (Pitzer et al., 2009), England and Wales (Atchison et al., 2010; Atkins et al., 2012), Mexico (Shim and Castillo-Chavez, 2009), and Kyrgyzstan (de Blasio et al., 2010) – however, with less focus on prediction uncertainty. In Germany the cost-effectiveness of rotavirus vaccination was so far assessed only based on a static cohort-model (Aidelsburger et al., 2014). Thus, with our modelling approach we aim to provide further evidence regarding the recommendation of rotavirus vaccination in Germany.

This chapter is structured as follows. In Section 5.1 we propose a set of extended transmission models accounting for vaccination mechanisms and further aspects, which were not considered previously. The Bayesian inference framework and corresponding results for estimating the rotavirus vaccine effectiveness among other model parameters and computing the marginal likelihood of each proposed model will be presented in Section 5.2. The predictive sampling procedure together with measures for the indirect effects of vaccination are treated in Section 5.3. A discussion on the epidemiological results and statistical methods is given in Section 5.4

5.1 The rotavirus transmission model

For assessing the rotavirus vaccine effectiveness and epidemiological impact we developed an age-stratified dynamic transmission model featuring the known key aspects of rotavirus epidemiology, which is governed by a system of ordinary differential equations. The model is an extension of our model not accounting for vaccination which was described in detail in Chapter 4.1. A structural overview on the newly proposed model accounting for vaccination is given in Figure 5.1.

The main transmission dynamics which were already described in Section 4.1 can be briefly summarized as follows. Infants are born into the state M in which they are immune to infection due to maternal antibodies. With vanishing protection children move to the first of three susceptibility states S_1 from where they can become infected, moving either to state I_1 (symptomatic) or A_1 (asymptomatic), respectively. After recovery individuals

Figure 5.1: Structural overview of a single age-layer of the augmented transmission model. Compartments above the dotted line represent the original transmission model without vaccination developed in Section 4.1, whereas the lower part is the new vaccination extension. Not shown, although present, are the age-layer structure and the intermediate susceptibility compartments S_{iA} and S_{iB} as seen in Figure 4.2.

move into the second susceptibility state S_2. With each increasing state of susceptibility the risks for infection and developing symptoms decrease due to a better immunity induced by previous infections. In the last susceptibility state S_3 there remains only the possibility for developing an asymptomatic infection. To model waning immunity in the absence of further infections, individuals move from high to lower susceptibility states at a constant rate.

To consider an age-specific transmission the population is decomposed into $n_A = 19$ age groups, with a finer decomposition among the lower years of age. We defined 6 two-month-wide groups for the first year of age, 8 six-months-wide groups up to four years of age and 5 groups for the age from 5-19, 20-39, 40-59, 60-79, and 80+. This high age resolution was necessary to model the age of vaccination after 2 and 4 months as determined by the vaccine license. Thus, each of the states described in Figure 5.1 is decomposed into $n_A = 19$ age-specific states (e.g. $S_1^{(j)}, j = 1, \ldots, n_A$) where the j-th component corresponds to the j-th age groups.

5.1.1 Vaccination modelling

In the present context we introduced four new age-specific states $V_1^{(j)}, \ldots, V_4^{(j)}$ capturing vaccinated individuals in the model. The reason for introducing more than one vaccine-related group is that waning vaccine-induced protection has been observed (Soares-Weiser et al., 2012) which requires multiple groups to mimic the varying level of immunity. Each of the four vaccine states is associated with specific relative risks $\mathrm{RR}_I^{(k)}$ and $\mathrm{RR}_S^{(k)}$, $k = 1, \ldots, 4$, for becoming infected and developing symptoms, respectively, compared to the first natural infection $S_1 \to I_1$. These relative risks are determined by the parameters η_I and η_S. We assumed an exponential waning of the vaccine-induced immunity, both against infection and developing symptoms, with the highest protection among individuals in state V_4. Thus, we defined the relative risks for state V_k as follows:

$$\mathrm{RR}_I^{(k)} = \eta_I^{\frac{k}{4}}, \quad \mathrm{RR}_S^{(k)} = \eta_S^{\frac{k}{4}}, \quad k = 1, \ldots, 4.$$

Literature evidence suggests a high immunity gained by vaccination compared to the immunity after one natural infection (Soares-Weiser et al., 2012). Hence, we assumed that following recovery from a breakthrough infection (received within one of the vaccine states $V_1, \ldots, V_4$) children move to state S_3 whereas moving to the second susceptibility group S_2 could otherwise result in less immunity compared to prior the infection.

We modelled the vaccine to be administered in two doses, with the first and second dose given at the end of the second and fourth month of life, respectively. Thus, the compartement change representing vaccination is functionally different to all other movements in the model, since it is directly linked to the ageing process such that it happens at a fixed time point in life. This mechanism is presented in Figure 5.2. Among all children eligible for vaccination, a proportion $0 \leq \phi \leq 1$ equal to the momentary vaccination coverage rate receives the first vaccine dose at age 2 months, whereas all infants who received one dose also receive the second dose, unless an infection occurred between the two doses. Since protection after partial vaccination was found to be already high (Martinón-Torres et al., 2011), we assumed children move to state V_3 after the first dose administration. With the second dose children then move to state V_4, where they are benefiting from the highest possible protection provided by the vaccine.

The vaccination coverage in the model is parametrized by yearly coverage rates ϕ_i ($i = 2006, \ldots, 2013$). To derive values for the weekly coverage rates $\phi_t^{(w)}$ ($t = 0, \ldots, T$) we applied cubic spline-interpolation, using the *R*-function *spline*, assuming that the year-specific coverage rate ϕ_i is achieved at the end of the corresponding year and vaccination coverage is zero at the beginning of 2006. To perform the interpolation procedure we transformed the crude coverage rates $\phi_i \in [0, 1]$ onto a logit-scale, i.e. $\tilde{\phi}_i = \mathrm{logit}\,\phi_i$. These were used to set the weekly coverage rates at the end of each year via $\tilde{\phi}_{(i-2005)\cdot 52}^{(w)} = \tilde{\phi}_i$ for $i \in \{2006, \ldots, 2013\}$. For the sake of well-definedness we set the transformed coverage rate in the beginning of 2006 by $\tilde{\phi}_0^{(w)} = \mathrm{logit}\,0.001$ (i.e. the coverage is almost zero). The intermediate values $\tilde{\phi}_t^{(w)}$ ($t \in \{0, \ldots, 52 \cdot (2013 - 2005)\}$) are then obtained by cubic
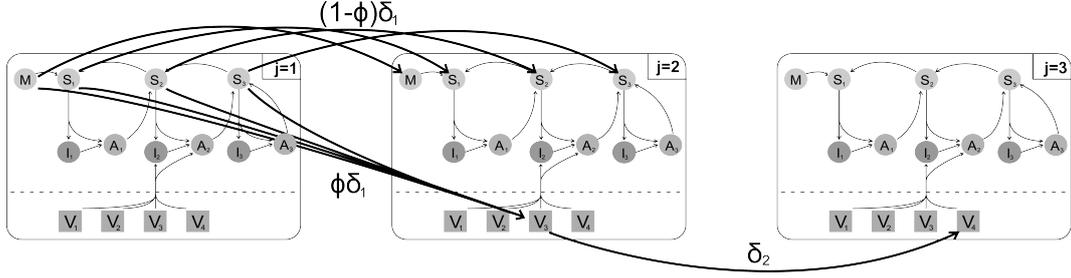
Figure 5.2: Overview of the vaccination process within the first three age groups representing the first 6 months of life. When moving from age group 1-2 months to 3-4 months a proportion $0 \leq \phi \leq 1$, i.e. the coverage, of all infants from the states M and $S_1 - S_3$ change the state to V_3 instead of jumping to the corresponding state in the higher age group, thus, receiving first dose vaccination. For the second vaccine dose, all infants from the states $V_1 - V_3$, i.e. who received first dose vaccination, change to V_4 while moving from age group 3-4 months to 5-6 months.

spline interpolation as described by Forsythe et al. (1977, Ch. 4.4). The final weekly momentaneous coverage rates are then computed by re-transforming

$$\phi_t^{(w)} = \text{expit}\, \tilde{\phi}_t^{(w)} = \text{logit}^{-1} \tilde{\phi}_t^{(w)}.$$

Waning immunity

In the clinical trials the rotavirus vaccine induced immunity was found to decline after already one year (Soares-Weiser et al., 2012). However, no specific evidence on the actual speed of immunity waning exists. Thus, we examined two different scenarios for the waning process.

In the first scenario immunity waning was modelled in yearly steps, such that all children move back from V_4 to V_3 at the age of 18 months. This proceeds with moving to states V_2 and V_1 after 2.5 and 3.5 years of life, respectively. Data on long term immunity after several years is not available from clinical studies yet. Thus, we assumed that after 4.5 years of life the vaccine induced immunity is completely vanished such that children move back to the first susceptibility state S_1 again. Within this scenario loss of immunity is coupled with the ageing process, similar to the modelling of vaccine administration as displayed in Figure 5.2. In doing so, a yearly loss of one immunity level is assured.

In a second model scenario, we assumed that waning occurs at a constant rate $\eta_W > 0$, at which individuals move from one vaccine state V_k to the prior one V_{k-1}. This also adds η_W as a third component to the parameter vector $\boldsymbol{\eta} = (\eta_I, \eta_S, \eta_W)$ governing the effectiveness related parameters. The alternative scenario allows for waning rates suggesting a faster or slower loss of immunity. Moreover, since only the mean waning rate is modelled, the waning process is more heterogeneous such that some children lose the vaccine-induced protection faster than others.

5.1.2   Age-related immunity

To explain the increasing rotavirus incidence around the age of 60 years, there are two competing approaches. In the first one, which was implemented in the model from Chapter 4.1, the dynamics allow for loss of natural immunity over time in absence of contact to the virus. Thus, people may become resusceptible to symptomatic infection, e.g. due to fewer virus contacts in higher age.

An alternative possibility is to implement an age-specific relative risk for developing symptoms directly into the model, which is reasonable since the natural immunity not accounting for potential previous infections, changes with age. Of course, a combination of the two approaches is also possible. Both approaches may be able two explain the rotavirus pre-vaccination incidence (2001-2008) equally well – however, they might suggest different results when introducing a vaccination program.

To account for both possibilities, our model was modified with an age-specific factor $\Psi(a)$ multiplied on the probability of developing symptoms in each infection stage. This factor $\Psi(a)$ was modelled as θ_1^{-1} times the expit (i.e. inverse logit) of a quadratic polynomial with respect to age a, parametrized by the additional parameters ψ_i, $i = 1, 2, 3$, i.e.

$$\Psi(a) = \frac{1}{\theta_1} \operatorname{expit}\left(\operatorname{logit}(\theta_1) + \psi_1 + \psi_2 a + \psi_3 a^2\right) \tag{5.1}$$

This definition is inspired by logistic regression analysis (Hosmer et al., 2013) as the parameters ψ_i correspond to regression coefficients measuring the impact of age on the relative risk for developing symptoms. Thereby, age is allowed to have a non-monotonic effect due to the second order polynomial in the regression term. In the default setting $\psi_i = 0$ for $i = 1, 2, 3$, $\Psi(a)$ equals $\theta_1 \cdot \theta_1^{-1} = 1$, i.e. no age-specific differences. Otherwise, $\Psi(a)$ takes values within the range $[0, \theta_1^{-1}]$ when modifying the parameters ψ_i such that the overall chances for developing symptoms in each stage $\theta_j \cdot \Psi(a)$ ($j = 1, 2, 3$) are well defined.

5.1.3   Full model-equations

The previous descriptions can be translated into the following system of differential equations given by (5.2) and (5.3) representing the unobserved transmission dynamics. Here, the given ODE system corresponds to the model scenario with variable vaccine immunity waning η_W. Note again, that the upper index (j), $j = 1, \ldots, n_A$, denotes the age group and that $\mathbb{I}_{\{A\}}(x)$ refers to the indicator function. $N^{(j)}$ and a_j denote the total size and the lower age bound of age group j, respectively.

This first part of the ODE system given by (5.2) governs the non vaccine compartments as seen in the upper half of Figure 5.1. The graphical model overview in Figure 5.1 does not contain the intermediate susceptibility groups $S_{2A}, S_{2B}, S_{3A}, S_{3B}$ which induce a time to loss of natural immunity, that corresponds to a $\Gamma(3, \beta)$-distribution. The variables κ, $\delta^{(j)}$, $\gamma^{(j)}$ and $m^{(j)}$ represent the time-specific birth, ageing, death and migration rates, respectively. See Table 5.2 on page 144 for an explanation of all model parameters. The second part of the system (5.3) (page 134)covers the vaccine groups $V_1, \ldots, V_4$.

$$\frac{dM^{(j)}}{dt} = \mathbb{I}_{\{1\}}(j)\kappa_t + \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}M^{(j-1)} - \left(\omega_0 + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)M^{(j)}$$

$$\frac{dS_1^{(j)}}{dt} = \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}S_1^{(j-1)} + \omega_0 M^{(j)} + \eta_W V_1^{(j)} + \beta S_{2B}^{(j)} - \left(\lambda^{(j)}(t) + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_1^{(j)}$$

$$\frac{dI_1^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}I_1^{(j-1)} + \theta_1\Psi(a_j)\lambda^{(j)}(t)S_1^{(j)} - \left(\mu + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)I_1^{(j)}$$

$$\frac{dA_1^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}A_1^{(j-1)} + (1-\theta_1\Psi(a_j))\lambda^{(j)}(t)S_1^{(j)} - \left(\omega + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)A_1^{(j)}$$

$$\frac{dS_2^{(j)}}{dt} = \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}S_2^{(j-1)} + \omega A_1^{(j)} + \beta S_{3B}^{(j)} - \left(\alpha_2\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_2^{(j)}$$

$$\frac{dS_{2A}^{(j)}}{dt} = \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}S_{2A}^{(j-1)} + \beta S_2^{(j)} - \left(\alpha_2\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{2A}^{(j)}$$

$$\frac{dS_{2B}^{(j)}}{dt} = \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}S_{2B}^{(j-1)} + \beta S_{2A}^{(j)} - \left(\alpha_2\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{2B}^{(j)}$$

$$\frac{dI_2^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}I_2^{(j-1)} + \theta_2\Psi(a_j)\alpha_2\lambda^{(j)}(t)\left(S_2^{(j)} + S_{2A}^{(j)} + S_{2B}^{(j)}\right) - \left(\mu + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)I_2^{(j)}$$
$$+ \sum_{k=1}^{4}\left(\theta_1 RR_S^{(k)}\Psi(a_j)\right)RR_I^{(k)}\lambda^{(j)}(t)V_k^{(j)}$$

$$\frac{dA_2^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}A_2^{(j-1)} + (1-\theta_2\Psi(a_j))\alpha_2\lambda^{(j)}(t)\left(S_2^{(j)} + S_{2A}^{(j)} + S_{2B}^{(j)}\right) - \left(\omega + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)A_2^{(j)}$$
$$+ \sum_{k=1}^{4}\left(1-\theta_1 RR_S^{(k)}\Psi(a_j)\right)RR_I^{(k)}\lambda^{(j)}(t)V_k^{(j)}$$

$$\frac{dS_3^{(j)}}{dt} = \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}S_3^{(j-1)} + \omega\left(A_2^{(j)} + A_3^{(j)}\right) - \left(\alpha_3\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_3^{(j)}$$

$$\frac{dS_{3A}^{(j)}}{dt} = \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}S_{3A}^{(j-1)} + \beta S_3^{(j)} - \left(\alpha_3\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{3A}^{(j)}$$

$$\frac{dS_{3B}^{(j)}}{dt} = \left((1-\phi_t)\mathbb{I}_{\{2\}}(j) + \mathbb{I}_{\{3,\dots,19\}}(j)\right)\delta^{(j-1)}S_{3B}^{(j-1)} + \beta S_{3A}^{(j)} - \left(\alpha_3\lambda^{(j)}(t) + \beta + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)S_{3B}^{(j)}$$

$$\frac{dI_3^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}I_3^{(j-1)} + \theta_3\Psi(a_j)\alpha_3\lambda^{(j)}(t)\left(S_3^{(j)} + S_{3A}^{(j)} + S_{3B}^{(j)}\right) - \left(\mu + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)I_3^{(j)}$$

$$\frac{dA_3^{(j)}}{dt} = \mathbb{I}_{\{2,\dots,19\}}(j)\delta^{(j-1)}A_3^{(j-1)} + (1-\theta_3\Psi(a_j))\alpha_3\lambda^{(j)}(t)\left(S_3^{(j)} + S_{3A}^{(j)} + S_{3B}^{(j)}\right) - \left(\omega + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}}\right)A_3^{(j)}$$

$$(5.2)$$

The force of infection $\lambda(t)$ is defined as in Chapter 4.1.3, i.e.

$$\lambda^{(j)}(t) = \frac{1}{N(t)}\lambda_{\text{seas}}(t)\sum_{i=1}^{n_A}c_{i,j}\lambda_{\text{inf}}^{(i)}(t),$$

$$\frac{dV_1^{(j)}}{dt} = \delta^{(j-1)}V_1^{(j-1)} + \eta_W V_2^{(j)} - \left(RR_I^{(1)}\lambda^{(j)}(t) + \eta_W + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}} \right) V_1^{(j)}$$

$$\frac{dV_2^{(j)}}{dt} = \delta^{(j-1)}V_2^{(j-1)} + \eta_W V_3^{(j)} - \left(RR_I^{(2)}\lambda^{(j)}(t) + \eta_W + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}} \right) V_2^{(j)}$$

$$\frac{dV_3^{(j)}}{dt} = \mathbb{I}_{\{2\}}(j)\phi_t\delta^{(j-1)} \left(M^{(j-1)} + S_1^{(j-1)} + \sum_{k=2}^{3} S_k^{(j-1)} + S_{kA}^{(j-1)} + S_{kB}^{(j-1)} \right) \tag{5.3}$$

$$+ (1 - \mathbb{I}_{\{3\}}(j))\delta^{(j-1)}V_3^{(j-1)} + \eta_W V_4^{(j)} - \left(RR_I^{(3)}\lambda^{(j)}(t) + \eta_W + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}} \right) V_3^{(j)}$$

$$\frac{dV_4^{(j)}}{dt} = \mathbb{I}_{\{3\}}(j)\delta^{(j-1)}V_3^{(j-1)} + \delta^{(j-1)}V_4^{(j-1)} - \left(RR_I^{(4)}\lambda^{(j)}(t) + \eta_W + \delta^{(j)} + \gamma^{(j)} - \frac{m^{(j)}}{N^{(j)}} \right) V_4^{(j)}$$

where

$$\lambda_{\text{seas}}(t) = \exp\left\{ a_1 \cos 2\pi \left(\frac{t}{52} - b_1 \right) + a_2 \cos 2\pi \left(\frac{t}{26} - b_2 \right) \right\},$$

$$\lambda_{\text{inf}}^{(j)}(t) = \sum_{k=1}^{3} \left(pI_k^{(j)}(t) + A_k^{(j)}(t) \right).$$

Regarding the contact pattern, here we consider only the contact matrix $\boldsymbol{C}_6$ as given in Chapter 4.2.3. Thus, the contact parameters $c_{i,j}$ are defined by

$$c_{i,j} = \begin{cases} c_1 & \text{if } i, j \in \{1:14\} \text{ (within first 5 years of age)} \\ c_2 & \text{if } i, j \in \{15:17\} \text{ (within 5 -59 years of age)} \\ c_3 & \text{else} \end{cases}$$

The number of new symptomatic cases occurring in week t according to the model is computed by

$$Y^{(j)}(t) = \int_t^{t+1} H^{(j)}(u)\, du.$$

with

$$H^{(j)}(t) = \theta_1 \Psi(a_j)\lambda^{(j)}(t)S_1^{(j)}(t) + \alpha_2\theta_2\Psi(a_j)\lambda^{(j)}(t) \left(S_2^{(j)}(t) + S_{2A}^{(j)}(t) + S_{2B}^{(j)}(t) \right)$$

$$+ \sum_{k=1}^{4} \left(\theta_1 RR_S^{(k)}\Psi(a_j) \right) RR_I^{(k)}\lambda^{(j)}(t)V_k^{(j)} + \alpha_3\theta_3\Psi(a_j)\lambda^{(j)}(t) \left(S_3^{(j)}(t) + S_{3A}^{(j)}(t) + S_{3B}^{(j)}(t) \right).$$

The number of reported cases is assumed to be a fraction h of all symptomatic cases $\boldsymbol{Y}$ where the time-dependent detection rate $h(t)$ is defined by

$$h(t) = \begin{cases} h & \text{if } t \leq t_{2005} & \text{(until 2005)} \\ h + (qh - h)\frac{t - t_{2005}}{26} & \text{if } t_{2005} < t \leq t_{2005} + 26 & \text{(first half year 2005)} \\ qh & \text{if } t > t_{2005} + 26 & \text{(after first half year 2005)} \end{cases}$$

Again, see Chapter 4.1.4 and Table 5.2 for further details.

5.1.4 Investigated model scenarios

We constructed various model scenarios to investigate which model features are required to explain the pre- and post-vaccination incidence data (2001-2008 and 2009-2013, respectively) and which could be omitted to obtain a parsimonious model. Model scenarios were constructed by either including or excluding age-specific immunity levels, the possibility of natural immunity waning, symptomatic infections within the third infection cycle $(S_3 \to I_3)$, irregular waning of vaccine protection, and a simpler seasonality term. Thus, the parameters $\boldsymbol{\psi}$, β, θ_3, η_W, and (a_2, b_2) are either subject of estimation or are being kept fixed at neutral values, respectively. The set of considered model scenarios $\mathcal{M}$ is defined as follows:

- "BC": base-case model

- "AI": BC with age-specific immunities $\Psi(a)$

- "NW": AI without waning of natural immunity $(\beta = 0)$

- "I_3": AI allowing symptomatic infections in third level $(\theta_3 > 0)$

- "WV_β": AI with non-constant waning of vaccine protection η_W

- "WV_β^{1S}": "WV_β" with only one seasonality term $(a_2, b_2 = 0)$, i.e.

$$\lambda_{\text{seas}}(t) = \exp\left\{ a_1 \cos 2\pi \left(\frac{t}{52} - b_1 \right) \right\}$$

- "WV": WV_β without waning of natural immunity $(\beta = 0)$

- "FM": full model accounting for all possible aspects

Thus, the final parameter vector $\boldsymbol{\vartheta}$ to estimate differs for each model scenario in $\mathcal{M}=\{$BC, AI, NW, I_3, WV_β, WV_β^{1S}, WV, FM$\}$. An overview of all model scenarios and their included parameters is given in Table 5.1.

5.2 Model inference and estimation of vaccine effectiveness

With the licensure of the two rotavirus vaccines in 2006, vaccination has been increasingly applied in Germany, especially in the eastern federal states (EFS) and even without being recommended by the Standing Committee on Vaccination (STIKO). Thus, we already observe a considerable vaccination coverage from 2008 onwards together with an incidence decrease among children less than five years of age in that region (see Figures 5.3). Using

Table 5.1: Summary table over all potential model parameters and the 8 model scenarios. A "∅" denotes that the parameter is not included in the respective scenario. Thus, in scenarios not including the parameter β waning of natural immunity is not possible. Scenarios not including θ_3 do not allow for symptomatic infections within the third infectivity stage. Scenarios not including η_W have a fixed time for waning vaccine protection as described in Section 5.1.1. Scenarios not including ψ do not account for age-specific immunities. Scenarios not including (a_2, b_2) account for only one harmonic wave in the seasonality.

parameter	interpretation	model scenario BC	AI	NW	I_3	WV_β	WV_β^{1S}	WV	FM
μ	rate of recovery from symptomatic infection	·	·	·	·	·	·	·	·
ω	rate of recovery from asymptomatic infection	·	·	·	·	·	·	·	·
ω_0	rate of waning maternal protection	·	·	·	·	·	·	·	·
p	relative infectiousness of sympt. infected	·	·	·	·	·	·	·	·
β	rate of waning natural immunity	·	·	∅	·	·	·	∅	·
c_1, c_2, c_3	age-specific contact rates	·	·	·	·	·	·	·	·
a_1, b_1	amplitude/phase shift of seasonality	·	·	·	·	·	·	·	·
a_2, b_2	amplitude/phase shift of 2nd order seasonality	·	·	·	·	·	∅	·	·
$h^{(e)}$	detection rate before 2005	·	·	·	·	·	·	·	·
$q^{(e)}$	relative increase of detection rate	·	·	·	·	·	·	·	·
d	overdispersion of observation distribution	·	·	·	·	·	·	·	·
θ_3	probability of symptomatic infection in 3rd stage	∅	∅	∅	·	∅	∅	∅	·
$\phi_{2006,\ldots,2013}$	yearly vaccine coverage proportion	·	·	·	·	·	·	·	·
η_I	vaccine protection against infection	·	·	·	·	·	·	·	·
η_S	vaccine protection against symptoms	·	·	·	·	·	·	·	·
η_W	rate of waning vaccine protection	∅	∅	∅	∅	·	·	·	·
ψ_1, ψ_2, ψ_3	age-specific immunities	∅	·	·	·	·	·	·	·
	Total number of parameters to estimate	25	28	27	29	29	27	28	30

this data on increasing coverage rates and decreasing incidence we aim to estimate the effectiveness of the rotavirus vaccination through the transmission model.

In this section we will describe the employed Bayesian inference framework for parameter estimation. In our previous modelling from Chapter 4, we considered case notification data from both, eastern and western federal states, subject to the assumption that the true underlying incidence is equal in both regions until 2008 and the observed differences arise from a differing reporting behaviour (Rosner et al., 2010). However, the vaccination coverages (in 2009/10: WFS 22-28%; EFS 56-59%) and thus also the resulting decrease of the true incidences differed significantly between the two regions of eastern and western federal states (Dudareva et al., 2012). Hence, for this analysis the transmission model is restricted to the EFS within this chapter, since the observed impact was more pronounced in this region.

5.2.1 Epidemiological and demographic data

We again obtained data on the number of rotavirus cases from the nationwide surveillance system of the Robert Koch Institute (Krause et al., 2007) as described in Section 4.1.1. We analysed the weekly number of reported rotavirus cases in the EFS from the beginning of 2001 till June 2013 ($t \in \{0, \dots, T = 650\}$), stratified by ten age groups corresponding to the original model from Chapter 4 ($<$1, 1, 2, 3, 4, 5-19, 20-39, 40-59, 60-79, $\geq$80 years of age). The weekly rotavirus incidence, i.e. cases per 100,000 people, for three aggregated age groups is displayed in Figure 5.3. The decreasing incidence within the youngest age group with beginning of 2010 is clearly visible.
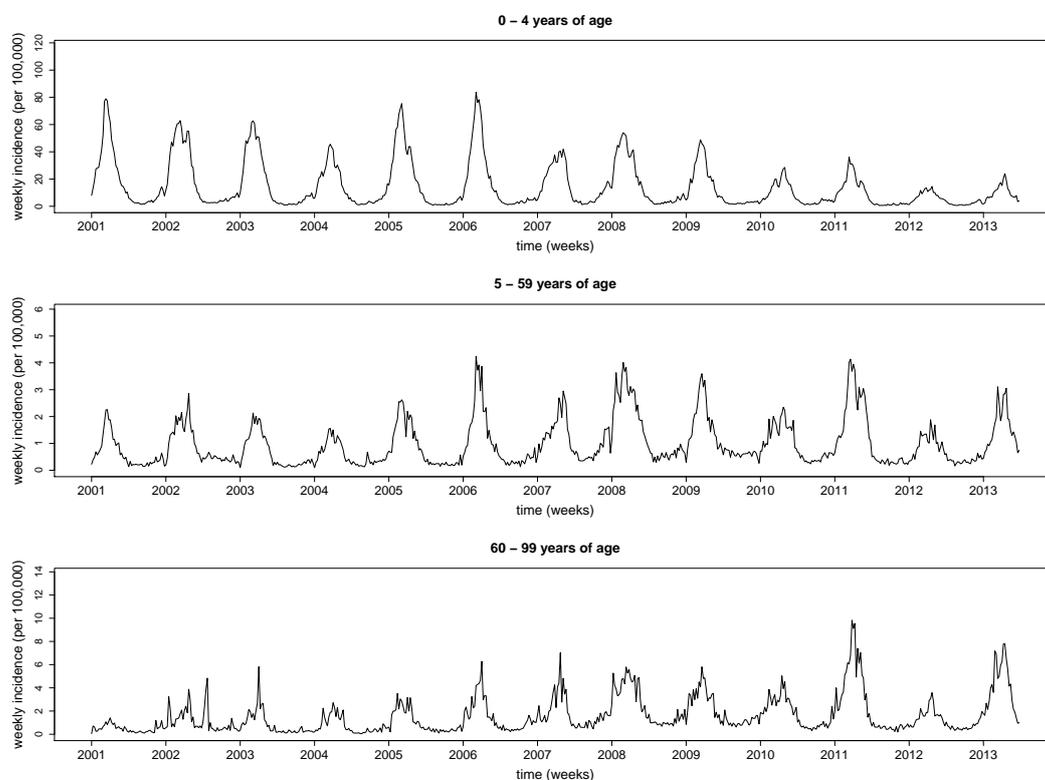

Figure 5.3: Weekly reported rotavirus incidence in the EFS for the age groups 0-4, 5-59, and >60 years of age from 2001 until mid 2013. Note the different scaling of the y-axes.

Estimates for the yearly rotavirus vaccine coverage rates in the EFS since licensure of the two vaccines were obtained from a questionnaire survey (Dudareva et al., 2012). Within this setting we defined the yearly vaccine coverage as the proportion of children receiving the vaccination among all children eligible for vaccination in that specific year, i.e all children reaching three months of age during that year. Vaccine efficacy data for both rotavirus vaccines was available from a meta analysis including over 60 clinical trials (Soares-Weiser et al., 2012).

In the model we again assumed a non-constant population with time-varying birth rates, death rates and migration. The necessary data regarding these demographic aspects were obtained from the federal bureau of statistics (Federal Bureau of Statistics, 2013).

5.2.2 Bayesian inference procedure

Within a Bayesian framework the parameters are inferred for each model scenario, given the extended incidence data set $\boldsymbol{D}$ from the pre-vaccination era (2001-2008) and the post-vaccination era (2009-2013). As prior knowledge on the new parameters regarding vaccination coverage and vaccine effectiveness is available from the survey and the meta-analysis, respectively, this further supports the use of a Bayesian approach. While all parameters are subject of estimation, we are especially interested in the posterior distribution of the parameter vector $\boldsymbol{\eta} = (\eta_I, \eta_S, \eta_W)$ determining the vaccine effectiveness.

The statistical model corresponds to that presented in Section 4.2. The likelihood $f(\boldsymbol{D}\,|\boldsymbol{\vartheta})$ of the data $\boldsymbol{D}$ given the model parameter vector $\boldsymbol{\vartheta}$ is defined via the dynamic transmission model as described in Section 5.1. Subject to $\boldsymbol{\vartheta}$ the model computes the corresponding time series for the expected number of new rotavirus cases $\boldsymbol{Y} = (Y^{(j)}(t))$ for each age group $j = 1, \ldots, n_A$ and each week $t = 1, \ldots, T$. Note that $\boldsymbol{Y}$ is indeed a deterministic function of the input parameter vector $\boldsymbol{\vartheta}$, which is governed by the ODE systems (5.2) and (5.3). However, $\boldsymbol{Y}$ is not a deterministic variable due to the input parameter $\boldsymbol{\vartheta}$ being a random variable within the Bayesian framework.

The actual number of reported cases $(X_t^{(j)})$ in week t and age group j is then assumed to be negative binomially distributed with expectation $h_{\boldsymbol{\vartheta}}(t) \cdot Y^{(j)}(t)$ and dispersion parameter $d_{\boldsymbol{\vartheta}} \in \boldsymbol{\vartheta}$, where $h_{\boldsymbol{\vartheta}}(t)$ is the time specific reporting rate depending on parameters from $\boldsymbol{\vartheta}$ (see Section 5.1.3), i.e.

$$X_t^{(j)} \sim \text{NegBin}\left(h_{\boldsymbol{\vartheta}}(t)Y^{(j)}(t), d_{\boldsymbol{\vartheta}}\right).$$

The vector $\boldsymbol{X}$ is the observed quantity within our setting, i.e. corresponds to the rotavirus case data. Thus, the likelihood function $f(\cdot\,|\boldsymbol{\vartheta})$ is defined as

$$f(\boldsymbol{D}\,|\boldsymbol{\vartheta}) = \prod_{j=1}^{n_D} \prod_{t=1}^{T} p_{\text{NegBin}}\left(D_t^{(j)}, h_{\boldsymbol{\vartheta}}(t)Y^{(j)}(t), d_{\boldsymbol{\vartheta}}\right)$$

with p_{NegBin} being the probability mass function of the negative binomial distribution. To account for the dependencies among the observed data points we again apply the CA-based adjustments of the likelihood function, which was employed in Section 4.2. Thus the adjusted likelihood function f_{CA} is defined by

$$f_{\text{CA}}(\boldsymbol{D}\,|\boldsymbol{\vartheta}) = \prod_{j=1}^{n_D} \prod_{t=1}^{T} p_{\text{NegBin}}\left(D_t^{(j)}, h_{\boldsymbol{\vartheta}}(t)Y^{(j)}(t), d_{\boldsymbol{\vartheta}}\right)^{1/\widehat{\text{CA}}^{(e,j)}},$$

where $\widehat{\text{CA}}^{(e,j)}$ are the estimated cumulative autocorrelations for the ten age specific time series from the EFS as described in Section 4.2.2. The final posterior distribution of the

whole parameter vector is then given by

$$\pi\left(\boldsymbol{\vartheta}\,|\,\boldsymbol{D}\right) \propto f_{\mathrm{CA}}\left(\boldsymbol{D}\,|\,\boldsymbol{\vartheta}\right)\pi\left(\boldsymbol{\vartheta}\right),$$

where $\pi\left(\boldsymbol{\vartheta}\right)$ denotes the prior distribution of the parameter vector $\boldsymbol{\vartheta}$, which will be discussed in the following section. Note again, that the likelihood function and prior distribution depend on the model scenario under consideration, where especially the prior probability may be concentrated in one point for some components due to parameters being fixed in certain scenarios (see Section 5.1.4). Thus, to point out this dependence we will mark the respective object with an upper index (i), e.g. $\boldsymbol{\vartheta}^{(i)}$ denotes the parameter vector corresponding to model i in $\mathcal{M}$.

Prior elicitation

Prior distributions for the parameter vectors $\boldsymbol{\vartheta}_{\mathrm{NV}}$, $\boldsymbol{\eta}$, $\boldsymbol{\phi}$, $\boldsymbol{\psi}$, and θ_3 (see Table 5.1) were defined separately, where $\boldsymbol{\vartheta}_{\mathrm{NV}}$ denotes the vector of parameters used in the previous model version without vaccination, i.e.

$$\boldsymbol{\vartheta}_{\mathrm{NV}} = \left(\omega_0, \mu, \omega, \beta, p, c_1, c_2, c_3, a_1, a_2, b_1, b_2, h^{(e)}, q^{(e)}, d\right).$$

Thus, as prior π_{NV} for $\boldsymbol{\vartheta}_{\mathrm{NV}}$ we used the same prior distribution as done in the transmission modelling for the pre-vaccination era in Germany described in Chapter 4, which utilized skew normal distributions ($\mathcal{SN}$) for prior elicitation. One modification was made to the prior of the contact parameters c_i $(i = 1, 2, 3)$. In order to avoid implausible deviations between the age-dependent contact behaviour as discussed in Section 4.2.2 we additionally required

$$\pi_{\mathrm{NV}}\left(c_1, c_2, c_3\right) \propto \exp\left\{-\exp\left(\left(\log\frac{c_1}{c_2}\right)^2\right) - \exp\left(\left(\log\frac{c_1}{c_3}\right)^2\right) - \exp\left(\left(\log\frac{c_2}{c_3}\right)^2\right)\right\},$$

thus imposing a stricter prior, which heavily penalizes too large relative differences between the contact parameters c_i.

To construct a prior distribution for the parameter $\boldsymbol{\eta} = (\eta_I, \eta_S, \eta_W)$ concerning the vaccine effectiveness we utilized data from the meta-analysis in Soares-Weiser et al. (2012) considering the vaccine efficacy of the two rotavirus vaccines Rotarix and RotaTeq with respect to specific clinical outcomes. Within the meta-analysis, both vaccines were found to yield protection against symptomatic infection of any severity after one year corresponding to a risk ratio of 0.26 (95% CI: 0.17-0.39). Assuming this risk ratio to be a log-normally distributed random variable T, the point estimator and confidence interval translate into an expected value of $\mathbb{E}[T] = 0.26$ and variance $\mathrm{Var}(T) = 0.019$. Within the model context a symptomatic infection of any severity was thereby defined as any symptomatic infection. Thus, in our model T is the product of the relative risks for infection and developing symptoms in case of infection, respectively.

$$T = \eta_I \cdot \eta_S$$

Since no further prior information was available, we derived individual prior distributions for the parameters η_I and η_S by assuming their priors to be independently and identically distributed. Hence, the distributions of both η_I and η_S have expected value

$$\mathbb{E}[\eta_S] = \mathbb{E}[\eta_I] = \sqrt{\mathbb{E}[T]} = 0.51$$

and variance

$$\mathrm{Var}\,(\eta_S) = \mathrm{Var}\,(\eta_I) = \sqrt{\mathrm{Var}\,(T) + \mathbb{E}[\eta_I]^4} - \mathbb{E}[\eta_I]^2 = 0.034.$$

Assuming a logit-normal distribution for both parameters η_I and η_S, from numerical optimization we obtain that

$$\mathrm{logit}\,(\eta_I) \sim \mathcal{N}\,(-0.12, 0.87)$$

with η_S being identically distributed. Considering η_W in scenarios allowing a continuous immunity waning, we assumed that loss of one immunity level, e.g. moving from V_k to V_{k-1}, happens on average once per year. As the evidence on the speed of waning is sparse, we defined an uninformative prior by

$$\log\,(\eta_W) \sim \mathcal{N}\,\left(\log\frac{1}{52}, 0.5\right).$$

The final joint prior distribution $\pi_{\boldsymbol{\eta}}$ is constructed by assuming independence of the parameters in $\boldsymbol{\eta}$.

From a nationwide questionnaire survey (Dudareva et al., 2012) point estimates and confidence intervals for the yearly rotavirus vaccine coverage among infants in the EFS from 2006 till 2011 were ascertained. Our prior estimates for the years 2012 and later were based on the least known information from 2011. In contrast to other parameters we a priori restricted the vaccine coverage $\boldsymbol{\phi}$ to the range provided by the survey. In preliminary studies we found that allowing the whole $[0, 1]$-range for the yearly coverage, the model tries to explain the seasonal incidence variation by extreme changes of the yearly coverages ϕ_i, $i \in 2006, \ldots, 2013$. Since the observed incidence fluctuations are likely to be not solely a product of varying coverage rates but also other random effects such as seasonally differing genotypes and environmental conditions, we restricted the coverage parameters to a reasonable range to obtain more stable results. Therefore, prior expectations and ranges were defined as

$$\left(\mu_{2006}^{(C)}, \ldots, \mu_{2013}^{(C)}\right) = (0.04, 0.07, 0.40, 0.59, 0.56, 0.56, 0.56, 0.56),$$

and

$$\left(\mathrm{R}_{2006}^{(C)}, \ldots, \mathrm{R}_{2013}^{(C)}\right) = (0.06, 0.09, 0.17, 0.16, 0.23, 0.23, 0.23, 0.23),$$

respectively, such that the yearly vaccine coverage ϕ was bound to the interval $[\mu_i^{(C)} - R_i^{(C)}/2, \mu_i^{(C)} + R_i^{(C)}/2]$, $i \in 2006, \ldots, 2013$. Assuming a logit-normal distribution on the given ranges leads to the following priors

$$\mathrm{logit}\,\left(\frac{1}{2} + \frac{\phi_i - \mu_i^{(C)}}{\mathrm{R}_i^{(C)}}\right) \sim \mathcal{N}\,(0, 1), \qquad i = 2006, \ldots, 2013$$

Again, the joint prior distribution π_ϕ is constructed by assuming independence between the parameters. This enables also non-monotonic increases in the yearly vaccine coverage rate. The constructed prior parameter ranges are displayed in Figure 5.6 on page 147.

Regarding the parameters $\boldsymbol{\psi}$ governing age-specific immunity, we chose a prior that penalizes heavy immunity variations with changing age by restricting especially the first and second order coefficient ψ_2, ψ_3. Thus, we set

$$\psi_i \sim \mathcal{N}\left(0, \sigma_i^2\right), \qquad i = 1, 2, 3$$

with $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 0.01, 0.0001)$. Again, assuming independence yields π_ψ.

For parameter θ_3 we wanted to secure that $\theta_3 < \theta_2$, where $\theta_2 = 0.25$ denotes the fixed baseline probability of developing symptoms in the second stage. Thus, we assumed

$$\mathrm{logit}\left(\frac{\theta_3}{\theta_2}\right) \sim \mathcal{N}\left(0; 100\right)$$

to define a vague prior over the possible interval.

Finally, the total prior distribution π with respect to the whole input parameter $\boldsymbol{\vartheta}$ is obtained by assuming independence between the components, e.g.

$$\pi(\boldsymbol{\vartheta}) = \pi\left(\boldsymbol{\vartheta}_{\mathrm{NV}}, \boldsymbol{\phi}, \boldsymbol{\eta}, \boldsymbol{\psi}, \theta_3\right) = \pi_{\mathrm{NV}}\left(\boldsymbol{\vartheta}_{\mathrm{NV}}\right) \pi_\phi\left(\boldsymbol{\phi}\right) \pi_\eta\left(\boldsymbol{\eta}\right) \pi_\psi\left(\boldsymbol{\psi}\right) \pi_{\theta_3}\left(\theta_3\right).$$

Analogously to $\boldsymbol{\vartheta}^{(i)}$ also the prior density depends on the considered model scenario in $\mathcal{M}$, where the corresponding prior $\pi^{(i)}$ is defined as the marginal distribution of π with respect to the relevant non-fixed parameters in $\boldsymbol{\vartheta}^{(i)}$.

Posterior computation and implementation

To generate a sample from each models posterior distribution $\pi^{(i)}(\boldsymbol{\vartheta}^{(i)}|\boldsymbol{D})$, $i \in \mathcal{M}$, we applied the adaptive MCMC algorithm which was presented in Chapter 3.2.3 (Algorithm 2). The adaptive algorithm provides a suitable tool to sample from intractable posterior distributions on high-dimensional parameter spaces.

For each of the eight models considered within our application we computed a preliminary chain $\tilde{\boldsymbol{\Theta}}^{(i)} = (\tilde{\boldsymbol{\vartheta}}_j^{(i)})_{j=1,\dots,J}$. For a simpler notation we will drop the index (i) in the following, but keep in mind that prior $\pi^{(i)}$, posterior $\pi^{(i)}\left(\cdot|\boldsymbol{D}\right)$, likelihood $f^{(i)}\left(\boldsymbol{D}|\cdot\right)$, and parameter vector $\boldsymbol{\vartheta}^{(i)}$ differ for each model.

As initial proposal variance Σ_0 we used the prior covariance

$$\tilde{\Sigma}_0 = \frac{1}{1000}\,\mathrm{Cov}_\pi\left(\boldsymbol{\vartheta}\right),$$

which was downscaled by the factor 1000 as the prior covariance is much larger than the posterior covariance we also aim to approximate within the adaptive algorithm. Thus, starting with a proposal covariance matrix which suggests too small steps is a safe choice

to guarantee that chain starts moving such that the adaptive algorithm can do its work. The initial chain value $\tilde{\boldsymbol{\vartheta}}_0$ was sampled from the corresponding prior π. We set the initial period to $K = 3,000$ and computed $J = 20,000$ samples within each chain. Based on the preliminary $\tilde{\boldsymbol{\Theta}}$ we decided for each model $i \in M$ by visual diagnostic if it was capable of describing the incidence data. This was done by plotting the rotavirus incidence data against the model output corresponding to the respective sample mode from each chain, i.e. the chain's parameter vector $\tilde{\boldsymbol{\vartheta}}_j$ yielding the highest posterior value. If the sample mode from a specific model $i \in \mathcal{M}$ could not mimic the age and season specific characteristics of the rotavirus incidence we did not pursue that model scenario for further analysis.

Then for all models capable of explaining the data according to the preliminary results we computed a proper sample $\boldsymbol{\Theta}^{(i)}$ by running three separate chains for each model. For these we choose the chain length $J = 100,000$ and the initial period $K = 10,000$. As the initial transition variance we defined

$$\Sigma_0 = s_d \widehat{\text{Cov}} \left(\tilde{\boldsymbol{\vartheta}}_{10,000}, \ldots, \tilde{\boldsymbol{\vartheta}}_{20,000} \right)$$

based on the preliminary results with $s_d = 2.4^2/d$ being the optimal scaling factor according to Gelman et al. (1996). The initial value $\boldsymbol{\vartheta}_0$ was set to $\tilde{\boldsymbol{\vartheta}}_{20,000}$ to guarantee that the chain starts within a high posterior region. The first 20.000 components of each chain were discarded as burn-in. We used convergence diagnostic tools proposed by Geweke (1992) to test whether the chains represent samples from the same distributions. The final chain $\boldsymbol{\Theta}^{(i)}$ corresponding to model $i \in \mathcal{M}$ was then constructed by combining the three separate chains. This full chain consisting of 240,000 parameter vectors was used for all further analyses.

Model selection

One aim of our modelling approach was to find the transmission model among our set of considered models $\mathcal{M}$ given in Section 5.1.4 that is most suitable to explain the rotavirus incidence data while also accounting for parsimony. As a criteria to select "a best" model we computed the posterior model probability $\pi^{(\mathcal{M})}(i\,|\,\boldsymbol{D})$ for each model $i \in \mathcal{M}$, which we introduced in Chapter 2.1.2. To do so, we assumed equal prior probabilities for all considered models. Since the marginal likelihoods $f^{(\mathcal{M})}(\boldsymbol{D}\,|\,i)$ necessary for computing the model probabilities can not be assessed analytically within our complex setting, we applied the newly introduced Algorithm 4 from Chapter 3.3.3 to estimate the marginal likelihoods. To apply the algorithm for each model we used the respective posterior samples $\boldsymbol{\Theta}^{(i)}$ generated from the adaptive MCMC procedure.

In Bayesian model selection the model probabilities $\pi^{(\mathcal{M})}(i\,|\,\boldsymbol{D})$ would yield a final mixture of models. However, as our results later suggest and since the dominance of one single model was also already apparent in the modelling from Chapter 4.3, we instead decided to further consider only the best model yielding the highest posterior probability. This model was used for the subsequent incidence forecast presented in Section 5.3.

5.2.3 Results

Model selection

The adaptive sampling algorithm was applied to all eight models $\mathcal{M} = \{$BC, AI, NW, I_3, WV$_\beta$, WV$_\beta^{1S}$, WV, FM$\}$ defined previously. As a result from the preliminary posterior sampling we excluded both models not accounting for waning of natural immunity (NW, WV) since those models yielded very low likelihoods and a distinct discrepancy between model output and data. This confirms our initial assumption that waning of natural immunity acquired by infection is an important aspect of the rotavirus transmission dynamics, especially among elderly.

Altogether, this left six models for further analysis. Conducting the comprehensive MCMC run and computing the marginal likelihoods for each model yielded the following model probabilities assuming equal prior probabilities:

$$\pi^{(\mathcal{M})}\left(\text{BC}\,|\,\boldsymbol{D}\right) < 0.001, \quad \pi^{(\mathcal{M})}\left(\text{AI}\,|\,\boldsymbol{D}\right) < 0.001, \quad \pi^{(\mathcal{M})}\left(\text{I}_3\,|\,\boldsymbol{D}\right) < 0.001,$$

$$\pi^{(\mathcal{M})}\left(\text{WV}_\beta\,|\,\boldsymbol{D}\right) = 0.973, \quad \pi^{(\mathcal{M})}\left(\text{WV}_\beta^{1S}\,|\,\boldsymbol{D}\right) < 0.001, \quad \pi^{(\mathcal{M})}\left(\text{FM}\,|\,\boldsymbol{D}\right) = 0.026.$$

Four noteworthy conclusions can be drawn from this. Firstly, models including an age-specific immunity are better able to explain nuances in the age-stratified incidence, particularly considering the age groups greater than 40 years of age, while the only model assuming constant age-independent immunity (BC) yielded the lowest marginal likelihood.

Secondly, assuming a continuous waning of vaccine induced immunity provides a higher posterior probability (see models WV$_\beta$, FM, 1S) than assuming that immunity loss is strictly linked to age. Moreover, the estimated mean waning rate suggests that the loss of vaccine protection happens more slowly than one immunity level per year (see Figure 5.7).

Thirdly, accounting for symptomatic infections within the highest susceptibility stage does not improve the explanatory potential of the model (FM). The actual ratio of symptomatic infection in this stage θ_3 was estimated to be negligible in this model scenario. Hence, according to the model, the primary cause for infections among elderly remains to be the absence of virus contact with higher age and the resulting loss of immunity.

Lastly, the model WV$_\beta^{1S}$ using solely one harmonic wave within the transmission seasonality λ_{seas} could not fully reproduce the seasonal peaks in the rotavirus incidence. This implies that the underlying seasonal variation in transmission is even more concentrated with respect to the winter months as one might expect.

All further analyses, e.g. prediction of vaccination impact, single parameter estimates, are computed based on model scenario WV$_\beta$ which yielded the highest model probability and thus provides the best fit to the rotavirus incidence data.

Posterior distribution

To check for convergence of the generated sample chain, we exemplarily looked at the chain trajectories for the parameters regarding the vaccine effectiveness parameters as given in

Table 5.2: Summary table over all model parameters. For parameters which were fixed in advance corresponding values and literature evidence is provided. For parameters estimated during model calibration mean estimates together with equitailed 95% credibility intervals from the sampled posterior distribution corresponding to the best fitting model scenario WV$_\beta$ are given.

parameter	interpretation	estimated or fixed value	prior distribution	literature evidence
α_1, α_2	reduced susceptibility after previous infections	(0.6; 0.4)	fixed	(Fischer et al., 2002; Velásquez et al., 1996; Mrukowicz et al., 1999)
$\theta_1, \theta_2, \theta_3$	probability of developing symptoms	(0.5; 0.25; 0)	fixed	(Fischer et al., 2002; Velásquez et al., 1996; Mrukowicz et al., 1999; Anderson and Weber, 2004)
μ	rate of recovery from symptomatic infection	1.58[1.14; 1.85]	$\log \mu \sim \mathcal{SN}(0.69, 0.31, 83)$	(Anderson and Weber, 2004)
ω	rate of recovery from asymptomatic infection	1.45[0.92; 1.81]	$\log \omega \sim \mathcal{SN}(0.08, 0.34, -1.38)$	(Pickering et al., 1988; Wilde et al., 1991)
ω_0	rate of waning maternal protection	0.037[0.025; 0.051]	$\log \omega_0 \sim \mathcal{SN}(-2.11, 0.35, 0)$	(Clark et al., 2010; Clemens et al., 1993; Pickering et al., 1995)
p	relative infectiousness of sympt. infected	4.58[1.80; 10.04]	$\operatorname{logit} p \sim \mathcal{SN}(0, 1.12, 0)$	(Assumption)
β	rate of waning natural immunity	0.008[0.007; 0.01]	$\log \beta \sim \mathcal{SN}(-3.30, 0.43, -0.13)$	(Anderson and Weber, 2004; Fischer et al., 2002; Ward et al., 1990)
c_1, c_2, c_3	contact rates	3.36 [1.70; 5.75] 4.43 [2.76; 5.57] 2.21 [1.43; 2.96]	$\log c_i \sim \mathcal{SN}(0, 0.82, 0)$	(Assumption)
a_1, a_2	amplitude of seasonality	0.10 [0.08; 0.16] 0.04 [0.03; 0.06]	$a_i \sim \mathcal{SN}(0, 1.02, 0)$	(Assumption)
b_1, b_2	phase shift of seasonality	0.01 [0.0; 0.02] 0.15 [0.11; 0.20]	$\operatorname{logit} b_i \sim \mathcal{SN}(0, 1.87, 0)$	(Assumption)
h	detection rate before 2005	0.13 [0.11; 0.16]	$\operatorname{logit} h \sim \mathcal{SN}(-1.87, 0.47, -2.94)$	(Soriano-Gabarró et al., 2006)
q	relative increase of detection rate	1.22 [1.10; 1.34]	$\log q \sim \mathcal{SN}(0, 0.36, 0)$	(Assumption)
d	overdispersion of observation distribution	4.25 [3.37; 4.85]	$\log d \sim \mathcal{SN}(-0.01, 0.82, 0)$	(Assumption)
ϕ_t	vaccine coverage rate	see Figure 5.6	see Section 5.2.2	(Dudareva et al., 2012)
η_I	vaccine protection against infection	0.82[0.48; 0.96]	$\operatorname{logit}(\eta_I) \sim \mathcal{N}(-0.12, 0.87)$	(Soares-Weiser et al., 2012)
η_S	vaccine protection against symptoms	0.75[0.36; 0.95]	$\operatorname{logit}(\eta_S) \sim \mathcal{N}(-0.12, 0.87)$	(Soares-Weiser et al., 2012)
η_W	rate of waning vaccine protection	0.005[0.001; 0.012]	$\log(\eta_W) \sim \mathcal{N}\left(\log \frac{1}{52}, 1\right)$	(Soares-Weiser et al., 2012)
ψ_1, ψ_2, ψ_3	age-specific immunity coefficients	0.86 [0.24; 1.61] −0.12 [−0.14; −0.10] 0.0013 [0.0011; 0.0017]	see Section 5.2.2	(Assumption)
κ_t	birth rates	time-dependent	fixed	(Federal Bureau of Statistics, 2013)
δ	ageing rates	age-dependent	by width of age group	
γ	death rates	age-time-dependent	fixed	(Federal Bureau of Statistics, 2013)
m	migration rates	age-time-dependent	fixed	(Federal Bureau of Statistics, 2013)

Figure 5.4. Each of the composed chains appears to have mixed well, with a vaninish-
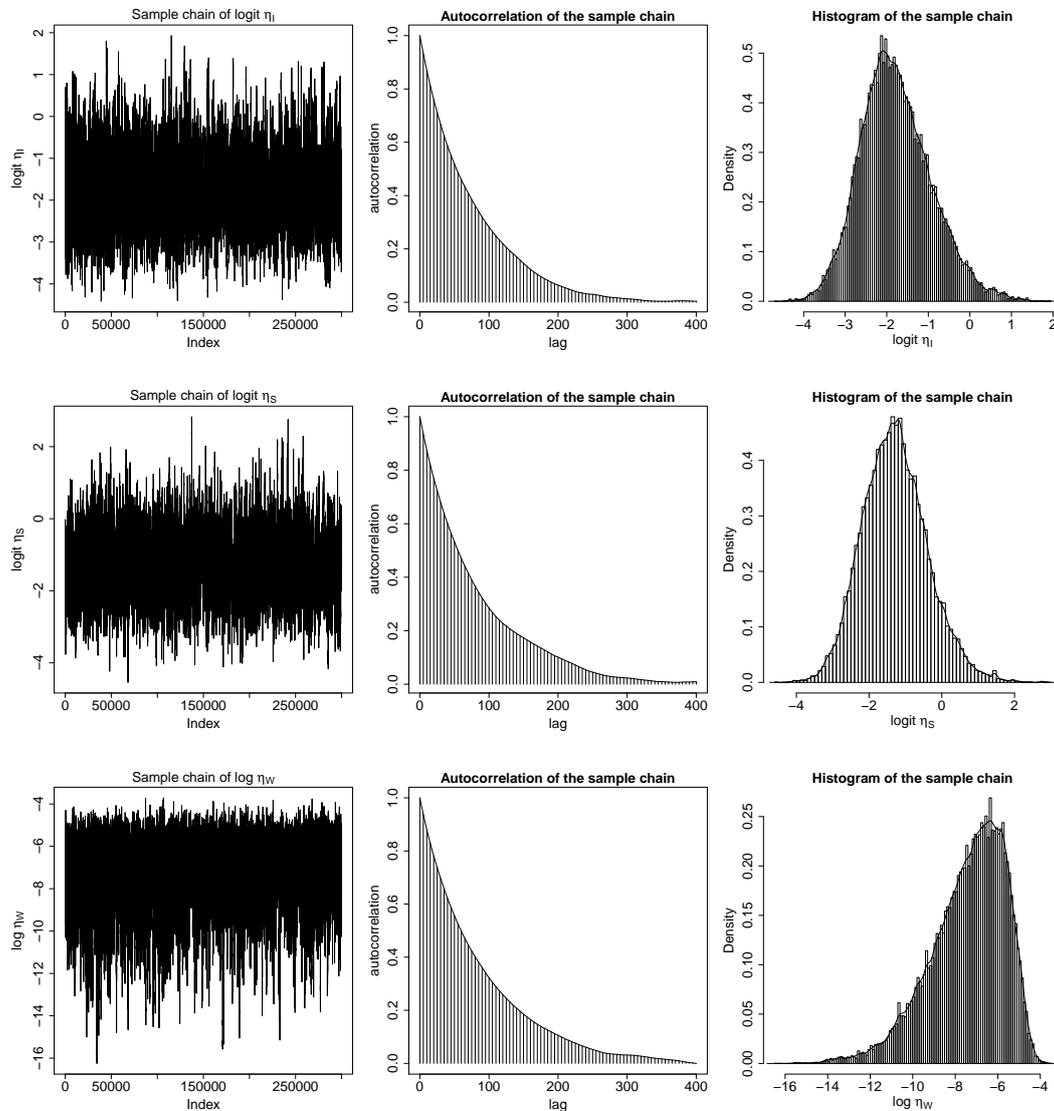
Figure 5.4: Trajectories, autocorrelation functions and histograms of the sample chains regarding the vaccine effectiveness parameters $\boldsymbol{\eta} = (\eta_I, \eta_S, \eta_W)$.

ing sample autocorrelation after about 300 steps. Considering the total chain length of 300,000, the resulting sample certainly provides a good approximate to the posterior distribution, although even longer chains might improve approximation as the kernel-based density estimates are still not entirely smooth. However, conducting the convergence criteria according to Geweke (1992), the respective Z-scores were 0.31, 0.37 and -1.91 for the three displayed chains, suggesting the first and last 120,000 chain values to represent the same distribution (for two samples origining from the same distribution the Z-score statis-

tic is standard normally distributed). This also implies that the three generated MCMC chains sampled from the same distribution. Applying Geweke's convergence diagnostic for all other estimated parameters yielded comparable results.

According to the MCMC sampling output for the model scenario WV_β we computed the empirical posterior means and 95% credibility intervals for each parameter to be estimated. The corresponding results are given in Table 5.2. As the distributions for the single parameters (ψ_1, ψ_2, ψ_3) governing the age-specific relative risk of developing symptoms $\Psi(a)$ (see Equation (5.1)) are difficult to interpret, the resulting age-specific relative risks for each model age group are shown in Figure 5.5.
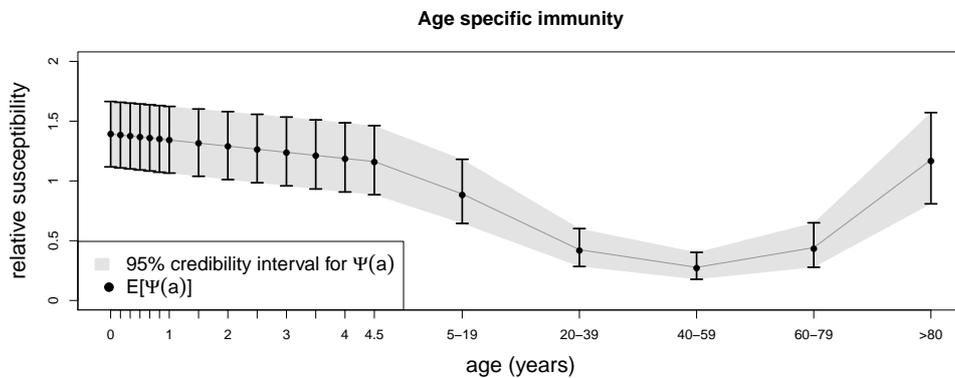

Figure 5.5: Posterior distribution of age specific susceptibilities $\Psi(a)$ for the $n_A = 19$ age groups defined in the model (ticks on the X-axis). The distribution is illustrated by the posterior mean and equitailed 95% credibility intervals.

Epidemiological results

Estimates for epidemiological parameters and vaccine effectiveness from fitting the model to the reported incidence data are given in Table 5.2. Regarding vaccination coverage in the EFS before 2013, our model suggests a substantial increase from nearly zero in 2007/08 to levels of around 60% in the subsequent years (see Figure 5.6).

The relative risk for acquiring rotavirus infection following vaccination was estimated to be 0.18 (95% credibility interval (CI): 0.04 – 0.52) as given in Figure 5.7. The corresponding relative risk for developing symptoms if infected was estimated at 0.25 (95% CI: 0.05 – 0.64). Together this yielded a combined vaccine effectiveness estimate for acquiring symptomatic infection of 96% (95% CI: 91% – 99%). The annual waning rate of vaccine-induced protection was estimated at 0.24 which suggests a median time to loss of one immunity level of 3.11 years (95% CI: 1.43 – 5.80). Thus, protection begins to decrease after three years and is fully waned after around twelve years due to the four vaccine immunity levels considered in the model.
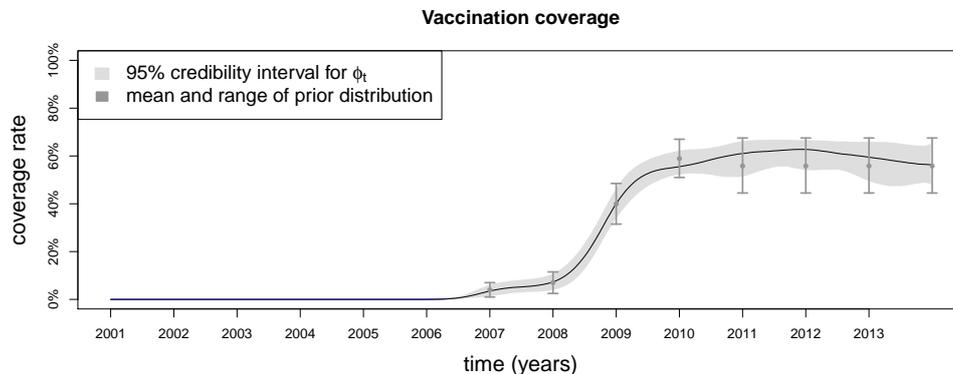
Figure 5.6: Posterior estimate of the yearly rotavirus vaccination coverage rate in the EFS from 2004 till 2013 among children aged three months. The black line denotes the posterior mean, the shaded area illustrates the 95% equi-tailed credibility intervals. Prior distribution ranges and means are marked grey. Prior distributions for the years following 2012 and later were based on the coverage estimates from 2011 (Dudareva et al., 2012).

5.3 Epidemiological impact of rotavirus routine vaccination

In the second part of this study we utilize the fitted model to compute the predictive distribution of the future age-stratified rotavirus incidence subject to an increase of the vaccination coverage among infants less than 6 months of age. To calculate such forecasts we extend the time horizon of our model maintaining the estimated parameter distributions but steadily increasing the coverage rate up to a certain value. Thus, we obtain an estimate for the epidemiological impact of routine vaccination assuming that the driving transmission parameters do not change over time.

To specify a future progress of the coverage rate for the years 2014 and onwards we choose a constant coverage rate ϕ_{long} in the range from 0% to 100%. We assumed that this level will be reached within two years at the end of 2015 by using a sigmoid shaped function assuring that the coverage rate increases smoothly from its estimated level in 2013 to its new long term level ϕ_{long}. For the base scenario we choose a long term coverage level of 90%. The maximum time horizon for incidence forecast was set to year 2030 as the seasonal incidence reached a new steady state by then.

5.3.1 Incidence prediction sampling

To calculate incidence forecasts $K = 1,000$ model parameter vectors were sampled from the posterior distribution $\pi\left(\cdot \mid \boldsymbol{D}\right)$, i.e. they were taken from the MCMC posterior sample, subject to the best fitting model WV_β. Considering future migration, death and birth
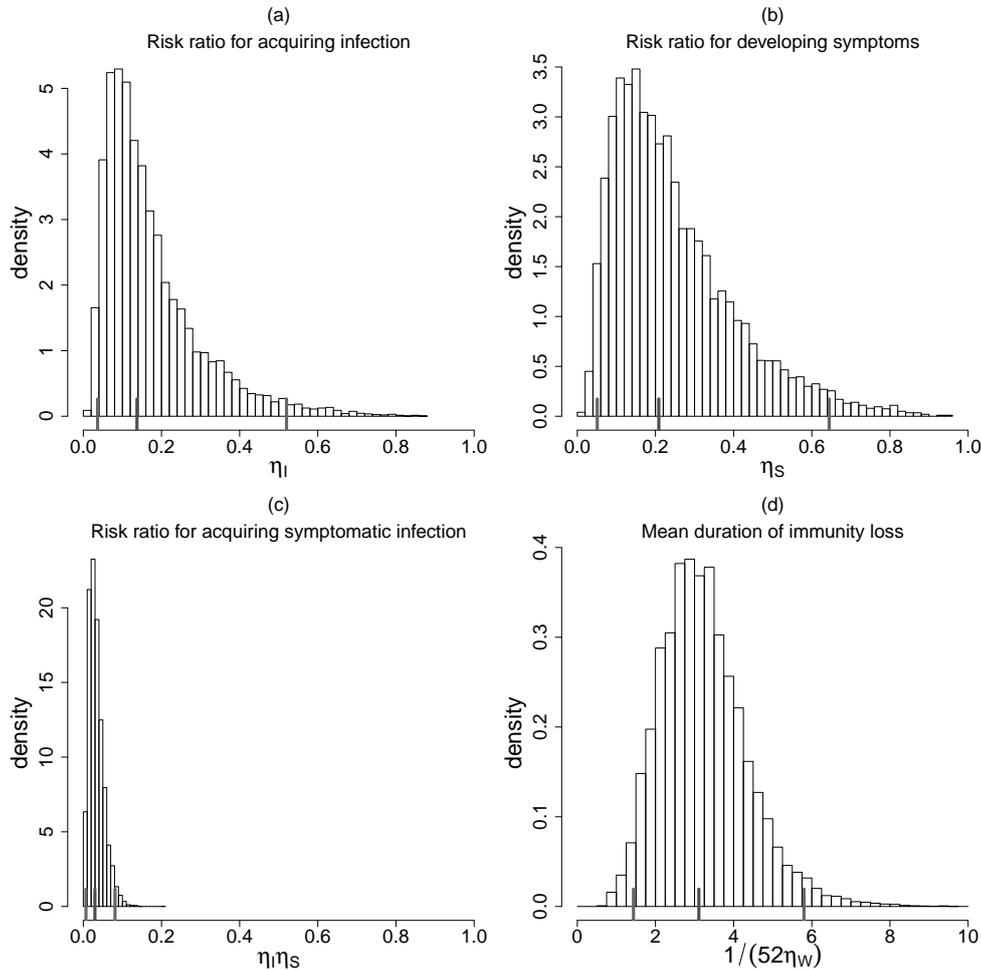
Figure 5.7: Histogram of vaccine effectiveness parameter distribution according to the posterior sample. Posterior medians and 95% equitailed credibility intervals for each parameter are indicated on the x-axis, respectively. Figures (a) and (b) display the vaccine protection against acquiring rotavirus infection (η_I) and developing symptoms (η_S). Figure (c) provides the combined risk ratio for acquiring symptomatic infection ($\eta_I \eta_S$). Figure (d) shows the mean duration of immunity loss in years $\left(52 \cdot \eta_W^{-1}\right)$, i.e. loss of one immunity level.

rates, we assumed that the respective demographic processes would evolve stochastically subject to their past trend until 2013.

For each sampled parameter vector and demographic development both the expected and observed incidence for each week and age group was computed, where the latter was sampled from the negative binomial distribution of the observations. Finally, 95% prediction intervals were calculated for the expected and observed incidence based on their respective sampled distribution. The complete sample procedure for model prediction is

given by Algorithm 6.

Algorithm 6: Predictive incidence sampling from a posterior sample of model parameters

Input: $\boldsymbol{\Theta}$: sample from the model parameter's posterior distribution $\pi\left(\cdot \,|\, \boldsymbol{D}\right)$

Input: F_κ, F_γ, F_m: distributions for the future development of birth, death and migration rates.

Input: ϕ_{long}: long term vaccination coverage level.

Input: K: size of the predictive sample.

Output: $\left\{\boldsymbol{h}_{(k)} \cdot \boldsymbol{Y}_{(k)}^{(j)}\right\}_{k=1,\dots,K} \left\{\boldsymbol{X}_{(k)}^{(j)}\right\}_{k=1,\dots,K}$: predictive samples (of size K) for the expected and reported incidence time series for each age group

- **for** $k = 1$ **to** K **do**

  1. Draw a parameter vector $\boldsymbol{\vartheta}_k$ from the posterior sample $\boldsymbol{\Theta}$ and draw samples for the demographic processes subject to F_κ, F_γ, F_m.

  2. Compute the expected incidences $h_{\boldsymbol{\vartheta}_k}(t) \cdot Y_{\boldsymbol{\vartheta}_k}^{(j)}(t)$ resulting from ODE system corresponding to model WV_β subject to parameter $\boldsymbol{\vartheta}^{(k)}$ and the sampled demographic processes

  3. set $\boldsymbol{h}_{(k)} \cdot \boldsymbol{Y}_{(k)}^{(j)} = \left\{h_{\boldsymbol{\vartheta}_k}(t) \cdot Y_{\boldsymbol{\vartheta}_k}^{(j)}(t)\right\}_{t=1,\dots,T}$

  4. draw a sample for the observed incidence
     $X^{(j)}(t) \sim \mathrm{NegBin}\left(h_{\boldsymbol{\vartheta}_k}(t) \cdot Y_{\boldsymbol{\vartheta}_k}^{(j)}(t), d_{\boldsymbol{\vartheta}^{(k)}}\right)$
     for $t = 1, \dots, T$ and $j = 1, \dots, n_D$

  5. set $\boldsymbol{X}_{(k)}^{(j)} = \left\{X^{(j)}(t)\right\}_{t=1,\dots,T}$

  end

5.3.2 Model validation using WFS data

In order to investigate whether the estimated vaccine effectiveness and the overall calibrated model is transferable to other settings, we used the posterior parameter distribution obtained using the EFS-data to compute the predictive distribution for the notified rotavirus incidence of the western federal states (WFS) from 2001 until 2013 (using Algorithm 6) and compared with the available data. To do so, we acquired the necessary data on demographics (Federal Bureau of Statistics, 2013), vaccine coverage (up to 28% (Dudareva et al., 2012)) and the corresponding notification data (Krause et al., 2007). The only model parameters requiring adjustment were those regarding the reporting rates in the WFS (h, q), for which we used posterior estimates from the model inference in Chapter 4

as seen in Table 4.4.

The resulting predictive distribution together with the data from the WFS is displayed in Figure 5.8. Overall, 87% of the data points are contained in the calculated 95% pointwise prediction bands. The good model fit implies that the employed model structure and parameter estimates are able to reproduce the rotavirus epidemiology also from external settings, which suggests a general applicability of our model and confirms the estimated distributions of our parameters.
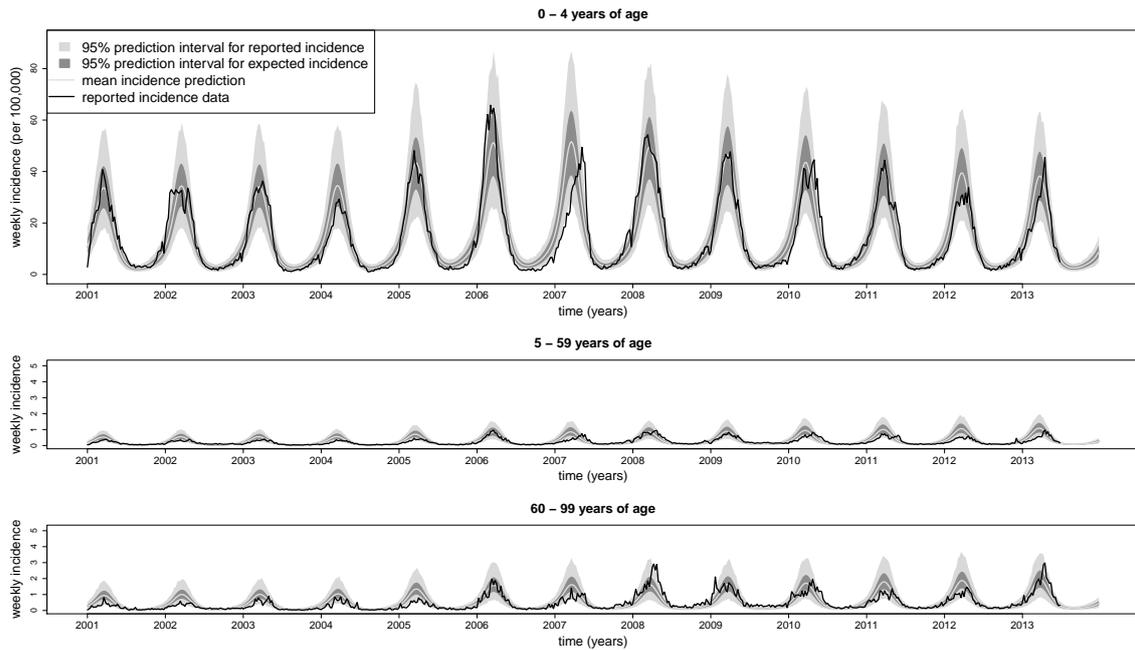

Figure 5.8: Model prediction for expected and observed number of weekly reported rotavirus incidence including pointwise 95% prediction intervals together with observed rotavirus incidence for the three age groups 0-4, 5-59 and 60+ years of age in the WFS from 2001 until 2013.

5.3.3 Investigating demographic uncertainty

To assess the impact of uncertainty around the future demographic development determined by migration, birth and death rates, we fitted suitable stochastic processes to the available demographic data up to 2013.

Considering future migration, we assumed that for each upcoming year the yearly migration rate per age group follows an autoregressive process

$$m_t - \bar{m} = \alpha^{(m)} \left(m_{t-1} - \bar{m} \right) + \epsilon_t, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_m^2),$$

with parameters $\bar{m}$, $-1 < \alpha^{(m)} < 1$ and $\sigma_m > 0$ being calibrated by the corresponding age-specific migration rates in the years 2001-2012. As the migration rates already exhibit

heavy fluctuation even in short periods (Statistisches Bundesamt, 2009) and since the assumption of migration being Markovian appears reasonable we did not account for any higher order correlations within the linear model.

Regarding the age group specific death rates, we modelled the yearly log death rates $\log(\gamma_t)$ to follow a random walk, i.e. the increments of $\log(\gamma_t)$ were assumed to be independently normally distributed

$$\log \gamma_t - \log \gamma_{t-1} = \epsilon_t, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_\gamma, \sigma_\gamma^2),$$

again with mean μ_γ and variance $\sigma_\gamma > 0$ being estimated by the death rates from the years 2001-2012. The linear trend of the death rate process is thus captured in the estimated mean of the increments μ_γ. Simulating from this fitted model yields a time series for the future log death rates, which continue the linear trend from the past years but with an added random walk controlled by the estimated variance σ_γ^2. More advanced mortality models can be found in, e.g., Lee and Carter (1992).

Finally, for the birth rates κ_t we used a model which also accounts for seasonality and fertility. As fertility y_t we defined the crude number of births per week κ_t divided by the population aged 20-39 years. We then fitted the following linear model to the weekly fertility rates $\{y_t\}_{t=1,\dots,T}$ from 2002 till 2012

$$\log y_t = \alpha_1^{(y)} + \alpha_2^{(y)} t + \alpha_3^{(y)} \cos\left(\frac{2\pi t}{52}\right) + \alpha_4^{(y)} \sin\left(\frac{2\pi t}{52}\right) + \epsilon_t,$$

where t refers to the week number and the residuals $\{\epsilon_t\}$ follow an autoregressive process of order one, i.e.

$$\epsilon_t = \beta \epsilon_{t-1} + \xi_t, \quad \xi_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_y^2).$$

A posterior mode estimate for $(\boldsymbol{\alpha}^{(y)}, \beta)$ and σ_y under the assumption of flat priors was then obtained by using the R function *arima()*. The log fertility rates y_t for future times $t > T$ were simulated from the fitted model enhanced with an added random walk to account for changing trends.

$$\log y_t = \hat{\alpha}_1^{(y)} + \hat{\alpha}_2^{(y)} t + \hat{\alpha}_3^{(y)} \cos\left(\frac{2\pi t}{52}\right) + \hat{\alpha}_4^{(y)} \sin\left(\frac{2\pi t}{52}\right) + \sum_{k=T+1}^{t} W_k + \epsilon_t, \quad W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_W^2),$$

where $\hat{\alpha}_i^{(y)}$ denotes the estimated coefficients and ϵ_t was simulated from the fitted autoregressive process defined by $(\hat{\beta}, \hat{\sigma}_y)$. The future weekly number of births was then recalculated by multiplying the simulated fertility rates with the momentaneous model population aged 20-39.

Note that the further the transmission model predicts into the future the more those predictions are subject to uncertainty regarding the demographic development. For a graphical representation of the demographic data and simulations from the fitted models see Figure 5.9.
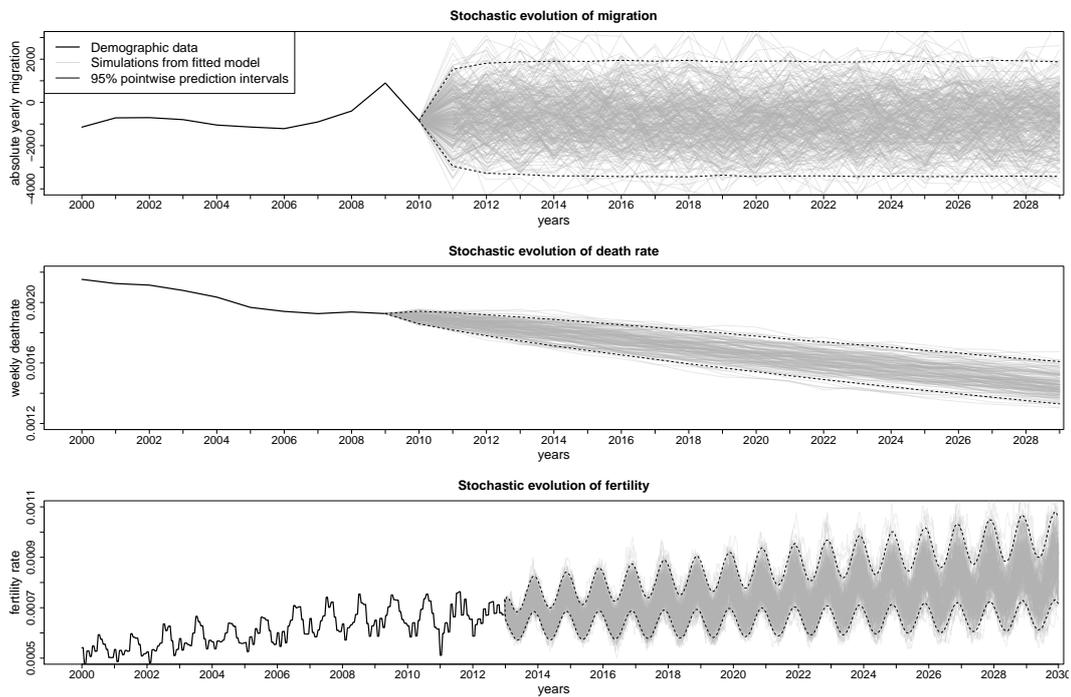
Figure 5.9: Demographic data on yearly migration (top), weekly death rates in the age group >80 years (middle), and weekly fertility rates (bottom) together with simulations from the corresponding time series regression models fitted to that data. For model fitting all available data (black) from 2001 until 2013 was used. For each of the three time series 200 sample paths (grey) were simulated into the future from the corresponding models.

The impact of the additional demographic uncertainty was measured by comparing the 95%-prediction bands from the original incidence sampling with model predictions assuming the stochastic demographic development instead of fixed rates remaining at the respective levels from 2013.

5.3.4 Herd immunity

A major goal of the dynamic modelling approach is to assess the indirect effects of a routine vaccination program, i.e. the additional incidence decrease in certain age groups which is not attributable to the direct protection granted by the vaccine but rather to the decreased risk of infection due to the changed prevalence.

To measure the indirect epidemiological effects it is necessary to know which direct protection can be expected from the vaccination programme, i.e. by which proportion is the original incidence expected to decrease. An approximation for this expected decrease can be obtained using the estimates for the vaccine effectiveness parameter $\boldsymbol{\eta} = (\eta_I, \eta_S, \eta_W)$ regarding protection from symptomatic infection and waning. Since the vaccine induced

immunity wanes over time the expected direct effects differ for various age groups. Therefore, the expected relative incidence decrease was calculated for each age a ($a = 0, \dots, 99$).

The vaccine protection against symptomatic infection corresponding to compartment V_k ($k = 1, \dots, 4$) was calculated according to combined estimated risk ratios $\mathrm{RR}_I^{(k)}$ and $\mathrm{RR}_S^{(k)}$ for acquiring infection and developing symptoms in case of infection compared to being fully susceptible, respectively. As stated in Section 5.1.1 this relative risk was defined by

$$\mathrm{RR}^{(k)} = \mathrm{RR}_I^{(k)} \cdot \mathrm{RR}_S^{(k)} = (\eta_I \eta_S)^{\frac{k}{4}} \,.$$

Since the effectiveness parameters η_I, η_S are to be considered as random variables within the Bayesian framework we defined the estimator $\widehat{\mathrm{RR}}^{(k)}$ as the posterior mean

$$\widehat{\mathrm{RR}}^{(k)} = \mathbb{E}\left[(\eta_I \eta_S)^{\frac{k}{4}} \right],$$

where the expectation is with respect to the posterior distribution corresponding to the best fitting model scenario WV_β. Thus, the estimated compartment-specific relative risks against symptomatic infection are as follows:

$$\widehat{\mathrm{RR}}^{(4)} = 0.041, \quad \widehat{\mathrm{RR}}^{(3)} = 0.085, \quad \widehat{\mathrm{RR}}^{(2)} = 0.175, \quad \widehat{\mathrm{RR}}^{(1)} = 0.414.$$

However, considering the underlying vaccination model and the most likely model scenario WV_β, there is no unique vaccine compartment V_k linked to each age a. Instead individuals of the same age are distributed over the different vaccine compartments V_k ($k = 1, \dots, 4$) while a complete loss of protection is also possible. The exact distribution is determined by the ODE system (5.3) and depends on the waning rate η_W and the time passed since complete vaccine administration at 4 months of age. We assumed that the mean time since vaccine administration for individuals of age a was approximately a years. Assuming that individuals would not leave the vaccine compartments in case of infection, the population distribution over the four vaccine compartments (and eventually S_1) after a years coincides with the probability masses of a Poisson distribution with parameter $52a\eta_W$, since the waning rate can be interpreted as the jump intensity of a Poisson process (counting the number of jumps) and we are interested in the distribution of this process. Thus, the ratio $p_a(V_k)$ of people aged a years being in V_k is given by

$$p_a\left(V_4\right) = \mathbb{P}_{\mathrm{Poiss}}\left(0; 52a\eta_W\right), \quad p_a\left(V_3\right) = \mathbb{P}_{\mathrm{Poiss}}\left(1; 52a\eta_W\right), \quad p_a\left(V_2\right) = \mathbb{P}_{\mathrm{Poiss}}\left(2; 52a\eta_W\right),$$

$$p_a\left(V_1\right) = \mathbb{P}_{\mathrm{Poiss}}\left(3; 52a\eta_W\right), \quad p_a\left(S_1\right) = \mathbb{P}_{\mathrm{Poiss}}\left(\geq 4; 52a\eta_W\right),$$

where $\mathbb{P}_{\mathrm{Poiss}}\left(\cdot; \lambda\right)$ denotes the probability mass function of the Poisson distribution with rate λ and $p_a\left(S_1\right)$ is the probability of the vaccine immunity being completely vanished. Using this age-specific compartment distribution and the compartment-specific relative risks against symptomatic infection one can calculate the overall vaccine protection for a given age a by

$$\widehat{\mathrm{RR}}(a) = \sum_{k=1}^{4} p_a\left(V_k\right) \cdot \widehat{\mathrm{RR}}^{(k)} + p_a\left(S_1\right) \cdot 1.$$

For larger age groups the expected protection against symptomatic infection was averaged over the corresponding age range, e.g. for the age group 5 till 19 years of age

$$\widehat{\text{RR}}_{5-19} = \frac{1}{15} \sum_{a=5}^{19} \widehat{\text{RR}}(a).$$

Finally, this yields expected relative incidence decreases for each age group that was contained in the data as well as in the model:

$$\widehat{\text{RR}}_0 = 0.042, \ \ \widehat{\text{RR}}_1 = 0.075, \ \ \widehat{\text{RR}}_2 = 0.126, \ \ \widehat{\text{RR}}_3 = 0.199, \ \ \widehat{\text{RR}}_4 = 0.286,$$

$$\widehat{\text{RR}}_{5-19} = 0.799, \ \ \widehat{\text{RR}}_{20-39} = 0.998, \ \ \widehat{\text{RR}}_{40-59} = \widehat{\text{RR}}_{60-79} = \widehat{\text{RR}}_{80-99} = 1.$$

Note that these stated relative decreases are based on a complete vaccination coverage. For incomplete coverage the expected decrease $\widehat{\text{RR}}_{(a)}(\phi)$ scales with the vaccine coverage rate ϕ:

$$\widehat{\text{RR}}_{(a)}(\phi) = 1 - \phi \cdot \left(1 - \widehat{\text{RR}}(a)\right).$$

Herd immunity was then calculated as the relative difference, i.e. the ratio, between the model predicted and the expected direct decrease of the annual incidence. The model prediction was measured for the year 2025, 10 years after long term vaccination coverage was achieved, assuming that the annual incidence reached a steady state by then.

5.3.5   Epidemiological results

In the base case we assumed that vaccination coverage for each birth cohort would increase to a level of 90% until the end of 2015 (see Figure 5.10). Considering children aged <5 years, formerly causing 60% of all notified cases, the model predicted an rotavirus incidence decrease to 437 cases per 100,000 population for the year 2020 (95% prediction interval (PI): 283 – 701) whereas the corresponding incidence without vaccination was 2,709 (95% PI: 1677 – 4053) cases per 100,000 yielding an 84% decrease. In the age group 5-59 years, the model predicted incidences of 107 (95% PI: 84 – 135) and 101 (95% PI: 61 – 154) with and without vaccination program, respectively, suggesting a slight increase. In the group of adults aged >60 years, the annual incidence was predicted to increase from 171 (95% PI: 96 – 265) to 199 (95% PI: 158 – 248) cases per 100,000. The resulting overall reduction in rotavirus incidence was predicted at 35%.

The rotavirus seasonality was predicted to shift after adoption of routine rotavirus vaccination as seen in Figure 5.11. According to our model, peak incidence before vaccine introduction occurred during March with week 12 of each year whereas after introduction the maximum incidence was predicted to be reached in week 15. Thereby, the relative decrease was predicted to be much more pronounced from January to April compared to the off-season.

From analysing the effect of demographic uncertainty (Section 5.3.3), we found that the width of the weekly prediction intervals for the reported incidence increased by up to
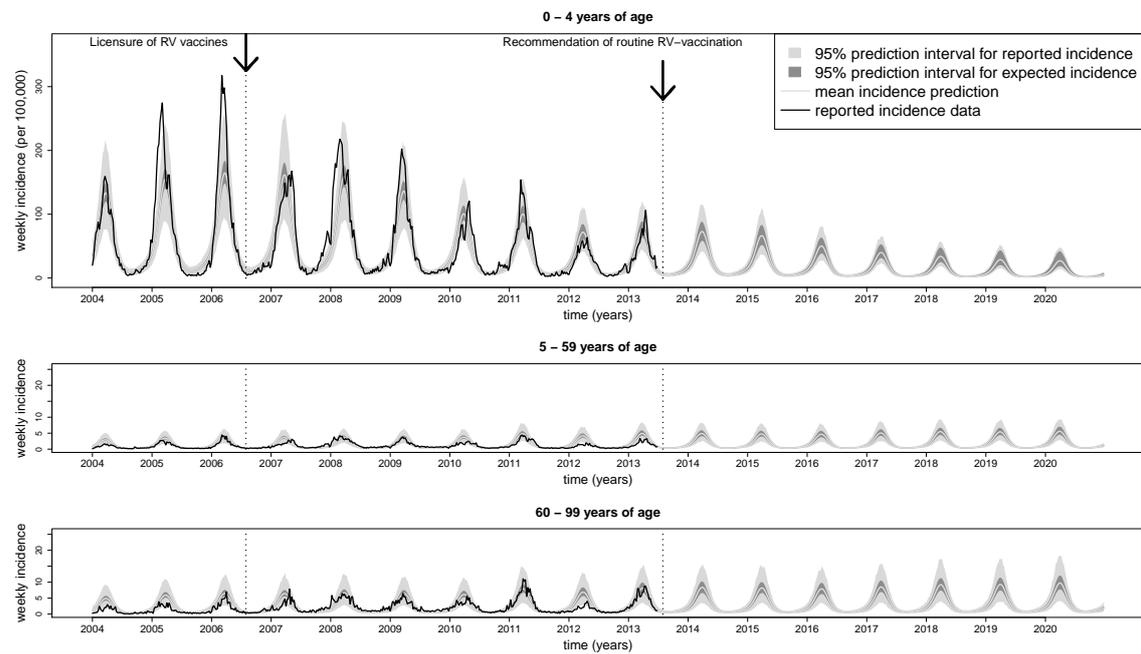
Figure 5.10: Model prediction of the weekly rotavirus incidence in the EFS for the age groups 0-4, 5-59 and 60+ years of age from 2004 to 2020 with introduction of routine rotavirus vaccination in 2013. The dark grey area provides 95% prediction intervals for the expected incidence $\boldsymbol{h} \cdot \boldsymbol{Y}^{(j)}$ incorporating only parameter uncertainty. The light grey area which provides the reported incidence $\boldsymbol{X}^{(j)}$ additionally includes uncertainty from seasonal fluctuations. Notified incidence data up to 2013 is given in black.

4% compared to a scenario with the demographic rates remaining at their respective levels from 2013.

The indirect effects of routine rotavirus vaccination were predicted to be most prominent in young children (Figure 5.12). At a vaccination coverage of 90% herd protection was predicted to prevent 14% of those childhood rotavirus cases remaining when only accounting for direct protection, whereas at 80% coverage herd protection prevents 11% of these cases. In the age-groups 5-59 and >60 years of age indirect effects lead to increased incidences of 14% and 18% above the expected level, respectively.

5.4 Discussion of the results

Within this chapter we estimated rotavirus vaccine effectiveness and predicted the population-level impact of routine rotavirus vaccination in Germany using an ODE-based dynamic model for rotavirus transmission and reporting.

By modelling vaccinated people in separate compartments and applying specific infection risk parameters, as done similarly in Atkins et al. (2012) and de Blasio et al. (2010),
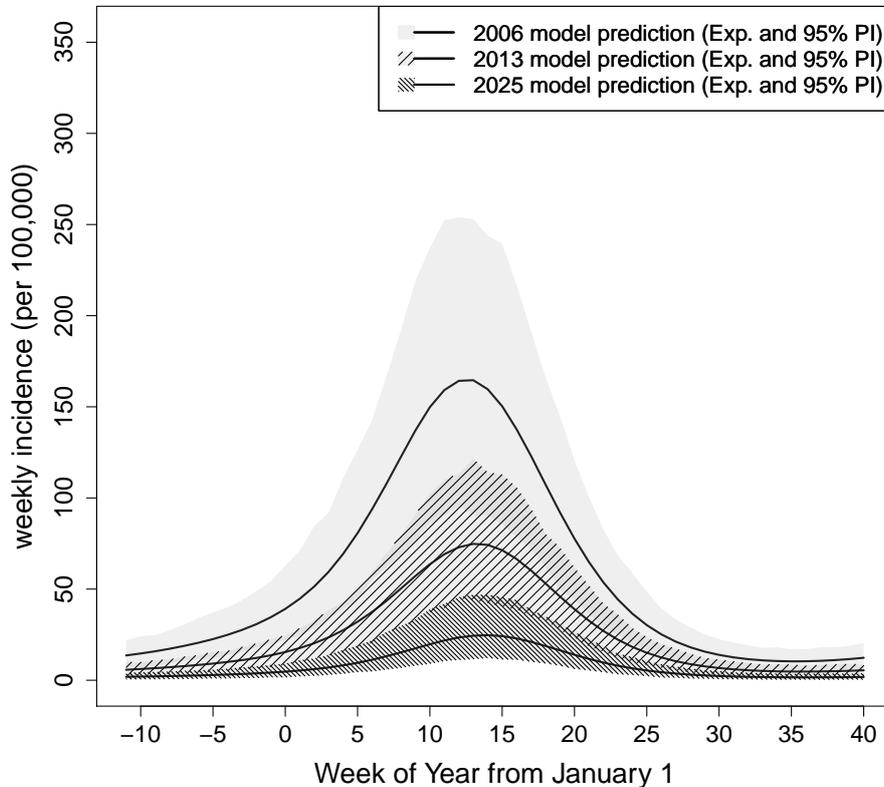
Figure 5.11: Model prediction for weekly reported rotavirus incidence in children <5 years of age in the EFS for three distinct season years (each beginning with october): before vaccine introduction (2006); the most recent season for which data is available (2013); ten years after introduction of routine vaccination with 90% coverage (2025). The x-axis refers to the week of the corresponding year (week 1 denotes the first week of January).

enabled the estimation of vaccine effectiveness via the transmission model. This was performed within a Bayesian inferential framework using adaptive MCMC sampling based on incidence data from the communicable disease reporting system in the EFS, where rotavirus vaccines were already widely used for several years.

Our estimates suggest an initial protection of 96% against symptomatic infection irrespective of its severity for the first three years after complete vaccination. Compared to results from randomized controlled trials (RCT) our estimated VE resembles the efficacy of 91% against acquiring severe rotavirus gastroenteritis (RVGE) but exceeds the efficacy of 74% with respect to the prevention of RVGE of any severity (Koch et al., 2013). A main cause for the discrepancy between our VE estimates and figures from the RCTs may be that the passive surveillance system, where the rotavirus incidence data used for model
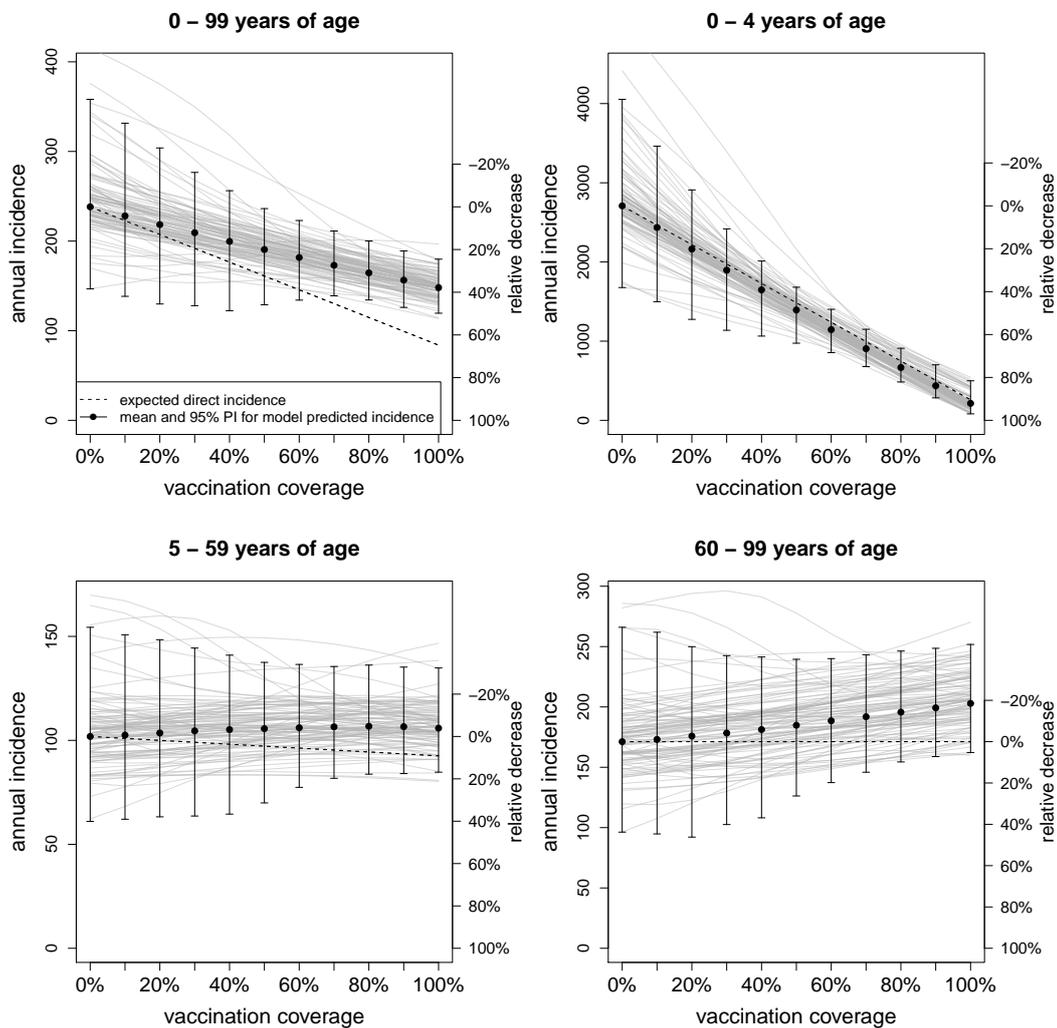
Figure 5.12: Model predictions for age stratified incidence for different vaccine coverage rates together with the expected direct incidence decrease according to the estimated vaccine effectiveness (dashed lines). Solid black lines give the predicted mean and 95% prediction intervals for the annual incidence ten years after introduction of routine vaccination (2025). Grey lines display the predicted incidences along different vaccine coverages for otherwise unchanged demographic processes and parameter vectors sampled from the posterior, which represent possible scenarios of the overall model.

fitting were derived from, primarily captures severe RVGEs while mild RVGE cases either do not seek healthcare or are less frequently notified by physicians, which implies that our estimated VE might in fact rather represent the protection against severe RVGE. Furthermore, a low incidence in the post-vaccination seasons due to disregarded exterior effects or a possible underestimation of the vaccination coverage in the population might also have

led to an overestimation of VE within our study.

Based on our model, the maximum vaccine-induced immunity is estimated to last three years and being fully waned after twelve years. In this regard the model estimate provides novel insight as the true duration of vaccine-induced immunity has yet to be fully assessed in observational studies. Note that an entirely confident assessment of the duration of protection is constrained by our limited observation period of six years. The waning of protection has a considerable impact on the long-term effects of routine rotavirus vaccination as it was found in another modelling approach (Atkins et al., 2012).

Given the estimated parameter posterior distribution our model was well capable of reproducing the incidence data up to 2013 including its seasonal variations. Especially the incidence decrease observed in the age-group <5 years was represented very well in the model as all data from 2009 onwards lied within the pointwise 95% prediction intervals. Prediction intervals for the groups 5–59 and >60 years of age also contained the majority of data points despite the seasonal fluctuations being more pronounced in these higher age-groups. Moreover, by applying data on the WFS population and vaccination coverage, we were able to validate our model structure and parameter estimates.

We found routine vaccination to cause considerable incidence reduction of 84% among young children. This impact prediction is consistent with the observed incidences after introduction of routine rotavirus vaccination in Austria and the US (Payne et al., 2011; Paulke-Korinek et al., 2013). For the higher age groups, however, our model predicted an incidence increase of more than 10% ten years after recommendation, whereas incidence reductions due to herd effects were also observed for those age groups shortly after vaccine introduction (Lopman et al., 2011). More data on the long term incidence development might clarify this discrepancy.

Other modelling studies predicted incidence reductions of 56% (de Blasio et al., 2010), 85% (Pitzer et al., 2009, 2012), 68% (Atkins et al., 2012), and 70% (Atchison et al., 2010) for the age group <5 years. In contrast, our results are less optimistic considering groups >5 years of age, as only the model by Atchison et al. (2010) also predicted an incidence increase. Thus, the potential loss of immunity over time seems to be the primary cause for an increasing incidence among older children and adults, which can be explained by mass vaccination leading to decreased immunity boosting. Our model scenarios analysis showed that models not including this potential immunity loss were not able to explain the EFS incidence data affected by starting vaccination (see Section 5.1.4). Besides a lesser immunity boosting, the increase can also be explained by a shift of childhood infections to older age groups. The corresponding uncertainty intervals are largest for predictions regarding older age groups such that both lower and higher herd protection may also be possible. However, considering the total population higher coverage rates always lead to an overall decreased incidence.

Regarding the applied statistical methodology, we found that the adaptive sampling procedure was able to provide a representative parameter sample from the posterior, while the required number of model evaluations could be kept low. Utilizing the employed algorithms for marginal likelihood estimation offered a suitable tool for identifying the

most likely model scenario among a set of considered models, which could then be used for later incidence predictions.

Altogether, our Bayesian modelling approach provides new insights into key components of rotavirus transmission and vaccine effectiveness. By introducing routine rotavirus vaccination overall rotavirus incidence is expected to be considerably reduced, especially among young children. Within our data-driven approach we took the impact of relevant sources of uncertainty into account: estimation of model parameters, stochasticity in reporting and demographic development. In other epidemiological works these aspects are often neglected or just rudimentarily treated. We think that addressing these issues is an important step towards a more honest assessment of uncertainty in mathematical transmission modelling, in particular with the goal to support health policy decision-making.

Chapter 6

Model implementation in R

This chapter contains documentation of the implementation of our methods for parameter inference and incidence sampling using the employed transmission models described in Chapters 4 and 5. For implementation we used the statistical software environment **R** (R Development Core Team, 2010). All necessary code and data was uploaded on

https://github.com/weidemannf/thesis_code

ready for computation. We present the implemented methods as a collection of files including R-functions, main scripts and data. However, the developed code is not provided within an **R**-package, since we believe that our application and especially the employed ODE-based model is too specific in order to provide a generally applicable program in other contexts. Still, the presented code might give some inspiration into approaches for the computational realization of inference procedures for this particular model class.

Within this section our aim is to provide an overview on the implemented functions, the overall flow-structure of how the main scripts for parameter estimation and incidence sampling integrate these single functions, and the incorporation of necessary data. Additionally, all code is described in detail within the **R**-scripts made available online.

Section 6.1 addresses the **R**-implementation of the inference approach based on asymptotic normality around the posterior mode as presented in Chapter 4 including the functional realization of the underlying ODE system and how it is invoked within the inference procedure. Of particular interest is the flexibility of the inference framework when dealing with variable parameter spaces. In Section 6.2 we will present the implementation of the adaptive Metropolis-Hastings algorithm for Bayesian estimation of the transmission model parameters as applied in Chapter 5.

6.1 Inference based on asymptotic normality accounting for autocorrelated data

The here presented scripts implement the methods presented in Chapter 4 in which the inference procedure was based on posterior approximation incorporating the likelihood adjustment based on the cumulative autocorrelation of the data. The code requires the **R** packages **deSolve** (Soetaert et al., 2010) and **optimx** (Nash and Varadhan, 2011) for evaluating the posterior function and computing its mode, respectively. All functions providing the necessary data on incidence and demographics of the underlying population as well as further required functions are contained within additional script files, e.g. **demographicdata.R** , **functions.R** and **rota.R**. See Table 6.1 for an overview of all implemented functions and the main scripts utilized within Chapter 4.

Table 6.1: Overview of the main scripts and functions regarding parameter inference, model averaging and incidence sampling as presented in Chapter 4.

Name	task	lines of code
Main scripts		
parameterinferenceESS.R	Posterior computation	81
ESScomputation.R	Estimation of cumulative autocorrelation	54
weightcomputation.R	Computation of model probabilities	64
incidencesampling.R	Sampling from model ensemble	115
plotincidence.R	Plotting of sampling results	53
Internal functions		
rota	ODE system	43
paraposteriorESS	Posterior density function	62
loglikelihoodESS	Loglikelihood function for ODE output	24
effectivesamplesize	Cumulative autocorrelation of residuals	71
anscomberesiduals	Anscombe residuals for incidence data	103
optimgrad	Optimization along a vector	7
contactmatrix	Predefined contact matrices	26
Data functions		
initial	Initial condition of ODE system	9
births	Weekly number of births	6
deathrate	Age-stratified death rates	12
ageing	Ageing rates for age groups	3
migration	Age-stratified migration counts	8
EFSdata	Notification data from EFS	9
WFSdata	Notification data from WFS	9

The function **rota** computes the left hand side of the ODE system and the cumulative number of new infections, thus representing the core of the transmission model. After importing the necessary functions a call to **rota** evaluates the ODE-defining equation sys-

tem for a specified model state **N**, model time **t** and model parameters which are provided within a list object **params**. For consistency with the ODE solvers implemented in **deSolve** the output of a call to **rota** must be a list containing one vector of length equal to the number of model states. After rearranging as a $15 \times n_A$ matrix, the output yields the derivatives of each model state component (first 14 rows) and of the cumulative number of infections (last row) for each of the n_A considered age groups (in our case $n_A = 19$), e.g.

```
> LHS <- rota(t=1,N=initial(),
+       params=list('alpha'=c(0.6,0.667),'theta'=c(0.5,0.25,0),
+          'mu'=7/3,'omega'=7/9,'omega0'=1/8,'beta'=1/50,
+          'p'=0.5,'sais'=c(0.2,0.2,0,0),
+          'contact'=contactmatrix(cpara=rep(1,3),pattern=1),
+          'births'=births(),'death'=deathrate(),
+          'mig'=migration(),'ageing'=ageing()))[[1]]
> dim(LHS) <- c(15,n_A)
> LHS[15,]
 [1]  56.57998 101.52175 134.95100 157.39381 170.31788 175.85514
 [7] 495.53931 433.37199 363.76865 297.26943 238.60450 189.30845
[13] 149.19820 117.27841 348.72896 248.63416 193.92024 308.79207
[19]  76.28781
```

The models posterior function **paraposteriorESS** utilizes the ODE system defining function **rota** in order to compute the negative posterior log density of a given parameter vector. Within **paraposteriorESS** the ODE system is solved by applying a Runge-Kutta order 4 scheme using **rk** with **method="rk4"** available in the package **deSolve**, where the solution is stored within the matrix **sol**.

```
> sol <- rk(N, times, rota, params, method = "rk4")
```

Hereby, **N** refers to the initial condition, **times** is the time horizon for which the ODE system has to be solved, **rota** provides the ODE system and **params** is the list of further parameters passed to **rota** as explained above. After some rearranging the ODE solution **sol** is an array of dimension $(T, 15, n_A)$ with T being equal to **length(times)**. The first 14 matrices **sol[,i,]** $(i = 1, \ldots, 14)$ of dimension $T \times n_A$ contain the solution processes of the 14 model compartments stratified by the n_A age groups. The last matrix **sol[,15,]** contains the age-stratified cumulative number of weekly cases occurring over time, from which the model predicted number of cases can be computed as the increments with respect to time. Based on the matrix containing the predicted case numbers for each week and each age group, the loglikelihood of the observed incidence data is evaluated by the function **loglikelihoodESS**, which requires as additional arguments the incidence data **datEFS** or **datWFS**, a detection rate **h** and a relative increase of the detection rate **fact**, the overdispersion of the observational negative binomial distribution **disp**, the effective sample size information **essEFS** or **essWFS** (more on this later), and the region the data belongs to which is either **'EFS'** or **'WFS'** (eastern of western federal states).

```
> sol      <- sol[times,-1]
> dim(sol) <- c(length(times),15,n_A)
> inci     <- sol[-1,15,1:n_A]-sol[-length(times),15,1:n_A]
> logLLEFS <- loglikelihoodESS(inci[dattimes,1:n_A],datEFS,
+ h[1],fact[1],disp,essEFS,region='EFS')
> logLLWFS <- loglikelihoodESS(inci[dattimes,1:n_A],datWFS,
+ h[2],fact[2],disp,essWFS,region='WFS')
```

Here, **dattimes** denotes the vector of time points corresponding to the weeks for which data is available. Within the function **paraposteriorESS** the log prior density is then further evaluated based on prior distribution parameters stored externally, such that the final output is the (negative) unnormalized log posterior density which is the (negative) sum of loglikelihood and log prior density evaluated for the parameters initially passed to **paraposteriorESS**.

Thereby, the posterior function uses two arguments as input, one being a list object **infolist** containing all parameters and data which are required for computing the ODE solution and the corresponding likelihood of the incidence data, e.g. among others the initial state **N**, the time horizon **times**, model parameters and data **datEFS** and **dataWFS**. The second argument is the vector of parameters **para** for which the prior distribution is to be evaluated, i.e. these are considered to be the variable parameters of the model and are thus passed twice to the function **paraposteriorESS** since they are already contained in **infolist**. Hereby, parameters contained in **para** overwrite respective values from **infolist**. For instance, after defining a suitable list object **infolist**, a call to **paraposteriorESS** yields different results depending on which parameters and respective values are included in **para**. which might also be empty.

```
> paraposteriorESS(info=infolist)
[1] 656434.3
> paraposteriorESS(para=c('mu'=0),info=infolist)
[1] 495476.6
> paraposteriorESS(para=c('mu'=0,'omega'=0),info=infolist)
[1] 1088821
```

In this example the components **'mu'** and **'omega'** control the duration of the infectious periods (on a log scale).

The reasoning behind this construction is that the optimisation methods implemented in the **R**-package **optimx** require the target function to have a numerical vector as its first argument, which is then subject of variation. Since one part of our modelling was to investigate the effect of fixing certain model parameters within inference, we wanted the posterior function to be allowing different parameter space, i.e. different parameter vectors which are subject of estimation. Therefore, for conducting parameter inference for different models one may still use the same posterior function **paraposteriorESS**. Thus, the variable parameters can be passed within the vector **para** whereas all fixed parameters together with the demographic and incidence data are passed within **infolist**.

Because one feature of our inference approach is to account for the autocorrelation structure in the data by using the measure of the cumulative autocorrelation (or effective sample size), the posterior function scales the loglikelihood using the information on the effective sample size provided in **infolist** through the vectors **infolist[['essEFS']]** and correspondingly **infolist[['essWFS']]**. These contain the scaling factors $1/\mathrm{CA}^{(j,i)}$ applied to the loglikelihood of the ten age-specific data series ($j = 1, \ldots, 10$) from the EFS and WFS ($i = e, w$), respectively, as explained in Section 4.2.2. For instance, changing the scaling parameters from 1 to 0.5 for each age group in the EFS yields a different posterior density value.

```
> infolist[['essEFS']] <- rep(0.5,10)
> paraposteriorESS(info=infolist)
[1] 474250.3
```
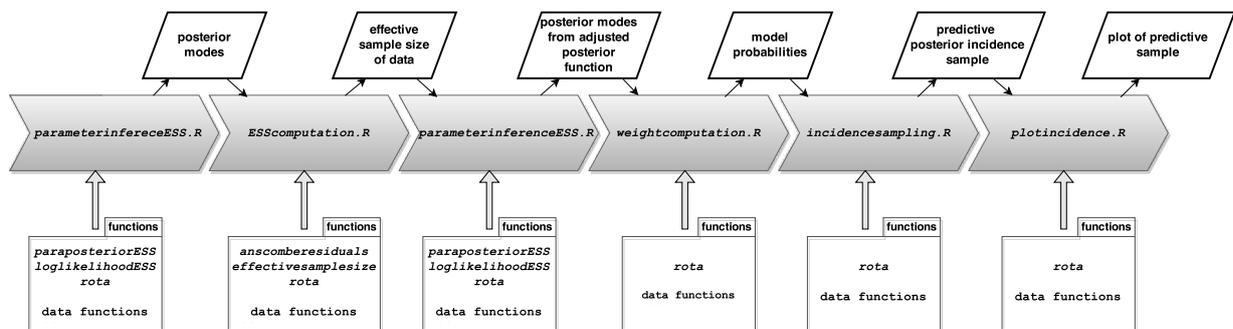

Figure 6.1: Work flow and corresponding scripts for posterior mode based inference procedure, model averaging, predictive incidence sampling and results plotting as performed within Chapter 4.

With the full statistical model being implemented as it is given through the posterior function **paraposteriorESS**, the overall work flow for inference, predictive sampling and results plotting is given by Figure 6.1. Each of the main tasks are conducted within separate **R**-scripts (see also Table 6.1), which are further described in the following.

6.1.1 Optimization procedure for posterior mode computation

The main script for conducting parameter inference based on the Gaussian approximation of the posterior distribution is **parameterinferenceESS.R**. The aim of this script is to find the posterior mode according to a specified parameter vector being subject of estimation and to compute the observed Fisher information matrix as described in Section 4.2.2. Recall, that the employed procedure is a two-step method where the posterior function was either unadjusted in the first step or scaled by the effective sample size of the data in the second step. Both steps are executed within **parameterinferenceESS.R**. To do so,

after fetching the necessary data, the script checks whether there is already information on the effective sample sizes available within a data file **meaness.Rdata** and, if so, applies these values or otherwise sets the scaling factors to 1, i.e.

```
> essdata.ia <- try(load(file="meaness.Rdata"))
> if (class(essdata.ia)=='character'){
+ essEFS<-meanessEFS; essWFS<-meanessWFS
+ }else{
+ essEFS<-essWFS<-rep(1,10)}
```

Hereby, the file **"meaness.Rdata"**, if available, contains the vectors **meanessEFS** and **meanessWFS** and was generated by the script **ESScomputation.R**, which in an intermediate step calculates and averages the effective sample size for each of the 20 data series over the 18 considered model as described in Section 4.2.2. To compute the effective sample size for one time series, the script applies the function **anscomberesiduals** which normalizes the observed residuals of the data with respect to the model predicted expectation according to the posterior mode using an Anscombe transformation. Then for each residual time series the cumulative autocorrelation as a measure for the effective sample size is computed by the function **effectivesamplesize**. Finally, the averaged cumulative autocorrelation estimates for each age group and region are stored within the vectors **meanessEFS** and **meanessWFS** to be used for scaling of the loglikelihood in the second optimisation step.

However, the remaining optimisation procedure is the same for both steps. After fetching the necessary functions and defining the list object **infolist** of fixed parameters, data and possibly information on the effective sample size, the parameter vector to estimate is defined within the script, e.g. by

```
> para <- c('omega0'=omega0,'beta'=beta,'p'=p,'a1'=sais[1],
+         'b1'=sais[2],'a2'=sais[3],'b2'=sais[4],
+         'contactpara'=contactpara,'h'=h,'fact'=fact,
+         'disp'=disp)
```

This vector **para** contains the model parameters which are variable within the posterior optimisation whereas other parameters contained in **infolist** remain fixed. The optimisation procedure is based on an alternating sequence of the Nelder-Mead algorithm (Press et al., 2007) within the function **optim** (package **optimx**) and intermediate minimisation along the gradient using the function **optimgrad**, with each algorithm being applied K times.

```
> for (i in 1:K){
+   optpara   <- optim(para,paraposteriorESS,
+         method='NelderMead',info=infolist,
+         control=list('trace'=1,maxit=2000,reltol=1e-15))
+   para      <- optpara$par
+   pointpost <- paraposteriorESS(para,info=infolist)
```

```
+   grad.vec  <- grad(paraposteriorESS,para,method='Richardson',
+       method.args=list(r=6),info=infolist)
+   grad.vec  <- grad.vec/(10000*mean(abs(grad.vec)))
+   opt.val   <- optimgrad(paraposteriorESS,para,
+    grad.vec,info=infolist)
+   para      <- para-(opt.val[[1]]*grad.vec)
+ }
```

The reason for the alternating methods is that the Nelder-Mead algorithm, which does not require any information on the gradients, generally constitutes a suitable choice optimising ODE-based models since gradient computation is often tedious for those models. However, the algorithm might get stuck or its searching area might become to small such that intermediate optimisation steps along the gradient improve the overall optimisation performance.

In a final step, the Hessian of the negative log posterior is computed for later assessment of uncertainty based on the observed Fisher information. Computation of the gradient and the Hessian is performed by functions from the package **numDeriv** (Gilbert and Varadhan, 2012).

```
> hess <- hessian(paraposteriorESS, para,method="Richardson",
+              method.args=list(r=6),info=infolist);
```

The results on the optimal parameter vector **para**, i.e. the posterior mode, together with its corresponding Hessian **hess** are stored for further processing as these provide the parametrisation of the approximate Gaussian posterior distribution.

6.1.2   Predictive incidence sampling

What we obtain from the inference procedure are the posterior mode and the observed Fisher information matrix evaluated at the posterior mode for each of the 18 considered models. In order to generate a predictive incidence sample from this model ensemble the script **weightcomputation.R** first calculates the corresponding model probabilities based on the marginal likelihoods as described in Section 4.2.3 and stores these within **weights.Rdata**.

Sampling from the predictive distribution of the incidence according to the ensemble of models is done by **incidencesampling.R** which performs the algorithm presented in Section 4.2.4. Subject to the model probabilities fetched from **weights.Rdata** the script **incidencesampling.R** iteratively samples a model which determines a mean parameter vector and its respective covariance matrix. The algorithm then samples from the corresponding multivariate normal distribution utilizing the package **mnormt** (Azzalini and Genz, 2014).

```
> model <- sample(seq_len(n_Models),size=1,prob=totalweights)
```

```
> para  <- c(rmnorm(n = 1, mean = paralist[[model]],
+  varcov=covarlist[[model]]))
> names(para) <- names(paralist[[model]])
```

Here, **paralist** and **covarlist** are lists containing the posterior modes and covariance matrices for the 18 considered model presented in Section 4.2.3. For the sampled parameter vector **para** the script then computes the corresponding solution to the ODE system and stores the resulting expected reported number of cases occurring in the $n_D = 10$ data age groups for each week from 2001-2008 within the $n_D \times (8 \cdot 52)$ matrices **incistrataEFS** and **incistrataWFS** for the regions of EFS ans WFS, respectively. Random samples **sampleinciEFS** and **sampleinciWFS** for the observed reported number of cases are then generated by sampling from the negative binomial distribution with mean given by **incistrataEFS** and **incistrataWFS**, respectively, and dispersion **disp** contained in the parameter vector **para**, e.g.

```
> sampleinciEFS  <- rnbinom(n=n_D*length(lastyears),
+                   mu=incistrataEFS,size=para[['disp']])
```

for the observed incidence in the EFS and analogously for the WFS. This procedure is repeated for a pre-specified number J of model samples. The final results are stored within the file **incidencesample.Rdata** consisting of the four $(n_D, 8 \cdot 52, J)$-arrays **aggreginciEFS**, **aggreginciABWFS**, **aggregsampleEFS** and **aggregsampleWFS** which contain the samples for the expected and observed number of reported cases in the EFS and WFS, respectively.

6.1.3 Plotting of the incidence sample

Plotting of the predictive sample is performed by the script **plotincidence.R**, which utilizes the previously generated incidence samples stored in **incidencesample.Rdata** to calculate pointwise quantiles of the predictive distribution for the expected and observed weekly number of reported cases in each age group. Hereby, the plotting and thus the computation of quantiles is not necessarily done for each of the ten age groups available in the data, but for user specified pooled age groups as determined by a list **agestrata** which defines how many and which age groups should be plotted and how the aggregation should be executed, e.g.

```
> load(file="incidencesample.Rdata")
> agestrata=list(1:2,3:5)
> plotyears <- seq_len(dim(aggreginciWFS)[2])
> probs <- c(0.975,0.025)
> for (j in 1:length(agestrata)){
+   for (i in 1:length(plotyears)){
+     sampleWFS.quantiles[j,i,]  <- quantile(colSums(
```
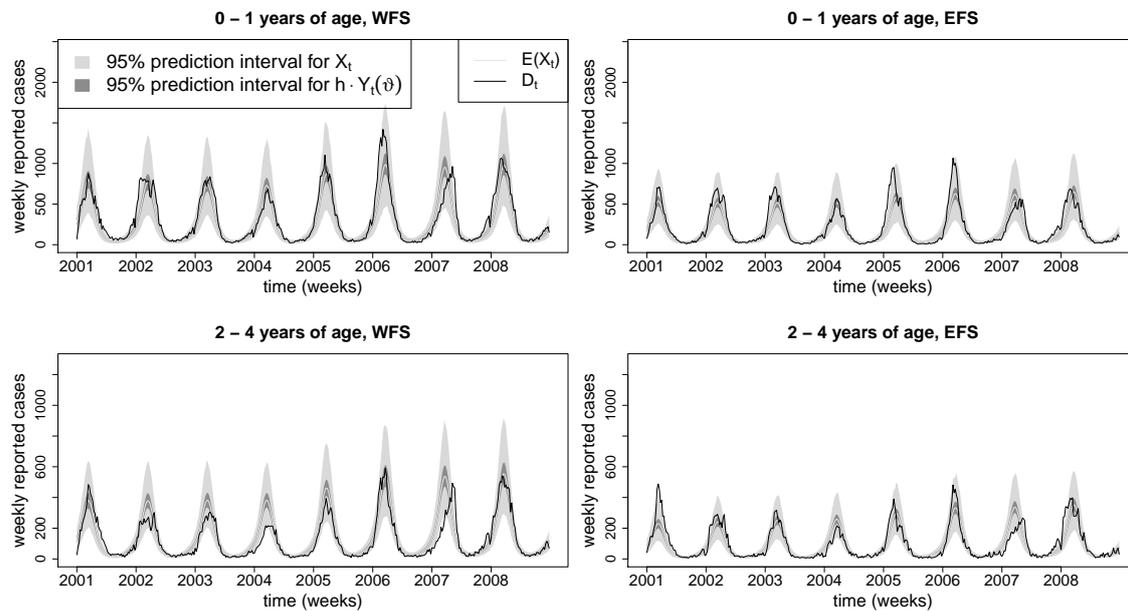
Figure 6.2: Predictive incidence plot generated by the script.

```
+               aggregsampleWFS[agestrata[[j]],i,]),
+               probs=probs,na.rm=TRUE)
+    meanWFS.quantiles[j,i,]  <- quantile(colSums(
+               aggreginciWFS[agestrata[[j]],i,]),
+               probs=probs,na.rm=TRUE)
+    meanWFS[j,i]    <- mean(colSums(
+               aggreginciWFS[agestrata[[j]],i,]),na.rm=TRUE)
+  }
+  D_WFS[j,] <- colSums(datWFS[agestrata[[j]],])
+ }
```

Here, **sampleWFS.quantiles** contains the predictive 0.975- and 0.025-quantiles for the weekly observed number of cases in the aggregated age groups, **meanWFS.quantiles** contains the corresponding quantiles for the expected number of observed cases (only considering parameter uncertainty) and **meanWFS** contains the predictive mean for the observed cases stratified by age group and week. For comparison the observed data is aggregated within **D_WFS**. Corresponding matrices and arrays are also computed for the EFS.

The final plot is then generated by first setting up the plot window, then drawing the prediction bands based on the computed pointwise quantiles, then drawing the prediction means and finally the observed data. Thus, setting **agestrata** as above, one obtains the plot from Figure 6.2.

6.2 Adaptive Metropolis-Hastings and marginal likelihood estimation

For the modelling of vaccination impact as described in Chapter 5 the employed transmission model had to be additionally augmented by mechanisms of vaccination. Here, we illustrate the implemented methods for parameter estimation, model selection and predictive incidence sampling using the extended ODE-based model. Table 6.2 provides an overview of the main scripts and utilized functions. Recall, that the extended model was applied to the population from EFS such that only data on demographics and incidence from the EFS was used within this modelling approach.

Table 6.2: Overview of the main scripts and functions regarding parameter inference, model selection and incidence sampling based on the vaccination-transmission model presented in Chapter 5.

Name	task	lines of code
Main scripts		
adaptiveMCMCinference.R	Posterior sampling	105
marginallikelihood.R	Computation of model probabilities	71
incidencesampling.R	Sampling from model ensemble	81
plotincidence.R	Plotting of sampling results	120
Internal functions		
rotavacc	ODE system	57
logprior	Prior density function	12
likelihoodEFS	Likelihood for specific parameters	23
loglikelihoodESS	Loglikelihood function for ODE output	12
vaccoverage	Loglikelihood function for ODE output	8
contactmatrix	Predefined contact matrices	26
Data functions		
initialEFS	Initial condition of ODE system	9
birthsEFS	Number of births in EFS	7
deathrate	Age-stratified death rates	12
ageing	Ageing rates for age groups	3
migrationEFS	Age-stratified migration counts	17
EFSdata	Notification data from EFS	9
Demographic projection		
projectdeath	Projected age-stratified deathrates	11
projectmigra	Projected migration counts	10
weeklyfertility	Projected fertility rates	8

For estimation of rotavirus vaccine effectiveness the compartmental ODE model was additionally augmented by compartments of vaccinated people. The resulting ODE system is implemented in the function *rotavacc* which again represents the core within all analyses.

Compared to the original ODE system provided by **rota** this new function requires as additional arguments parameters regarding vaccine efficacy, weekly vaccination coverage rates, and age specific susceptibilities (see Section 5.1.1). The output of **rotavacc** is a $19 \times n_A$ matrix containing the derivatives of the now 18 model states and the cumulative number of cases, each stratified by $n_A = 19$ age groups.

The function **likelihoodEFS** computing the loglikelihood of the data requires the package **deSolve** for solving the ODE system. Note also, that the necessary information list again contains an argument **infolist$essEFS** for the effective sample size of the data time series to accordingly adjust the likelihood. For this, the script fetches the corresponding results contained in **meanESS.Rdata** computed previously as described in Section 6.1. Furthermore, in contrast to the posterior function **paraposteriorESS** from the last section, all arguments within **likelihoodEFS** must be specified.

The work flow for posterior sampling, marginal likelihood estimation, predictive sampling and results plotting is displayed in Figure 6.3. The **R**-script performing the posterior sampling is explained in more detail in the following.
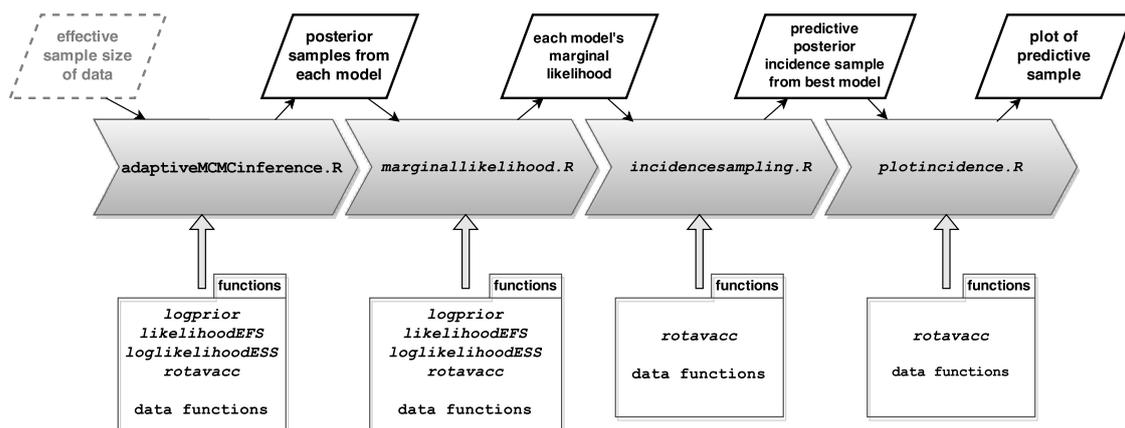

Figure 6.3: Work flow and corresponding scripts for MCMC-based posterior sampling, marginal likelihood estimation, predictive incidence sampling and results plotting as performed within Chapter 5.

6.2.1   Parameter inference using adaptive Metropolis-Hastings

The adaptive MCMC-algorithm generates a sample from the posterior function defined on the 29-dimensional parameter space consisting of the four parameter blocks **modelpara**, **vaccpara**, **coverpara** and **immupara**. The sampling procedure is performed within the script **adaptiveMCMCsampling.R**. After loading the necessary functions, packages and previous results, the script defines the log prior density **logprior** for the parameter vector to be estimated, which as for the log posterior has to be defined up to an additive constant only. The prior information for the first block **modelpara**, which is the same as used in

Section 6.1, is again fetched from an external file. In contrast, the prior means **para2**,..., **para4** and covariance matrices **covar2**,..., **covar4** of the three remaining (possibly transformed) parameter blocks are defined explicitly in the script.

```
> para2  <- c(-0.147,-0.147,log(1/52))
> covar2 <- diag(c(0.871,0.871,0.5))
> para3  <- rep(0,8)
> covar3 <- diag(,nrow=8)
> para4  <- c(0,0,0);
> covar4 <- diag(c(1,0.01,0.0001))
```

Since in the inference procedure all parameters are treated as one joint vector, we defined corresponding index vectors **block1**,..., **block4** to assign the parameter vector components to the single parameter blocks.

```
> parDims <- c(length(para1),length(para2),length(para3),length(para4))
> block1 <- 1:parDims[1]
> block2 <- (cumsum(parDims)[1]+1):(cumsum(parDims)[2])
> block3 <- (cumsum(parDims)[2]+1):(cumsum(parDims)[3])
> block4 <- (cumsum(parDims)[3]+1):(cumsum(parDims)[4])
> dimension <- sum(parDims)
```

Here, **para1** refers to the posterior mode regarding the transmission parameters from the modelling without vaccination.

To initialize the sample chain, i.e. to set initial parameter vectors **steppara1**, ..., **steppara4** and a joint proposal covariance matrix **transcovar**, the script checks if there is already a sample from the posterior available, stored within a file **adaptivemcmcresults.Rdata**. If so, the initial quantities are set based on this sample, otherwise the initial parameter vector is set to the prior mode and the proposal covariance is set to a downscaled version of the prior covariance using the **bdiag** function from the package **Matrix** (Bates and Maechler, 2014).

```
> sample.ia <- try(load(file='adaptivemcmcresults.Rdata'))
> if (class(sample.ia)=="character"){
+   steppara1 <- paramatrix1[,length(paramatrix1[1,])]
+   names(steppara1)<-names(para1)
+   steppara2 <- paramatrix2[,length(paramatrix1[1,])]
+   steppara3 <- paramatrix3[,length(paramatrix1[1,])]
+   steppara4 <- paramatrix4[,length(paramatrix1[1,])]
+   transcovar <- cov(t(rbind(paramatrix1,paramatrix2,
+                    paramatrix3,paramatrix4)))
+ }else{
+   optpara1 <- optim(para1,logprior,vaccpara=para2,
+         coverpara=para3,immupara=para4,
```

```
+          method="BFGS",control=list(fnscale=-1))$par
+   steppara1 <- optpara1; names(steppara1)<-names(para1);
+   steppara2 <- para2
+   steppara3 <- para3
+   steppara4 <- para4
+   hess    <- hessian(logprior,modelpara=optpara1,
+          vaccpara=para2,coverpara=para3,immupara=para4)
+   transcovar <- as.matrix(bdiag(-solve(hess),covar2,covar3,
+    covar4))/10000
+}
```

The script then computes the log posterior **stepLP** at the initial parameter and sets up
storage matrices **paramatrix1**, ..., **paramatrix4** and **LPsample** for the chain of parameter
vectors for each block and the corresponding log posterior values. After defining the overall
chain length **K** to be sampled and the length **J** of the initial period, the posterior sampling
procedure according to algorithm 2 from Section 3.2.3 is ready to go.

The adaptive step happens within the first part of the loop, where after the initial period
has passed the proposal covariance matrix **transcovar** is recalculated based on the so far
computed posterior sample. Using this proposal covariance the algorithm samples a new
parameter candidate which is then accepted as a new chain value with probability based
on its log posterior value **canLP** compared to the former log posterior value **stepLP**. For
drawing the multivariate chain step, the algorithm requires the package **mnormt** (Azzalini
and Genz, 2014).

```
> for (i in 2:K){
+   if (i>J){
+     transcovar <-  cov(t(rbind(paramatrix1,paramatrix2,
+                     paramatrix3,paramatrix4)[,1:(i-1)]))
+   }
+   increment <-  rmnorm(n = 1, mean = rep(0,dimension),
+               varcov=varscale*transcovar)
+   canpara1 <- steppara1 + increment[block1]
+   names(canpara1) <- names(para1)
+   canpara2 <- steppara2 + increment[block2]
+   canpara3 <- steppara3 + increment[block3]
+   canpara4 <- steppara4 + increment[block4]
+   canLL <- likelihoodEFS(modelpara=canpara1,vaccpara=canpara2,
+       coverpara=canpara3,immupara=canpara4,info=infolist)
+   canLP <- canLL + logprior(modelpara=canpara1,
+       vaccpara=canpara2, coverpara=canpara3,
+       immupara=canpara4)
+   if (!is.nan(canLP)){
+     if(runif(1) < exp(canLP-stepLP)){
```

```
+      steppara1 <- canpara1
+      steppara2 <- canpara2
+      steppara3 <- canpara3
+      steppara4 <- canpara4
+      stepLP   <- canLP
+    }
+  }
+  paramatrix1[,i] <- steppara1
+  paramatrix2[,i] <- steppara2
+  paramatrix3[,i] <- steppara3
+  paramatrix4[,i] <- steppara4
+  LPsample[i]    <- stepLP
+}
```

Within each iteration of the chain, for each block the current parameter vector and its log posterior value are stored within the predefined parameter matrices. After conclusion of the algorithm, the final sample is stored within the data file **adaptivemcmcresults.Rdata**.

As stated, the adaptive sampling algorithm is directly implemented within the script **adaptiveMCMCsampling.R**. However, more generally applicable software for ODE models is provided, e.g., within the package **fme** (Soetaert and Petzoldt, 2010) which also includes the here presented adaptive MCMC sampler.

6.2.2   Marginal likelihood estimation and incidence sampling

The estimation of the marginal likelihood is performed by the script **marginallikelihood.R** which implements Algorithm 4 proposed in Section 3.3.3. The script first fetches the necessary packages and data required by the log prior density **logprior** and the log likelihood function **likelihoodEFS** as well as a previously computed sample from the posterior distribution stored as **paramatrix1**,…, **paramatrix4** together with the respective (unnormalized) log posterior values **LPsample**. After defining the algorithm specification, i.e. the number of estimator evaluations **N** and the respective number newly proposed parameters **K**, the script computes **N** estimates for the log marginal likelihood **margLL** of the data according to the given model and eventually computes their mean **mean(margLL)** as its final output.

The sampling from the predictive incidence distribution is performed similarly as described in Section 6.1.2 and is implemented in **incidencesampling.R**. The main difference is that instead of drawing from an Gaussian approximation of the posterior distribution, here the sampling algorithm draws parameter vectors from the posterior sample stored in **adaptivemcmcsample.Rdata**. Additionally, a vector of long term vaccination coverage levels to be investigated **coverage.levels** has to be specified whereas demographic processes for the future progress of birth, death and migration rates are sampled from the respective functions **weeklyfertility**, **projectdeath** and **projectmigra**. The final samples consisting

of arrays for the age and time stratified expected and observed cases as well as the sampled population cases are stored within **_Rdata_** files which can be further progressed for results plotting, e.g. as displayed in Figure 5.10 on page 155.

Chapter 7

Conclusion

This thesis dealt with predicting the epidemiological effects of routine vaccination programmes. Our topic was initiated through the possible implementation of routine infant vaccination against rotavirus in Germany. The vaccination impact was assessed through a compartmental dynamic transmission model based on ordinary differential equations. In applied disease transmission modelling the standard approach is to fix most model parameters according to best available knowledge. However, in the present analysis model structure and parameters were inferred within a Bayesian framework incorporating prior epidemiological knowledge and routinely collected disease surveillance data. These data consisted of weekly aggregated rotavirus case counts which were provided by the Robert Koch Institute and stratified by age and region. Resulting structural and parametric uncertainty was taken into account when making model based predictions for the future epidemiological consequences of introducing routine vaccination.

The epidemiological results regarding the predicted rotavirus vaccination impact as well as parts of the methodological innovations were published within

Weidemann, F., M. Dehnert, J. Koch, O. Wichmann, and M. Höhle (2014). Bayesian parameter inference for dynamic infectious disease modelling: rotavirus in Germany. *Statistics in Medicine 33* (9), 1580–1599.

and

Weidemann, F., M. Dehnert, J. Koch, O. Wichmann, and M. Höhle (2014). Modelling the epidemiological impact of rotavirus vaccination in Germany – A Bayesian approach. *Vaccine 32* (40), 5250 – 5257.

In this concluding chapter we give a summary of these epidemiological outcomes together with the statistical methods which were applied or developed in order to obtain those results. Furthermore we provide starting points for future research.

7.1   Summary of the epidemiological results

An ODE-based SIS-type model was developed in order to represent the transmission dynamics of rotavirus disease within the German population (see Chapter 4). The model accounts for population-specific aspects such as the age distribution in Germany and a heterogeneous contact behaviour, but also for rotavirus specific characteristics like the possibility of both symptomatic and asymptomatic infections as well as an incremental immunity increase after each infection which also allows for multiple episodes of rotavirus illness per individual. Additionally, a stochastic observational component connects the unobserved transmission model dynamics with the available rotavirus case notification data. Model parameter estimates were obtained via a Bayesian inference framework and provide insight into some epidemiological aspects that are otherwise difficult to assess through clinical or observational studies, e.g. the degree of case underdetection in eastern and western federal states or the importance of asymptomatic infections for the transmission process. The most likely contact matrix – as determined by Bayesian model selection procedures – showed similarities to the contact patterns assessed within POLYMOD (Mossong et al., 2008), a European wide study on the frequency of social contacts. Considering the match with the observed rotavirus incidence data prior to increased vaccine administration beginning in 2009, the model was well capable of reproducing the annual mean incidence curve within each age group and region and also captures the inter-seasonal variation when assuming a negative binomial distribution of the observed case counts (Figure 4.3).

As a consequence of introducing routine rotavirus vaccination for infants in 2013, the transmission model predicted a substantial decrease of incidence among children less than five years of age whereas the incidence among older children, adults and elderly is likely to increase due to a potential age shift and other indirect effects (see Chapter 5). To obtain these results the transmission model was augmented with vaccination compartments and respective vaccine efficacies protecting vaccinated individuals against infection and development of symptoms. Estimating the vaccine effectiveness on the basis of time series data also including post-vaccination incidence data from 2009 till 2013 suggests, that vaccination yields high protection against acquiring infection and additional moderate protection against a symptomatic course (Figure 5.7). We detected that additionally accounting for age-specific natural immunity is necessary to explain the vaccination-related observed incidence development, especially among elderly. This was achieved by applying model selection criteria based on the marginal likelihood of the data. Although the incidence reduction among young children due to routine vaccination is predicted to be partially compensated by incidence increases in other age groups, the main epidemiological goal of removing the disease burden from the age group younger than five years – where rotavirus infection frequently leads to a severe course of illness requiring hospitalisation – is ensured (Figure 5.12). Within the predictions, the uncertainty regarding the epidemiological impact is most pronounced in the older age groups which are only indirectly affected by the vaccination. Thus, also an incidence decrease among adults and elderly presents a plausible

scenario.

7.2 Summary of the applied statistical methods

For inferring parameters of the complex ODE-based transmission model presented in Chapter 4 the posterior distribution of the model parameters is approximated through a multivariate Gaussian distribution, which is motivated by asymptotic properties in the case of many available data. This approach circumvents laborious inference based on sampling procedures and only requires the computation of the posterior mode. The Gaussian approximation provides a satisfying match with the true posterior when comparing corresponding single-parameter credibility intervals and two-parameter joint credibility regions (Figure 4.8). Marginal posteriors and the marginal likelihood of the data necessary for model averaging is immediately given due to the well known analytical properties of the Gaussian distribution. However, although calculation of the posterior mode appears to be a straightforward task using numerical optimisation, the overall procedure requires many thousands evaluations of the model likelihood and thus does not provide the fast posterior computation method as initially expected.

From a methodological perspective one of the major contributions of this thesis is the likelihood adjustment in the posterior based on the autocorrelation structure within the data (Chapter 2.2). By employing a power posterior distribution (see e.g. Friel and Pettitt (2008)) the approach suggests a down-scaling of the likelihood function when the likelihood is defined by, possibly erroneously, assuming the data points to be independent observations given the model. The scaling factor is hereby defined such that the impact of the data is reduced subject to their effective sample size, which can be measured through the cumulative residual autocorrelation of the data time series given the fitted model. It was shown that for the estimation of the mean of simple autoregressive processes this adjustment procedure yields equivalent inference results compared to directly modelling the dependency of subsequent observations into the likelihood function. Moreover, through simulation studies we could demonstrate that also in a more complex setting of an ODE-based transmission model, the proposed method yields an accurate assessment of parameter estimates and their respective uncertainty, while being very flexible with respect to the actually unknown dependency structure of the data. Thus, for inference of dynamic disease transmission models, the new adjustment approach provides a suitable procedure in order to address conditionally dependent observations while circumventing additional model complexity.

For the extended transmission model including vaccination mechanisms from Chapter 5 we left the asymptotic Gaussian framework – inference for the vaccine effectiveness and other model parameters was performed by an adaptive Metropolis-Hastings MCMC sampling algorithm (see Chapter 3.2). Whereas for an efficient sampling the classical MCMC-based procedures require the proposal distribution to capture the inherent model parameter correlation, the adaptive algorithm is able to improve its proposal distribution on the fly. By doing so, the self-learning algorithm optimizes its mixing behaviour in order

to reduce the total number of required model evaluations for generating a sufficiently large sample. For Bayesian inference of computationally expensive ODE-based transmission models, the adaptive algorithm thus offers a practical tool for efficient posterior sampling.

In order to identify necessary model components of the transmission model with vaccination, model selection was performed by estimating the marginal likelihood of the data for each considered model (Chapter 3.3). Since the marginal likelihood can not be computed analytically for models based on nonlinear ODEs, it is assessed via a newly enhanced algorithm for pointwise estimation of the normalized posterior density based on an existing MCMC sample output. This estimation approach – which was inspired by Chib and Jeliazkov (2001) – utilizes an expectation ratio identity derived from the detailed-balance condition of the preceding MCMC run, which can then be estimated by using Monte-Carlo methods. The methodological innovation presented in this thesis suggests to estimate an expectation ratio from a detailed-balance condition according to an alternative proposal distribution, although its corresponding Metropolis-Hastings algorithm may differ from that one that was used for generating the available posterior sample. Within a Gaussian posterior setting we have shown, that proposal distributions which are optimal with respect to MCMC sampling do not necessarily yield optimal results for the respective marginal likelihood estimation algorithm. We thus derived, that minimal variance and bias of the marginal likelihood estimation can be achieved by choosing a state-independent proposal density which approximately mimics the posterior. Results from a simulation-based analysis suggest (3.3.4), that the modified algorithm using proposal functions defined via an existing posterior sample can considerably improve accuracy of the marginal likelihood estimation while also requiring fewer additional model evaluations. The improved performance was stable over a wide range of model dimensions and possible shapes of the posterior density.

Regarding the general applicability of the here presented methods, the class of ODE-based models is widely used, not only in infectious disease epidemiology but also in physics, chemistry and systems biology. Thus, the statistical procedures provided in this thesis can be applied in those fields as well. Especially our innovations regarding the estimation of marginal likelihoods might yield new insights as this is a common issue in model selection, e.g. see Schmidl et al. (2012).

7.3 Future research

This thesis not only provides suitable methodology when dealing with complex dynamic models together with insights into the epidemiological consequences of routine rotavirus vaccination. We also detected remaining pitfalls when treating large ODE-based models and fitting those models to disease incidence data. Here, we propose some alternative approaches for performing parameter inference in ODE-based models as well as possible model alterations in order to better match the observed fluctuations in the incidence data.

The most time-consuming task when handling high-dimensional models defined via

nonlinear ODEs is the approximate computation or sampling from the posterior, since every single likelihood evaluation demands the solution of the full ODE system via numerical means. Algorithms requiring as few model evaluations as possible are thus crucial for an efficient assessment of the posterior. One novel sampling procedure is the adaptive hybrid independence and random walk Metropolis-Hastings algorithm proposed by Schmidl et al. (2013). The sampler constructs a partially state-independent proposal density by applying a vine-copula decomposition based on an earlier posterior sample. The resulting sampling algorithm regularly draws independent parameter candidates while simultaneously maintaining high acceptance rates. This approach might speed up the inference procedure for the here employed transmission models.

Since the iterated numerical solution of an ODE system is massively slowing down any parameter inference procedure when dealing with ODE-based models, an estimation algorithm which completely circumvents solving the ODE system through numerical methods might yield much faster inference results. So-called collocation methods provide the base of a two-step inference procedure, which does not require tedious ODE simulation. The approach was originally proposed by Varah (1982), whereas further developments are summarized in Brunel (2008). The idea is to first compute a nonparametric estimate of the ODE solution and its derivative using spline regression techniques. In a second step the ODE-defining parameter vector is estimated by minimizing the distance between the estimated derivative and the output from plugging the estimated solution into the ODE system function. Due to the various available types of spline regression methods the described collocation method can select from a rich assortment of estimation techniques. However, although the approach is suitable for multi-dimensional ODEs it is still unclear whether it is also applicable for incompletely observed systems as in our case.

In order to address the autocorrelation of the residuals from a fitted model, another natural approach might be to find an alternative model which is better able to explain the observed pattern in the data. For the case of our transmission model this could be done by either modifying the observational model component to permit dependent subsequent observed case counts given the mean again coming from the ODE system, or by replacing the underlying hidden transmission model with a stochastic model based on jump processes (e.g. Cauchemez and Ferguson (2008)) or respective diffusion approximations (e.g. Fuchs (2013)). Whereas the first option presumably only increases the number of parameters to estimate, the utilization of a stochastic model for the unobserved transmission process requires more sophisticated inference methods such as data augmentation (Cauchemez and Ferguson (2008)) or simulation-based techniques (McKinley (2009)). However, for the large population studied in the present work, another approach which maintains the employed ODE-based model but allows for more flexibility is to apply hierarchical modelling by fitting the model for each observed season separately (e.g. Baguelin et al. (2013)) and estimate hyper distributions for the season specific parameters.

7.4 Public health impact

This thesis has demonstrated, that through mathematical-statistical modelling of transmission dynamics it becomes possible to estimate the population wide epidemiological impact of new potential vaccination programmes before deciding upon recommendation of the vaccination.

In the past, model-based results on the possible epidemiological effects were considered by the STIKO only in single occasions – such as for the recommendation of HPV vaccination for teenage girls (Ständige Impfkommission, 2014). As the present PhD project was the first transmission modelling study which was conducted by the RKI Immunization Unit itself, the RKI and thus also the STIKO located at the Immunization Unit built their first internal expertise in this field and started to acknowledge the relevance of model-based evaluations. Since the beginning of this project, the Immunization Unit together with the STIKO initiated three further modelling studies conducted externally. These aim at analysing the epidemiological impact of *(i)* varicella zoster and herpes zoster vaccination among children and elderly, respectively, *(ii)* meningococcal B childhood vaccination, and *(iii)* different serotype-specific vaccines against pneumococcal disease. Additionally, the RKI is funding an internal project on evaluating the impact of a potential routine influenza childhood-vaccination with a newly developed quadrivalent live attenuated vaccine.

Yet another aspect that goes beyond the epidemiological effects of vaccination programmes is the additional evaluation from an economic perspective by performing a cost-effectiveness analysis – see e.g. Ultsch et al. (2013) assessing the cost-effectiveness of herpes-zoster vaccination in Germany. For such an analysis the additional vaccination costs for the healthcare system, which might be partially compensated by reduced treatment costs, are compared with achieved epidemiological effects, e.g. prevented cases or hospitalisations. This yields a so-called incremental cost-effectiveness ratio (ICER) which provides the additional health care costs per prevented case or hospitalisation, respectively. A Bayesian approach as employed in our modelling fits also very well for such an health economic analysis as it provides a natural framework for computing the ICERs and their respective predictive distribution.

Considering the increasing utilisation of model-based analyses, what is missing so far is a decision making policy for the STIKO which determines whether new vaccination recommendations should always require evidence from transmission models and if so, what weight should model conclusions have within the final decision-making. Moreover, such a policy should be accompanied with modelling guidelines to guarantee consistent methodological standards for all epidemiological questions. These should include, e.g., the types of models to apply, which kind of data (if available) has to be used for model fitting, and how to account for uncertainty regarding model choice and respective parameters. Considering this last point, the present thesis illustrated well-suited statistical procedures towards an honest assessment of uncertainty in mathematical transmission modelling, which is a basic requirement for supporting evidence-based health policy making.

Bibliography

Abbey, H. (1950). *An Examination of the Reed-Frost Theory of Epidemics.*

Abramowitz, M. and I. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications.

Aidelsburger, P., K. Grabein, K. Böhm, M. Dietl, J. Wasem, J. Koch, B. Ultsch, F. Weidemann, and O. Wichmann (2014). Cost-effectiveness of childhood rotavirus vaccination in Germany. *Vaccine 32*(17), 1964 – 1974.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory*, Budapest, pp. 267–281. Akadémiai Kiado.

Allen, L. (2003). *An Introduction to Stochastic Processes with Applications to Biology.* Pearson/Prentice Hall.

Anderson, E. J. and S. G. Weber (2004). Rotavirus infection in adults. *The Lancet Infectious Diseases 4*(2), 91 – 99.

Anderson, R. M., S. Gupta, and W. Ng (1990, 4). The Significance of Sexual Partner Contact Networks for the Transmission Dynamics of HIV. *Journal of Acquired Immune Deficiency Syndromes 3*, 317–329.

Anderson, R. M. and R. M. May (1991). *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press.

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

Atchison, C., B. Lopman, and W. J. Edmunds (2010). Modelling the seasonality of rotavirus disease and the impact of vaccination in England and Wales. *Vaccine 28*(18), 3118 – 3126.

Atkins, K. E., E. Shim, V. E. Pitzer, and A. P. Galvani (2012). Impact of rotavirus vaccination on epidemiological dynamics in England and Wales. *Vaccine 30*(3), 552 – 564.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*, 171–178.

Azzalini, A. and A. Genz (2014). *The R package* `mnormt`*: The multivariate normal and t distributions (version 1.5-1).*

Baguelin, M., S. Flasche, A. Camacho, N. Demiris, E. Miller, and W. J. Edmunds (2013, 10). Assessing optimal target populations for influenza vaccination programmes: An evidence synthesis and modelling study. *PLoS Med 10*(10), e1001527.

Bailey, N. and L. Bailey (1987). *The Mathematical Theory of Infectious Diseases.* Mathematics in Medicine Series. Oxford University Press.

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

Bartlett, M. S. (1956). Deterministic and stochastic models for recurrent epidemics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Health*, Berkeley, Calif., pp. 81–109. University of California Press.

Bates, D. and M. Maechler (2014). *The R package* `Matrix`*: Sparse and Dense Matrix Classes and Methods (version 1.1-4).*

Becker, N. (1989a). *Analysis of Infectious Disease Data.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Becker, N. (1989b). *Analysis of Infectious Disease Data.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Bilcke, J., P. Beutels, M. Brisson, and M. Jit (2011). Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: A practical guide. *Medical Decision Making 31*(4), 675–692.

Birrell, P. J., G. Ketsetzis, N. J. Gay, B. S. Cooper, A. M. Presanis, R. J. Harris, A. Charlett, X.-S. Zhang, P. J. White, R. G. Pebody, and D. De Angelis (2011). Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proceedings of the National Academy of Sciences 108*(45), 18238–18243.

Block, S. L., T. Vesikari, M. G. Goveia, S. B. Rivers, B. A. Adeyi, M. J. Dallas, J. Bauder, J. W. Boslego, P. M. Heaton, and for the Pentavalent Rotavirus Vaccine Dose Confirmation Efficacy Study Group (2007). Efficacy, immunogenicity, and safety of a pentavalent human-bovine (wc3) reassortant rotavirus vaccine at the end of shelf life. *Pediatrics 119*(1), 11–18.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods.* New York: Springer.

Brookhart, M. A., A. E. Hubbard, M. J. van der Laan, J. M. Colford, and J. N. S. Eisenberg (2002). Statistical estimation of parameters in a disease transmission model: analysis of a Cryptosporidium outbreak. *Statistics in Medicine 21*(23), 3627–3638.

Brooks, S. P., P. Giudici, and G. O. Roberts (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(1), 3–39.

Brunel, N. J.-B. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electron. J. Statist. 2*, 1242–1267.

Cairns, A. J. G. (2004). *Interest Rate Models: An Introduction.* Boca Raton: Princeton University Press.

Casella, G. and E. I. George (1992). Explaining the Gibbs Sampler. *The American Statistician 46*(3), 167–174.

Cauchemez, S. and N. M. Ferguson (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *Journal of The Royal Society Interface 5*(25), 885–897.

Chib, S. and I. Jeliazkov (2001). Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association 96*, 270–281.

Chong, K. and B. Ying Zee (2012). Modeling the impact of air, sea, and land travel restrictions supplemented by other interventions on the emergence of a new influenza pandemic virus. *BMC Infectious Diseases 12*(1).

Christensen, R. (2001). *Advanced Linear Modeling, Second Edition.* Springer.

Chung, K. and R. Williams (2014). *Introduction to Stochastic Integration.* New York: Springer.

Clark, H. F., A. Marcello, D. Lawley, M. Reilly, and M. DiNubile (2010). Unexpectedly high burden of rotavirus gastroenteritis in very young infants. *BMC Pediatrics 10*(1), 40.

Clemens, J., M. Rao, F. Ahmed, R. Ward, S. Huda, J. Chakraborty, M. Yunus, M. R. Khan, M. Ali, B. Kay, F. van Loon, and D. Sack (1993). Breast-Feeding and the Risk of Life-Threatening Rotavirus Diarrhea: Prevention or Postponement? *Pediatrics 92*(5), 680–685.

Coelho, F. C., C. T. Codeço, and M. G. M. Gomes (2011, 05). A Bayesian Framework for Parameter Estimation in Dynamical Models. *PLoS ONE 6*(5), e19616.

Cohn, D. L. (2013). *Measure Theory.* New York: Springer.

Czado, C., T. Gneiting, and L. Held (2009). Predictive Model Assessment for Count Data. *Biometrics 65*(4), 1254–1261.

Daley, D., J. Gani, and J. Gani (2001). *Epidemic Modelling: An Introduction.* Cambridge Studies in Mathematical Biology. Cambridge University Press.

de Blasio, B. F., K. Kasymbekova, and E. Flem (2010). Dynamic model of rotavirus transmission and the impact of rotavirus vaccination in Kyrgyzstan. *Vaccine 28*(50), 7923 – 7932.

Dorigatti, I., S. Cauchemez, A. Pugliese, and N. M. Ferguson (2012). A new approach to characterising infectious disease transmission dynamics from sentinel surveillance: Application to the Italian 2009–2010 A/H1N1 influenza pandemic. *Epidemics 4*(1), 9 – 21.

Dudareva, S., J. Koch, M. An der Heiden, D. Oberle, B. Keller-Stanislawski, and O. Wichmann (2012). Impact of rotavirus vaccination in regions with low and moderate vaccine uptake in Germany. *Human Vaccines & Immunotherapeutics 8*(10), 1407 – 1415.

Efron, B. and D. V. Hinkley (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika 65*(3), 457–483.

Elderd, B. D., V. M. Dukic, and G. Dwyer (2006). Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proceedings of the National Academy of Sciences 103*(42), 15693–15697.

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature 460*(687), 19356–19356.

Fahrmeir, L., T. Kneib, S. Lang, and B. Marx (2013). *Regression - Models, Methods and Applications.* Berlin: Springer.

Federal Bureau of Statistics (2013). GENESIS Online Database.

Finkenstädt, B. F. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 49*(2), 187–205.

Fischer, T. K., P. Valentiner-Branth, H. Steinsland, M. Perch, G. Santos, P. Aaby, K. Mølbak, and H. Sommerfelt (2002). Protective Immunity after Natural Rotavirus Infection: A Community Cohort Study of Newborn Children in Guinea-Bissau, West Africa. *Journal of Infectious Diseases 186*(5), 593–597.

Forsythe, G. E., M. A. Malcolm, and C. B. Moler (1977). *Computer Methods for Mathematical Computations.* Prentice Hall Professional Technical Reference.

Friel, N. and A. N. Pettitt (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(3), 589–607.

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

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and R. D. (2013). *Bayesian Data Analysis.* New Jersey: CRC Press.

Gelman, A., G. O. Roberts, and W. R. Gilks (1996). Efficient Metropolis jumping rules. *Bayesian statistics 5*, 599–607.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics*, pp. 169–193. University Press.

Ghosal, S. (1999, 04). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli 5*(2), 315–331.

Giaquinto, C., P. Van Damme, F. Huet, L. Gothefors, M. Van der Wielen, and on behalf of the REVEAL Study Group (2007). Costs of community-acquired pediatric rotavirus gastroenteritis in 7 european countries: The reveal study. *Journal of Infectious Diseases 195*(Supplement 1), S36–S44.

Gilbert, P. and R. Varadhan (2012). *numDeriv: Accurate Numerical Derivatives.* R package version 2012.9-1.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics 22*(4), 403 – 434.

Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics 115*(4).

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4), 711–732.

Haario, H., E. Saksman, and J. Tamminen (2001). An Adaptive Metropolis algorithm. *Bernoulli 7*, 223–242.

Hardin, J. W. and J. M. Hilbe (2007). *Generalized Linear Models and Extensions.* Stata Press.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika 57*(1), 97–109.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences 44*(1), 1–12.

He, D., E. L. Ionides, and A. A. King (2010). Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of The Royal Society Interface 7*(43), 271–283.

Held, L. and D. S. Bové (2014). *Applied Statistical Inference – Likelihood and Bayes.* Berlin: Springer.

Held, L., M. Hofmann, M. Höhle, and V. Schmid (2006). A two-component model for counts of infectious diseases. *Biostatistics 7*(3), 422–437.

Held, L., M. Höhle, and M. Hofmann (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling 5*(3), 187–199.

Hens, N., G. Ayele, N. Goeyvaerts, M. Aerts, J. Mossong, J. Edmunds, and P. Beutels (2009). Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight european countries. *BMC Infectious Diseases 9*(1), 187.

Heyde, C. and I. Johnstone (1979). On asymptotic posterior normality for stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological) 41*(2), 184–189.

Hosmer, D. W. J., S. Lemeshow, and R. X. Sturdivant (2013). *Applied Logistic Regression, 3rd edition.* New York: Wiley.

Ionides, E. L., A. Bhadra, Y. Atchadé, and A. King (2011, 06). Iterated filtering. *The Annals of Statistics 39*(3), 1776–1802.

Joint Committee on Vaccination and Immunisation (2013). Code of Practice June 2013.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Keeling, M. J. and K. T. Eames (2005). Networks and epidemic models. *Journal of The Royal Society Interface 2*(4), 295–307.

Keeling, M. J. and P. Rohani (2007). *Modeling Infectious Diseases in Humans and Animals.* Princeton University Press.

Klinkenberg, D., J. de Bree, H. Laevens, and M. de Jong (2002). Within- and between-pen transmission of Classical Swine Fever Virus: a new method to estimate the basic reproduction ratio from transmission experiments. *Epidemiology and Infection 128*(2), 293–299.

Koch, J. and M. Wiese-Posselt (2011). Epidemiology of rotavirus infections in children < 5 years of age: Germany, 2001-2008. *Pediatric Infectious Disease Journal 30*(2), 112–117.

Koch, J., M. Wiese-Posselt, C. Remschmidt, O. Wichmann, H. Bertelsmann, E. Garbe, H. Hengel, J. Meerpohl, A. Mas Marques, H. Oppermann, E. Hummers-Pradier, R. von Kries, and T. Mertens (2013). Background paper to the recommendation for routine rotavirus vaccination of infants in Germany. *Bundesgesundheitsblatt 2013*(56), 957–984.

Krause, G., D. Altmann, D. Faensen, K. Porten, J. Benzler, T. Pfoch, A. Ammon, M. H. Kramer, and H. Claus (2007). SurvNet Electronic Surveillance System for Infectious Disease Outbreaks, Germany. *Emerg Infect Dis 13*(10).

Lee, R. D. and L. R. Carter (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association 87*(419), 659–671.

Levin, D. A., Y. Peres, and E. Wilmer (2008). *Markov Chains and Mixing Times.* American Mathematical Society.

Lindsey, J. K. (1996). *Parametric Statistical Inference.* Oxford University Press.

Liu, H., J. D. Lafferty, and L. A. Wasserman (2007). Sparse nonparametric density estimation in high dimensions using the rodeo. In M. Meila and X. Shen (Eds.), *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS-07)*, Volume 2, pp. 283–290. Journal of Machine Learning Research - Proceedings Track.

Liu, J. S., F. Liang, and W. H. Wong (2000). The Multiple-Try Method and Local Optimization in Metropolis Sampling. *Journal of the American Statistical Association 95*(449), 121–134.

Lopman, B. A., A. T. Curns, C. Yen, and U. D. Parashar (2011). Infant Rotavirus Vaccination May Provide Indirect Protection to Older Children and Adults in the United States. *Journal of Infectious Diseases 204*(7), 980–986.

Lunelli, A., C. Rizzo, S. Puzelli, A. Bella, E. Montomoli, M. C. Rota, I. Donatelli, and A. Pugliese (2013). Understanding the dynamics of seasonal influenza in Italy: incidence, transmissibility and population susceptibility in a 9-year period. *Influenza and Other Respiratory Viruses 7*(3), 286–295.

Martinón-Torres, F., M. Alejandro, L. Collazo, J. Lastres, S. Díaz, M. Pillado, and J. Sánchez (2011). Effectiveness of rotavirus vaccination in Spain. *Human Vaccines 7*(7), 757–761.

Matis, J. and T. Kiffe (2000). *Stochastic Population Models: A Compartmental Perspective.* Lecture Notes in Statistics. Springer New York.

McBryde, E., I. Bergeri, C. Van Gemert, J. Rotty, E. Headley, K. Simpson, R. Lester, M. Hellard, and J. Fielding (2008). Early transmission characteristics of influenza A(H1N1) v in Australia: Victorian state, 16 May-3 June 2009. *Euro surveillance: bulletin Europeen sur les maladies transmissibles= European communicable disease bulletin 14*(42), 19356–19356.

McKinley, T. (2009). Inference in epidemic models without likelihoods. *International Journal of Biostatistics 5*(1), 24–24.

McMillan, H., J. Freer, F. Pappenberger, T. Krueger, and M. Clark (2010). Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes 24*(10), 1270–1284.

Mossong, J., N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds (2008, 03). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med 5*(3), e74.

Moulton, L. H., M. A. Staat, M. Santosham, and R. L. Ward (1998). The Protective Effectiveness of Natural Rotavirus Infection in an American Indian Population. *Journal of Infectious Diseases 178*(6), 1562–1566.

Mrukowicz, J. Z., J. Thompson, G. W. Reed, S. J. Tollefson, M. Kobayashi, K. Araki, and P. F. Wright (1999). Epidemiology of rotavirus in infants and protection against symptomatic illness afforded by primary infection and vaccination. *Vaccine 17*(7-8), 745 – 753.

Nash, J. C. and R. Varadhan (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software 43*(9), 1–14.

Neal, R. M. (2003, 06). Slice sampling. *The Annals of Statistics 31*(3), 705–767.

Nelson, K. E. and C. Williams (2014). *Infectious Disease Epidemiology - Theory and Practice.* Burlington: Jones & Bartlett Learning.

O'Neill, P. D. (2010). Introduction and snapshot review: Relating infectious disease transmission models to data. *Statistics in Medicine 29*(20), 2069–2077.

Pasarica, C. and A. Gelman (2010). Adaptively Scaling the Metropolis Algorithm Using Expected Squared Jumped Distance. *Statistica Sinica 20*, 343–364.

Paul, M., L. Held, and A. M. Toschke (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine 27*(29), 6250–6267.

Paulke-Korinek, M., H. Kollaritsch, S. W. Aberle, I. Zwazl, B. Schmidle-Loss, A. Vécsei, and M. Kundi (2013). Sustained low hospitalization rates after four years of rotavirus mass vaccination in Austria. *Vaccine 31*(24), 2686 – 2691.

Payne, D. C., M. A. Staat, K. M. Edwards, P. G. Szilagyi, G. A. Weinberg, C. B. Hall, J. Chappell, A. T. Curns, M. Wikswo, J. E. Tate, B. A. Lopman, U. D. Parashar, and the New Vaccine Surveillance Network (NVSN) (2011). Direct and Indirect Effects of Rotavirus Vaccination Upon Childhood Hospitalizations in 3 US Counties, 2006–2009. *Clinical Infectious Diseases 53*(3), 245–253.

Perez, L. and S. Dragicevic (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics 8*(1), 50.

Pickering, L. K., A. V. B. III, R. R. Reves, and A. Morrow (1988). Asymptomatic excretion of rotavirus before and after rotavirus diarrhea in children in day care centers. *The Journal of Pediatrics 112*(3), 361 – 365.

Pickering, L. K., A. L. Morrow, I. Herrera, M. O'Ryan, M. K. Estes, S. E. Guilliams, L. Jackson, S. Carter-Campbell, and D. O. Matson (1995). Effect of Maternal Rotavirus Immunization on Milk and Serum Antibody Titers. *Journal of Infectious Diseases 172*(3), 723–728.

Pitzer, V. E., K. E. Atkins, B. F. de Blasio, T. Van Effelterre, C. J. Atchison, J. P. Harris, E. Shim, A. P. Galvani, W. J. Edmunds, C. Viboud, M. M. Patel, B. T. Grenfell, U. D. Parashar, and B. A. Lopman (2012, 08). Direct and indirect effects of rotavirus vaccination: Comparing predictions from transmission dynamic models. *PLoS ONE 7*(8), e42320.

Pitzer, V. E., M. M. Patel, B. A. Lopman, C. Viboud, U. D. Parashar, and B. T. Grenfell (2011). Modeling rotavirus strain dynamics in developed countries to understand the potential impact of vaccination on genotype distributions. *Proceedings of the National Academy of Sciences 108*(48), 19353–19358.

Pitzer, V. E., C. Viboud, B. A. Lopman, M. M. Patel, U. D. Parashar, and B. T. Grenfell (2011). Influence of birth rates and transmission rates on the global seasonality of rotavirus incidence. *Journal of The Royal Society Interface*.

Pitzer, V. E., C. Viboud, L. Simonsen, C. Steiner, C. A. Panozzo, W. J. Alonso, M. A. Miller, R. I. Glass, J. W. Glasser, U. D. Parashar, and B. T. Grenfell (2009). Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics. *Science 325*(5938), 290–294.

Poole, D. and A. E. Raftery (2000). Inference for Deterministic Simulation Models: The Bayesian Melding Approach. *Journal Of The American Statistical Association*, 1244–1255.

Presanis, A. M., D. De Angelis, A. Goubar, O. N. Gill, and A. E. Ades (2011). Bayesian evidence synthesis for a transmission dynamic model for HIV among men who have sex with men. *Biostatistics 12*(4), 666–681.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007). *Numerical Recipes 3rd Edition*. Cambridge University Press.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. ISBN 3-900051-07-0.

Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika 83*(2), 251–266.

Raftery, A. E., M. A. Newton, J. M. Satagopan, and P. N. Krivitsky (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In *Bayesian Statistics*, pp. 1–45.

Railsback, S. F. and V. Grimm (2011). *Agent-Based and Individual-Based Modeling:A Practical Introduction*. Princeton University Press.

Richey, M. (2010). The evolution of Markov chain Monte Carlo methods. *The American Mathematical Monthly 117*(5), 383–413.

Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods (2nd edition)*. New York: Springer.

Roberts, G. O. and J. S. Rosenthal (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability 44*, 458–475.

Roberts, G. O. and R. L. Tweedie (1996, 12). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli 2*(4), 341–363.

Rosner, B., K. Stark, and D. Werber (2010). Epidemiology of reported Yersinia enterocolitica infections in Germany, 2001-2008. *BMC Public Health 10*(1), 337.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 319–392.

Ruiz-Palacios, G. M., I. Pérez-Schael, F. R. Velázquez, H. Abate, T. Breuer, S. C. Clemens, B. Cheuvart, F. Espinoza, P. Gillard, B. L. Innis, Y. Cervantes, A. C. Linhares, P. López, M. Macías-Parra, E. Ortega-Barría, V. Richardson, D. M. Rivera-Medina, L. Rivera, B. Salinas, N. Pavía-Ruz, J. Salmerón, R. Rüttimann, J. C. Tinoco, P. Rubio, E. Nuñez, M. L. Guerrero, J. P. Yarzábal, S. Damaso, N. Tornieporth, X. Sáez-Llorens, R. F. Vergara, T. Vesikari, A. Bouckenooghe, R. Clemens, B. De Vos, and M. O'Ryan (2006). Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. *New England Journal of Medicine 354*(1), 11–22. PMID: 16394298.

Salimans, T. and Knowles, D. (2013, 12). Fixed-form variational posterior approximation through stochastic linear regression. *Bayesian Analysis 8*(4), 837–882.

Schmidl, D., C. Czado, S. Hug, and F. J. Theis (2013, 03). A vine-copula based adaptive mcmc sampler for efficient inference of dynamical systems. *Bayesian Anal. 8*(1), 1–22.

Schmidl, D., S. Hug, W. Li, M. Greiter, and F. Theis (2012). Bayesian model selection validates a biokinetic model for zirconium processing in humans. *BMC Systems Biology 6*(1).

Schwarz, G. (1978, 03). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* New York: Wiley.

Shim, E. and C. Banks, H.T. Castillo-Chavez (2006). Seasonality of rotavirus infection with its vaccination. *AMS contemporary mathematics book series 410.*

Shim, E. and C. Castillo-Chavez (2009). The epidemiological impact of rotavirus vaccination in the United States and Mexico. *Mathematical and Statistical Estimation Approaches in Epidemiology*, 303–323.

Shim, E. and A. P. Galvani (2009). Impact of transmission dynamics on the cost-effectiveness of rotavirus vaccination. *Vaccine 27*(30), 4025 – 4030.

Soares-Weiser, K., H. MacLehose, H. Bergman, I. Ben-Aharon, S. Nagpal, E. Goldberg, F. Pitan, and N. Cunliffe (2012). Vaccines for preventing rotavirus diarrhoea: Vaccines in use. *Cochrane Database of Systematic Reviews* (2).

Soetaert, K. and T. Petzoldt (2010). Inverse modelling, sensitivity and monte carlo analysis in R using package FME. *Journal of Statistical Software 33*(3), 1–28.

Soetaert, K., T. Petzoldt, and R. W. Setzer (2010). Solving Differential Equations in R: Package deSolve. *Journal of Statistical Software 33*(9), 1–25.

Soriano-Gabarró, M., J. Mrukowicz, T. Vesikari, and T. Verstraeten (2006). Burden of Rotavirus Disease in European Union Countries. *The Pediatric Infectious Disease Journal 25*(1), 7–11.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

Ständige Impfkommission (2013). Empfehlungen der Ständigen Impfkommission (STIKO) am Robert Koch-Institut/Stand: August 2013. *Epidemiologisches Bulletin 2013*(34), 313–344.

Ständige Impfkommission (2014). Wissenschaftliche Begründung für die Änderung der Empfehlung zur Impfung gegen humane Papillomviren. *Epidemiologisches Bulletin 2014* (35), 343–347.

Statistisches Bundesamt (2009). Bevölkerung Deutschlands bis 2060: 12. koordinierte Bevölkerungsvorausberechnung.

Tate, J. E., A. H. Burton, C. Boschi-Pinto, A. D. Steele, J. Duque, and U. D. Parashar (2012). 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *The Lancet Infectious Diseases 12* (2), 136–141.

Thiébaux, H. J. and F. W. Zwiers (1984, May). The Interpretation and Estimation of Effective Sample Size. *Journal of Applied Meteorology 23*, 800–811.

Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface 6* (31), 187–202.

Ultsch, B., F. Weidemann, T. Reinhold, A. Siedler, G. Krause, and O. Wichmann (2013). Health economic evaluation of vaccination strategies for the prevention of herpes zoster and postherpetic neuralgia in Germany. *BMC Health Services Research 13* (1), 359.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Van Effelterre, T., M. M. Soriano-Gabarró, S. Debrus, E. Claire Newbern, and J. Gray (2010). A mathematical model of the indirect effects of rotavirus vaccination. *Epidemiology and Infection 138* (6), 884–897.

Varah, J. M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing 3* (1), 28–46.

Velásquez, F. R., D. O. Matson, and G. M. Ruiz-Palacios (1996). Rotavirus Infection in Infants as Protection Against Subsequent Infections. *Vaccine 335* (14), 1022–1028.

Vesikari, T. (2008). The role of scientific societies in the decision-making process to recommend new vaccines: the example of rotavirus in europe. *Journal of Public Health 16* (4), 287–290.

Vesikari, T., A. Karvonen, R. Prymula, V. Schuster, J. Tejedor, R. Cohen, F. Meurice, H. Han, S. Damaso, and A. Bouckenooghe (2007). Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in European infants: randomised, double-blind controlled study. *The Lancet 370* (9601), 1757 – 1763.

Vesikari, T., D. O. Matson, P. Dennehy, P. Van Damme, M. Santosham, Z. Rodriguez, M. J. Dallas, J. F. Heyse, M. G. Goveia, S. B. Black, H. R. Shinefield, C. D. Christie, S. Ylitalo, R. F. Itzler, M. L. Coia, M. T. Onorato, B. A. Adeyi, G. S. Marshall, L. Gothefors, D. Campens, A. Karvonen, J. P. Watt, K. L. O'Brien, M. J. DiNubile, H. F. Clark, J. W. Boslego, P. A. Offit, and P. M. Heaton (2006). Safety and Efficacy of a Pentavalent Human–Bovine (WC3) Reassortant Rotavirus Vaccine. *New England Journal of Medicine 354*(1), 23–33. PMID: 16394299.

Vynnycky, E., R. Pitman, R. Siddiqui, N. Gay, and W. J. Edmunds (2008). Estimating the impact of childhood influenza vaccination programmes in England and Wales. *Vaccine 26*(41), 5321 – 5330.

Ward, R. L., D. I. Bernstein, R. Shukla, M. M. McNeal, J. R. Sherwood, E. C. Young, and G. M. Schiff (1990). Protection of Adults Rechallenged with a Human Rotavirus. *Journal of Infectious Diseases 161*(3), 440–445.

Wei, G. C. G. and M. A. Tanner (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika 77*(3), 649–652.

Weidemann, F., M. Dehnert, J. Koch, O. Wichmann, and M. Höhle (2014a). Bayesian parameter inference for dynamic infectious disease modelling: rotavirus in Germany. *Statistics in Medicine 33*(9), 1580–1599.

Weidemann, F., M. Dehnert, J. Koch, O. Wichmann, and M. Höhle (2014b). Modelling the epidemiological impact of rotavirus vaccination in Germany – A Bayesian approach. *Vaccine 32*(40), 5250 – 5257.

Whitaker, H. J. and C. P. Farrington (2004). Estimation of infectious disease parameters from serological survey data: the impact of regular epidemics. *Statistics in Medicine 23*(15), 2429–2443.

Wilde, J., R. Yolken, R. Willoughby, and J. Eiden (1991). Improved detection of rotavirus shedding by polymerase chain reaction. *The Lancet 337*(8737), 323 – 326.

Wilkinson, D. (2006). *Stochastic Modelling for Systems Biology.* Chapman & Hall/CRC Mathematical & Computational Biology. Taylor & Francis.

Willem, L., K. Van Kerckhove, D. L. Chao, N. Hens, and P. Beutels (2012, 11). A Nice Day for an Infection? Weather Conditions and Social Contact Patterns Relevant to Influenza Transmission. *PLoS ONE 7*(11), e48695.

Wood, J., N. Zamani, C. MacIntyre, and N. Beckert (2007). Effects of internal border control on spread of pandemic influenza. *Emerg Infect Dis 13*, 1038–1045.

Eidesstattliche Versicherung

(siehe Promotionsordnung vom 12.07.11, §8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

_____

München, den 30.07.2015   Felix Weidemann