

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical
Engineering
Department of Mathematical Sciences

Hedda Hognedatter Bjørnebye Vik

Forecasting Child Mortality while Accounting for Complex Survey Designs

Master's thesis in Applied Pysics and Mathematics

Supervisor: Geir-Arne Fuglestad

June 2019



Norwegian University of
Science and Technology

Hedda Hognedatter Bjørnebye Vik

Forecasting Child Mortality while Accounting for Complex Survey Designs

Master's thesis in Applied Physics and Mathematics
Supervisor: Geir-Arne Fuglestad
June 2019

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

 **NTNU**
Norwegian University of
Science and Technology

Summary

Child mortality rates are key indicators of the health of a nation. The United Nations' (UN) millennium development goals (MDGs) and sustainable development goals (SDGs) focus on reducing child mortality. To measure progress toward the UN goals and to identify the need for interventions, accurate estimates and forecasts of neonatal mortality rate (NMR) and under-5 mortality rate (U5MR) are needed. The main goal of this thesis is to improve on current estimates and forecasts of NMR and U5MR in low-and-medium income countries by borrowing strength across time and from neighboring countries. The best performing models are used to make forecasts of NMR and U5MR for 2030 for Kenya and evaluate Kenya's progress towards the MDGs and SDGs.

In low- and medium-income countries, child mortality rates are primarily estimated using survey and census data. Traditional model-based approaches to estimation do not account for complex survey design. Not accounting for stratification and clustering leads to biased estimates and incorrect uncertainties. In this thesis, yearly design-based estimates are computed independently. Then, as mortality rates are expected to be slowly changing over time, a temporal model is applied. The model is latent Gaussian and has a linear trend in time, a stochastic temporal effect, and an effect correcting for systematic differences between surveys. Time-dependent covariates can be included, and joint modeling of neighboring countries is considered. The models are estimated using the integrated nested Laplace approximation (INLA).

The analysis concludes that the best model for forecasting NMR in Kenya has an autoregressive process of order two as the temporal effect, a linear trend in time, and the logarithm of the gross domestic product (GDP) as a covariate. It does not borrow strength from neighboring countries. The best model for forecasting U5MR in Kenya has a random walk of order one as the temporal effect, GDP as a covariate, and borrows strength from Uganda and Tanzania. Using that model, Kenya did not reach MDG target 4A: reduce U5MR by two thirds between 1990 and 2015. SDG target 3.2 is to reduce NMR to 12 per 1000 live birth and U5MR to 25 per live birth by 2030. The preferred models give Kenya a 29.63% probability of reaching SDG target 3.2 for NMR and a 99.97% probability of reaching SDG target 3.2 for U5MR. The probability of reaching SDG target 3.2 is sensitive to the choice of model.

Sammendrag

Barnedødelighetsrater er viktige indikatorer på en nasjons helse. FNs tusenårs mål (MDG) og bærekraftsmål (SDG) fokuserer på å redusere barnedødelighet. Man trenger nøyaktige estimater og prediksjoner for nyfødtdødelighet (NMR) og under-5 dødelighet (U5MR) for å måle fremdrift mot FNs mål og for å identifisere behovet for intervensjoner. Hovedmålet med denne oppgaven er å forbedre dagens estimater og prediksjoner for NMR og U5MR i lav- og medium-inntektsland ved å låne styrke over tid og fra naboland. De beste modellene brukes til å produsere prediksjoner for NMR og U5MR i Kenya frem til 2030 og evaluere Kenyas fremgang mot FN-målene.

I lav- og medium-inntektsland estimeres barnedødelighet primært ved hjelp av undersøkelser og folketelling data. Tradisjonelle modellbaserte tilnæringer til estimering tar ikke hensyn til de komplekse utformingene av undersøkelsene. Ignorere stratifisering og *clustering* fører til *biased* estimater og feil usikkerhet. I denne oppgaven beregnes årlige designbaserte estimater uavhengig. Ettersom det forventes at dødelighetsrater endres sakte over tid, blir en tidsmodell brukt. Modellen er latent Gaussisk og har en lineær trend i tid, en stokastisk temporal effekt og en korrigerende for systematiske forskjeller mellom undersøkelser. Tidsavhengige kovariater kan inkluderes, og felles modellering av naboland utforskes. Modellene er estimert ved hjelp av den integrerte nestede Laplace-tilnærmingen (INLA).

Analysen konkluderer med at den beste modellen for predikering av NMR i Kenya har en autoregressiv prosess av orden to som tids-effekt, en lineær trend i tid, logaritmen for bruttonasjonalproduktet (BNP) som kovariat og låner ikke styrke fra nabolandene. Den beste modellen for prediksjoner for U5MR i Kenya har en *random walk* av første orden som temporal effekt, BNP som kovariat og låner styrke fra Uganda og Tanzania. Basert på prediksjoner produsert av denne modellen nådde Kenya ikke MDG-mål 4A: redusere U5MR med to tredjedeler mellom 1990 og 2015. SDG-mål 3.2 er å redusere NMR til 12 per 1000 levendefødt og U5MR til 25 per levendefødt innen 2030. Den foretrukne modellen for prediksjon av NMR gir Kenya en sannsynlighet på 29.63% for å nå SDG mål 3.2 for NMR. Den beste modellen for U5MR gir Kenya en 99,97% sannsynlighet for å nå SDG mål 3.2 for U5MR. Sannsynligheten for å nå SDG mål 3.2 er sensitiv til valg av modell.

Preface

This thesis is written during the spring of 2019, and it is my final assignment in the five-year master program *Applied Physics and Mathematics* at the Norwegian University of Science and Technology (NTNU), with specialization *Industrial Mathematics*. The thesis has not only allowed me to combine and conceptualize knowledge I have acquired during the last five years, but also required me to gain a deeper understanding of new statistical concepts and statistical modeling.

The thesis aims to estimate and forecast child mortality rates in low- and medium-income countries in order to assess the progress towards the United Nations' sustainable development goals and millennium development goals. Survey data from the Demographic and Health Survey program is used for analysis. The methods are implemented in R, with the help of the `survey` and `INLA` packages.

I want to thank my supervisor, Geir-Arne Fuglstad, for helpful guidance, support, and encouragement. I would also like to thank all my friends whom without I would not have made it through the last five years.

Hedda Hognedatter Bjørnebye Vik,
Trondheim, June 2019.

Table of Contents

Summary	i
Sammendrag	ii
Preface	iii
Table of Contents	v
1 Introduction	1
1.1 State of the art	2
1.2 Approach	3
1.3 Child mortality in Kenya	6
2 Background Theory	9
2.1 Complex survey methodology	9
2.2 Temporal modeling	12
2.3 Bayesian hierarchical models	16
2.4 Bayesian inference	18
2.5 Multivariate temporal models	21
2.6 Model assessment and scoring rules	22
3 Methods	25
3.1 Producing yearly design-based mortality rate estimates	25
3.2 Estimating models using <code>inla()</code>	32
3.3 Extracting posterior inference	35
3.4 Including covariates	36
4 Simulation Study	39
4.1 Details of the simulation study	39
4.2 Autoregressive process of order 1	41
4.3 Autoregressive process of order 2	47

4.4	Discussion	52
5	Univariate Modeling of Child Mortality Rates in Kenya	53
5.1	Model selection	53
5.2	Estimation and forecasting	62
6	Joint Modeling of Child Mortality	69
6.1	Effect of borrowing strength from neighboring countries on forecast . . .	69
6.2	Effect of borrowing strength from neighboring countries on estimation . .	83
7	Discussion	85
	Bibliography	89
A	Additional Simulation Results	97
B	Cross Plots of Responses	107
C	Estimated Coefficients, Chapter 5	111
D	Additional Plots Section 6.1	115
E	Additional Plots Section 6.2	119

Introduction

Under-5 mortality rate (U5MR: the probability of dying before the age of five for live births) and neonatal mortality rate (NMR: the proportion of children who die within the first month of life for live births) are important indicators of the degree of poverty and deprivation in a country. The United Nations' (UN's) third sustainable development goal (SDG) is "good health and well-being". Target 3.2 reads: "By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births" (UN (2019a)). Similarly, the fourth UN millennium development goal (MDG) is "reduce child mortality", with target 4A: "Reduce by two thirds, between 1990 and 2015, the under-5 mortality rate" (UN (2019b)). Health and social-service planning should consider trends in mortality rates and health drivers, as argued by Foremann et al. (2018). Thus, estimates and forecasts of NMR and U5MR are necessary both to assess progress towards development goals and to identify and plan needed interventions.

It is complicated to estimate and forecast child mortality rates because countries that have not already met SDG target 3.2 are typically low- and medium-income countries where vital registration systems are deficient. If no complete birth and death data exists, NMR and U5MR must be estimated using data from sparse complex surveys and censuses. When using survey data, the design of the survey must be considered. If not, the estimates will be biased, and one will not account for the uncertainty caused by the sampling. However, traditional design-based estimation approaches do not borrow strength across time or space. Spatial or temporal models can be used to improve estimates, obtain estimates for small areas in which too little data is collected to produce design-based estimates, and produce forecasts. Several models have been developed, some of which are discussed in the next section. The goal of this thesis is to investigate models that improve estimates and forecasts by borrowing strength across time and from neighboring countries and at the same time, account for the complex survey design.

1.1 State of the art

The UN currently uses the Bayesian B-spline bias-reduction (B3, Alkema and New (2014)) model to estimate child mortality rates. Alkema and New (2014) produce child mortality rates for 194 countries using a Bayesian framework where information is exchanged across countries. They use Bayesian penalized spline regression to produce yearly estimates for U5MR. When estimating the model, “second-order differences in adjacent spline coefficients . . . are penalized for guaranteeing the smoothness of the resulting U5MR trajectory” (Alkema and New (2014), p.8). They estimate the model using a Markov Chain Monte Carlo (MCMC) algorithm, which involves six parallel chains with a total of 50 000 iterations in each chain. Thus, estimating the B3 model is computationally demanding.

To establish a baseline for the SDGs Golding et al. (2017) carried out a large scale study for 46 African countries. They used data from 235 household surveys and censuses to generate estimates for NMR and U5MR for 5×5 km cells for five-year periods. When computing estimates for U5MR, Golding et al. (2017) considered four separate age intervals: the first month of life, month 1-11, month 12-35, and month 36-59. They modeled the age-intervals independently and finally computed the U5MR as one minus the probability of not dying in any of the age intervals. Wakefield et al. (2018) argue against this approach. They compute U5MR estimates at a fine spatial grid and estimate different age intervals simultaneously.

Other approaches described by Wang et al. (2014) and Foremann et al. (2018) estimate and forecast child mortality rates. Wang et al. (2014) do so simultaneously for 188 countries, using a three-step modeling process. Step one is a “non-linear mixed effects model to examine the relationship between child mortality, lagged distributed income per person, maternal education, and the crude death rate from HIV/AIDS in the under-5 age group” (Wang et al. (2014), p. 3). The second step borrows strength across time and countries within the same region using spatiotemporal regression. The third and final step applies a Gaussian process regression model. Foremann et al. (2018) forecast life expectancy and cause-specific mortality for 250 causes of death for 195 countries. Their model consists of three components; the first due to change in risk factors, the second is the underlying mortality rate for each cause as a function of income per capita, education, fertility rate under 25 years and time, and the third component models the unexplained stochastic variation in time using an autoregressive integrated moving average (ARIMA) model. When modeling the first two components, they assume that the cause-specific mortality rate is normally distributed on the log scale. The mean is specified as a sum of relevant risks. After that, the residuals are fitted by ARIMA models.

The models described by Alkema and New (2014), Golding et al. (2017), Wang et al. (2014) and Foremann et al. (2018) require large amounts of data from many countries. It is often difficult to collect data from various sources, and some data might be inaccessible. Further, it requires extensive domain-knowledge and computational powers to be able to handle huge amounts of data. The data they use is collected by several different programs and governments and is bound to be of varying quality. While some data is accurate, other data might be introducing more errors than information.

In addition, the methods described by Alkema and New (2014), Golding et al. (2017), Wang et al. (2014) and Foremann et al. (2018) are computationally complex. If estimated and forecasted mortality rates are to be of value, the entities responsible for social/eco-

nomical interventions must trust them. Therefore estimates and forecasts should be easy to reproduce. User-friendly software should be accessible as well as the computational powers required to estimate the model should not be too large. The models discussed above do not meet these criteria.

Mercer et al. (2015) combine data from surveys by the Demographic and Health Survey (DHS) program and Health Demographic Surveillance System (HDSS) sites from one country. DHS surveys collect information from a large population, whereas the HDSS data focus on a smaller population and is more accurate and precise. HDSS data is collected by repeated visits to the same households. Please see Section 2.1 of Mercer et al. (2015) for further explanation. The data is used to produce design-based U5MR estimates, and then a space-time Bayesian hierarchical model is applied. Mercer et al. (2015) aim to produce U5MR estimates at regional levels. Because of the sparsity of data, they produce estimates for five-year intervals. A free R-package called `SUMMER`, designed to generate and model mortality rate estimates using the methods and models described in Mercer et al. (2015), is under development. `SUMMER` allows for transparent estimation of mortality rates. The estimating method and model are relatively interpretable and straightforward.

To be able to assess progress towards SDG target 3.2, models producing reliable forecasts are needed. However, producing reliable forecasts has so far not been a focus for `SUMMER`. This thesis aims to investigate, assess, and suggest possible temporal models that can be included in `SUMMER`. Furthermore, the effect of covariates and borrowing strength from neighboring countries is investigated. The best performing model will be used to assess Kenya's progress towards the SDGs and MDGs. The focus of the thesis is limited to temporal models of mortality rates. The models described in this thesis can be expanded to spatiotemporal models. However, estimating mortality rates at a finer spatial scale requires large amounts of data or model-based approaches. The scope of this thesis would have been too broad and unfocused if complex spatiotemporal model-based approaches were included in it.

1.2 Approach

Estimation of mortality rates in low- and medium-income countries relies on complex survey data and censuses. Surveys collect the full birth history of women, which includes the dates of birth of the woman's children and if the children have died, the dates of death. In contrast, a census typically only contains summary birth histories; the number of children born to a woman and the number of children who died before turning five. Summary birth histories require methods such as the Brass method (Brass and Coale (1968)) to determine the NMR or U5MR. The Brass method, however, is complicated and not within the scope of this thesis. Surveys are constructed to be representative of the population, often with complex survey design. When computing estimates, the designs of the surveys have to be taken into account. If not, the resulting estimates and corresponding variance will be biased and inaccurate as the observations are not weighted correctly and do not account for dependence (i.e., clustering). In practice, design-based estimates can be computed using the R-package `survey`¹, which is used by, among others, `SUMMER`. The `survey`

¹<http://r-survey.r-forge.r-project.org/survey/>

package is developed by Thomas Lumley and is meant to facilitate analysis of complex survey samples. If data from several different surveys are used, separate estimates should be computed for each survey. As the surveys are conducted by different people, at different times, there might be systematic differences between them. That should be accounted for in the final models.

When computing design-based estimates, yearly NMR and U5MR are estimated without assuming temporal dependency. However, as mortality rates are expected to be slowly changing in time, yearly mortality rates are expected to have a temporal dependency. By modeling the temporal dependency, improved estimates of mortality rates can be obtained, and it is possible to produce forecasts. Temporal smoothing effects used in Mercer et al. (2015) are limited to random walks of first and second order. They find that random walk of second order is preferred for modeling U5MR in Tanzania. However, forecasts based on random walks of second order have exploding credible intervals, as seen in Figure 1.1. The models in Figure 1.1 are fitted using design-based mortality rates and uncertainty for Kenya for 1980-2013, and predictions are made until 2030. Both for U5MR and NMR, the 95% credible interval for the predictions includes unlikely high estimates; values that are way higher than past estimates and values that are very close to zero are included in the intervals. In this thesis, temporal effects based on random walks of first and second order will be compared to temporal effects based on autoregressive models of first and second order. These four models are chosen because they are simple and transparent. As there is not that much knowledge about the dynamics of mortality rates, constructing complex temporal models will not make sense.

In addition to investigating different temporal effects, it is also interesting to investigate the effect of covariates. To produce forecasts of mortality rates, future values of the covariates must be available. In this thesis, the effect of historical data and reliable forecasts of population size and gross domestic product (GDP) is investigated. Population size and GDP are chosen as covariates primarily to exemplify how covariates can be included. However, it is reasonable to assume that they affect mortality rates and high-quality historical data and projections at a national level are available. The data is obtained from the Shared Socioeconomic Pathways (SSP) database (Riahi et al. (2017), Rogelj et al. (2018) and Gidden et al. (2018)) and forecasts are made for five different future scenarios. It is also reasonable to assume that mortality rates in similar countries behave in the same way so that borrowing strength across countries is a good way of taking advantage of available data and improving the models. To test this, models that jointly estimate mortality rates in neighboring countries, vector autoregressive models, are investigated.

Due to the complex correlation structure of the problem, hierarchical Bayesian models are well suited. MCMC methods are widely used to estimate Bayesian models and can be made to be arbitrarily accurate by letting the algorithm run for long enough time. Thus, they can achieve accuracy at the cost of computational time. Software named `RStan` (Carpenter et al. (2017)) is being developed to fit Bayesian models using MCMC methods. `RStan` could have been used to fit the models in this thesis, but the software is inconvenient. The user is required to make manual adjustments for all different temporal effects tested, potentially leading to demanding debugging. Besides, it is difficult to assess convergence and therefore problematic to determine for how long the algorithm should run.

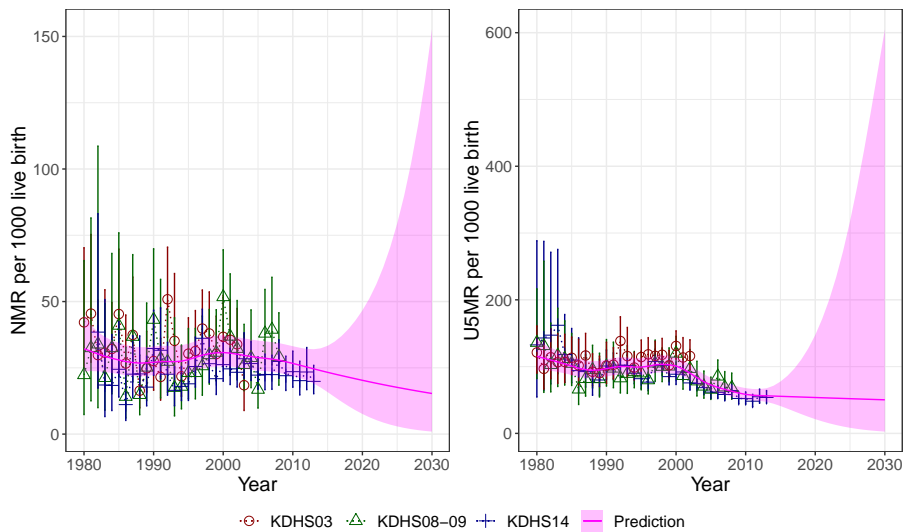


Figure 1.1: Estimated yearly NMR (left) and U5MR (right) for Kenya, given as expected deaths per 1000 live birth, with 95% credible intervals. Modeled with a random walk of second order as the temporal effect. The models are fitted using data for $t < 2014$, and then predictions are made until 2030. The red circles, error bars, and dotted lines are the estimated rates based on data from the Kenya DHS 2003 (KDHS03) survey. The green triangles, error bars, and dotted lines are the equivalent for estimates based on the Kenya DHS 2008-2009 survey (KDHS08-09), while the blue lines, error bars, and dotted lines correspond to estimates from the Kenya DHS 2014 survey (KDHS14). The solid purple lines are the predictions, and the purple areas are the 95% credible intervals of the predictions.

As in Mercer et al. (2015), the models developed in this thesis belongs to the class Bayesian hierarchical models called latent Gaussian models (LGMs). LGMs have an additive structure, which makes them easily interpretable. For LGMs, a recent estimation approach called the integrated nested Laplace approximation (INLA, Rue et al. (2009)) is developed. INLA is a quick, deterministic method that is very popular. INLA is implemented in the R-package `INLA`², and used by `SUMMER`. In contrast to `RStan`, INLA allows for easy use of several different temporal models and can estimate LGMs in a matter of seconds. Therefore, INLA is used in this thesis.

The explanatory and predictive strengths of the developed models are investigated using several simulated datasets. To evaluate the models, the Watanabe-Akaike information criterion (WAIC, Watanabe (2010)), marginal log-likelihood, mean absolute error, mean squared error, and the continuous ranked probability score (CRPS, Gneiting and Raftery (2007)) are calculated for every fitted model. Based on the models' performance using the simulated data sets, the best performing models are applied to design-based mortality rates in Kenya. The effect on forecasts of covariates and the effect of simultaneously calculating mortality rates for two neighboring countries, Uganda and Tanzania, is explored and assessed by CRPS.

1.3 Child mortality in Kenya

The models developed in this thesis are applied to survey data collected in Kenya. Kenya is an example of a country in which child mortality rates must be estimated from survey data. Even though it is chosen to analyze data from Kenya, the methods are also applicable to other low- and medium-income countries. Kenya is situated in East Africa, bordering Ethiopia (north), Somalia (northeast), Tanzania (south), Uganda (west), and South Sudan (northwest), as seen in Figure 1.2. Prior to the 2010 approved new Constitution, Kenya was divided into eight provinces subdivided into districts. After 2013 the country has seven regions subdivided into 47 counties.

Data from the three surveys undertaken by the Demographic and Health Surveys (DHS) program in 2003, 2008-09 and 2014 is used in this thesis. The surveys are referred to as KDHS03, KDHS08-09, and KDHS14, respectively. Census data is not considered, as it contains only summary birth histories. With some preprocessing, data from other sources can be included in the models presented here. The KDHS surveys provide enough data to produce estimates urban and rural area estimates at county/provincial level, but there is not enough data to produce yearly design-based mortality rates estimates at that level. Therefore the focus of this thesis is on national mortality rates.

The KDHS surveys have similar designs, described in short in the DHS reports (for example Kenya National Bureau of Statistics et al. (2015)) with more details in the DHS sampling manuals (for example MEASURE DHS and ICF International (2012)). The KDHS03 and KDHS08-09 are designed to be able to produce estimates for urban and rural areas for each of the eight former regions (except for Nairobi which does not have a rural area), while the KDHS14 is designed to produce estimates for urban and rural areas for each of the 47 counties (except for Nairobi and Mombasa which do not have rural

²www.r-inla.org



Figure 1.2: Map of Kenya, taken from Kenya National Bureau of Statistics et al. (2015), with permission from the DHS program.

areas). The sample size of the KDHS14 is 40300 households, whereas the size of the KDHS03 and the KDHS8-09 is approximately 10000 households.

The processes of planning, conducting, and processing the surveys are extensive and expensive. It includes several weeks of training, test questionnaires, translations, and workshops. They try to ensure high quality in questionnaires and performance to reduce the non-sampling error. The non-sampling error cannot be reduced simply by increasing the sampling size, whereas the sampling error can be reduced by increasing the sample size and is non-existent in a census. The uncertainties associated with survey-estimates often consider only the sampling error. The non-sampling error is difficult to quantify and often much larger than the sampling error. In the KDHS surveys, examples of sources of non-sampling error that are difficult to address are recollection bias, migration, changes in maternal mortality (HIV epidemics) and non-responses. In general, DHS surveys have a high response rate ($> 90\%$ for the used KDHSs). Everyone that slept in a sampled household the night before the interview is interviewed. More details about how the surveys are planned, conducted, and analyzed are given in the DHS reports (e.g. Kenya National Bureau of Statistics et al. (2015)).

The DHS program conducts surveys in over 90 countries, including several other countries in Africa. When investigating the effect of jointly modeling child mortality rates in neighboring countries, DHS data from the three most recent surveys undertaken in Tanzania and Uganda is used. These six surveys are similar to the ones conducted in Kenya;

Rate	1990	2015
NMR	29.7 [26.9, 32.8]	21.8 [17.7, 26.9]
U5MR	103.7 [97.6, 110.5]	48.7 [40.7, 58.7]

Table 1.1: NMR and U5MR in Kenya for 1990 and 2015, official estimates from IGME (2017). Presented as median, [lower level, upper level] for a 90% credible interval.

they are extensively planned for, tested, and efforts are made to reduce the non-sampling error.

Official estimates of NMR and U5MR in Kenya, obtained by the B3 model, is available for download at the website of the UN Inter-agency Group for Child Mortality Estimation (IGME)³. Official estimates for Kenya for key years are presented in Table 1.1. By evaluating the median values presented in Table 1.1, one can conclude that Kenya did not reach MDG target 4A.

This thesis will investigate how sensitive this conclusion is to the choice of model, and assess the probability of Kenya reaching SDG target 3.2. Using the KDHS data, models with autoregressive processes of first and second order as temporal effect are estimated and used to simulate new datasets. A simulation study investigates if the added model complexity of autoregressive processes helps describe and predict mortality rates better than less complex models with random walks. The best performing temporal effects are chosen to determine if including population size and GDP improves forecasts for child mortality rates in Kenya. Finally, it is investigated whether or not estimates and predictions for NMR and U5MR in Kenya are improved by borrowing strength from Uganda and Tanzania.

The organization of this thesis is as follows. Chapter 2 explains relevant theory, while Chapter 3 describes the methods used in this thesis in detail. The simulation study is presented in Chapter 4, in which the two most suitable temporal processes are determined. In Chapter 5 the models are applied to child mortality data from Kenya, the effect of including covariates is explored, and the most suitable temporal process determined. The effect of including information from neighboring countries is investigated in Chapter 6, before finally, the methods and findings are discussed in Chapter 7.

³https://childmortality.org/files_v22/download/UN%20IGME%20Child%20Mortality%20Report%202018.pdf

Background Theory

In this chapter, important concepts that are used in this thesis are outlined, and terminology and notation are introduced.

2.1 Complex survey methodology

Child mortality rates in low- and medium-income countries are calculated using survey data. In surveys, responses are collected from individuals in a subset of the full population of interest. The units from which information is collected are called observation units or simply units. When collecting information about child mortality, the observation units are women. The population of interest must be identified and can be all or some units in a given village, county, country or similar, depending on the scope of the survey. Observation units from the population are randomly chosen to be included in the sample. The sample is not independently drawn. The inclusion probability is determined by a predefined design, called the survey design. The choice of design is not arbitrary and can have a large impact on the time required and financial costs of conducting the survey as well as on the accuracy of the estimates generated from the survey. In this section, aspects of survey methodology relevant for handling DHS data are briefly outlined. More detailed books on the topic of survey methodology are written by for example Lohr (2010), Groves et al. (2004) and Lumley (2010).

In the most straightforward situation, the data is an independent identically distributed (iid) sample of an infinite population. The formulas used to obtain popular statistics from an iid sample from an infinite population are assumed well known to the reader. Let y_i , $i = 1, 2, \dots, n$ denote the data points, where n is the number of sampled data points. The expected value is simply the mean, $\frac{1}{n} \sum_{i=1}^n y_i$, and the variance can be obtained through the standard formula $\hat{V}(x) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$.

In a survey context, the population of interest is a real, finite population. A sample of the population is denoted by S , a set containing the indices of the units of the population included in the sample. Further, N denotes the number of units in the population while

n denotes the number of units in the sample. An important population quantity is the population total, t . The population total of y is $t_y = \sum_{i=1}^N y_i$. The true population total cannot be computed, as information is only available for units in the sample. Consequently, an estimator for the population total, T_y , must be used. As other interesting population quantities such as the population mean can be derived easily from the population total, the formulas in this section will focus on the population total.

The simplest form of sample design is a simple random sample (SRS), in which units are randomly selected from the population independently. Every unit has the same inclusion probability, so each unit is weighted equally. Thus, for an SRS $T_y = \frac{N}{n} \sum_{i \in S} y_i$. When calculating the estimated variance of T_y , $\hat{V}(T_y)$, a finite population correction (fpc) term must be included. The variance is a measure of the sampling error, and the fpc reflects the fact that the variance should decrease when n increases. The variance can be computed as $\hat{V}(T_y) = N^2 \left(1 - \frac{n}{N}\right) \frac{s^2}{n}$, where $\left(1 - \frac{n}{N}\right)$ is the fpc and s^2 is the sample variance of y . If $n = N$, the fpc ensures that the variance is 0. If the population is large compared to the sample size ($N \gg n$), the fpc is close to 1. In that case, there is no need to correct for the finite size of the population.

The DHS surveys must be able to produce estimates at regional levels, using limited resources, and are conducted by visiting selected households and performing face-to-face interviews. Therefore, SRS is not a practical design. An SRS could exclude several regions of a country and could require interviewers to travel for hours to visit one sampled household. Instead, the DHS sample designs combine stratification and cluster sampling. A stratified sample divides the population into mutually exclusive strata and includes observation units from every stratum, ensuring that estimates can be obtained for each stratum. Cluster sampling involves dividing the population into smaller groups often based on geographical location. Thereafter, units are selected only from a sample of the clusters, limiting the (geographical) spread of sampled observation units.

A population can be divided into strata based on any relevant characteristic such as age, race, education, etc. The DHS program needs to compute estimates for all administrative units within a country both for urban and rural areas. Thus, a natural stratification is regions and urban/rural areas. Let the population be divided into H strata, and let there be N_h units in total in stratum h . The simplest case is when an SRS of n_h units is sampled from each stratum h , denoted by S_h . Then $T_y = \sum_{h=1}^H T_h$, where $T_h = \frac{N_h}{n_h} \sum_{i \in S_h} y_{hi}$, where y_{hi} denotes unit i in stratum h . In other words, T_y is simply a sum of the estimated total in every stratum. Similarly, an estimator of the variance of T_y can be computed as a sum of the variance within each stratum, $V(T_y) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h}$, because strata are sampled independently. Here, $s_h^2 = \sum_{i \in S_h} \frac{(y_{hi} - \hat{y}_h)^2}{n_h - 1}$, where \hat{y}_h denotes the estimated mean of y in stratum h .

In addition to ensuring that units from every stratum are included in the sample, stratification can also be used to increase the precision of estimates by oversampling from some strata. See examples in Section 3.2 of Saghagen and Vik (2018). A potential problem with stratification is that the strata must be well-defined and their sizes known. For low- and medium-income countries, accurate registries of inhabitants might be difficult to obtain. The DHS estimates the size of the strata based on the last available census (censuses are done approximately every decade).

After stratifying the population, the DHS program applies a two-stage proportional-to-size sampling cluster design. The country is divided into small geographical areas called enumeration areas (EAs) using the last census. For example, in the KDHS14 survey, Kenya is divided into 96 251 EAs using the 2009 census. The EAs are what is called the primary sampling units (psus), and the DHS selects a sufficient number of psus in each stratum. An up-to-date list of the households is computed for every sampled psu. The households are what is called the secondary sampling units (ssus). In a one-stage cluster sample, all ssus within the sampled psus are included in the sample. In a two-stage cluster sample, only a subsample of the units in the sampled clusters are included in the sample. At the first stage, the DHS samples psus. The probability of sampling a psu is proportional to the estimated size of the psu. At the second stage, the DHS samples a fixed number of households from each sampled psu. In that way, the inclusion probability of a household is approximately proportional to the size of the enumeration area and approximately equal for all households in Kenya. All women aged 15-49 in a sampled household are interviewed.

Cluster sampling reduces the resources needed per sampled unit but increases the number of sampled units required to achieve a certain level of precision compared to an SRS. That is because units within a cluster often are more similar than units in different clusters. For example, all residents in one village might have access to prenatal care, while none of the residents in neighboring villages do. Consequently, the contribution of each unit is less valuable in a cluster sample than in an SRS. However, the benefits of reducing the geographical area, especially for the DHS who conducts face-to-face interviews in several thousand households, outweigh the cons.

The formulas for estimating the population total and its variance using a clustered sample will depend on the number of stages of clustering. The idea is to include the sampling weight from each level. Here, formulas for two-stage cluster samples are presented, and for simplicity for SRS sampling, not proportional-to-size sampling. Let there be N psus in a population, n of which are sampled. In the simplest form of cluster sampling, the n psus are sampled using SRS. Denote the sample of n psus by S . From each of the n clusters, an SRS of ssus is taken. Let S_i denote the sample of m_i ssus taken from psu n_i , $i \in S$, and let there be M_i ssus in total in psu n_i . The population total of quantity y can then be calculated as

$$T_y = \frac{N}{n} \sum_{i \in S} \sum_{j \in S_i} \frac{M_i}{m_i} y_{ij}.$$

Here, y_{ij} is an observation from ssu j in psu i . The corresponding variance is

$$V(T_y) = N^2 \left(1 - \frac{n}{N}\right) \frac{s_t^2}{n} + \frac{N}{n} \sum_{i \in S} \left(1 - \frac{m_i}{M_i}\right) M_i^2 \frac{s_i^2}{m_i},$$

as derived in Section 6.6 in the book by Lohr (2010). s_t^2 is the sample variance of the cluster totals, $s_t^2 = \frac{1}{n-1} \sum_{i \in S} (t_i - \hat{t}_y)^2$, where \hat{t}_y is the estimate of T_y . t_i is the estimate of the total in psu i , with estimator $T_i = \sum_{j \in S_i} y_{ij}$. s_i^2 is the sample variance in cluster i . Thus, the formula is simply a sum of the variance between the clusters and the variance within each cluster.

When several design strategies are combined, as for the DHS surveys, the formulas given in this section must be combined. As the DHS surveys are not SRS samples,

n/N must be replaced by inclusion probabilities and N/n must be replaced by sampling weights. The sampling weight is the inverse of the inclusion probability, and can be interpreted as the “number of units from the full population the sampled unit represents in the sample”.

In practice, the R package `survey` can be used to compute design-based estimates; more details are given in Chapter 3. To calculate NMR the function `svyratio()` is used. `svyratio()` calculates the ratio between two quantities, such as the number of neonatal deaths and number of live births, by first estimating the total of each quantity. Denote a birth by b_i and a let $d_i, i = 1, 2, \dots, N$ be 1 if the child died within the first 30 days of life and 0 if not. Let \hat{t}_b be the estimated total number of children born in the population, and \hat{t}_d be the estimated number of neonatal deaths. The variance of the ratio, $\hat{r} = \hat{t}_d/\hat{t}_b$ is derived in Example 9.2 in Lohr (2010) using Taylor expansions. The expression can be written as

$$V(\hat{r}) = \frac{V(\hat{t}_e)}{\hat{t}_b^2},$$

where \hat{t}_e is the estimated total of the residuals, $e_i = d_i - \hat{r}b_i$. When computing the U5MR the function `svyglm()` is used. `svyglm()` essentially performs weighted regression, in which the likelihood for each observation is weighted with the inclusion probability of that observation (Lumley et al. (2017)).

If yearly mortality rates are estimated in a way that accounts for the complex survey design, there is no need to make any further design considerations when moving forward with analysis. However, the design-based mortality rate estimates are computed for each year separately and do not borrow strength across time. It is desirable to take advantage of temporal dependency to produce more accurate estimates and forecasts. Therefore the design-based estimates are combined with a temporal model. Denote by η_t the logit transformation of the true mortality rate for year $t, t = 1, 2, \dots$. It is not possible to observe η_t , however the design-based estimates of η_t, y_t , are observed with noise with an estimated variance σ_t^2 . Following Mercer et al. (2015), assume $y_t \sim \mathcal{N}(\eta_t, \sigma_t^2)$. It is then possible to construct a temporal model for the underlying true mortality rate, η_t .

2.2 Temporal modeling

Assume the true mortality rate follows a trend, but that there is random variation around the trend. Let $\eta_t = \nu_t + a_t$, where ν_t is a deterministic term and a_t is a stochastic process. Assume that ν_t can be written as $\nu_t = \mu + \xi t + \mathbf{z}_t^T \boldsymbol{\beta}$, where μ is an intercept, ξ is a coefficient for a linear trend and \mathbf{z}_t is a vector of time-dependent covariates with coefficients collected in the vector $\boldsymbol{\beta}$. ν_t is important when forecasting mortality rates as it captures the overall trend, while a_t ensures reliable uncertainty estimates. It is expected that the stochastic variation is correlated in time, and there are many possible temporal processes to consider with different numbers of parameters. This thesis will explore the suitability of four simple stochastic processes: autoregressive models of first and second order and random walks of first and second order.

Autoregressive models (AR) is a popular class of temporal models. The output, a_t , of

an AR(p) model of the time series $\{a_t\}$, depends linearly on the past p time steps,

$$a_t = \sum_{i=1}^p \phi_i a_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad t = 1, 2, \dots \quad (2.1)$$

Here, ϕ_i $i = 1, \dots, p$ are lag-coefficients and ε_t is a noise term. Note that the indices are defined s.t. the first p terms of $\{a_t\}$ are $a_{-p+1}, a_{-p+2}, \dots, a_0$.

When performing modeling, it is desirable to have a stationary stochastic temporal process. An AR(1) process is stationary if $|\phi_1| < 1$. In general, an AR(p) process is stationary and has a unique stationary distribution if

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p \neq 0 \quad \text{for } |z| \geq 1. \quad (2.2)$$

Thus, an AR(2) process is stationary if $-1 < \phi_2 < 1 - |\phi_1|$.

As the conditions for stationarity changes with p and quickly become complicated, it is desirable to use a different parameterisation. The autocorrelation function (acf) and partial autocorrelation function (pacf) are often used. The acf, $\rho(\cdot)$, is defined as $\rho(h) = \gamma(h)/\gamma(0)$, $h \in \mathbb{Z}$, where $\gamma(\cdot)$ is the autocovariance function (acvf), defined as $\gamma(h) = \text{E}[a_{t+h}a_t]$, $t > 0$, $h \in \mathbb{Z}$. For a stationary time series, the acvf is only dependent on the lags h , not the time t . From the definition one gets that $\gamma(0) = \text{V}(a_t) := \sigma_a^2$, where σ_a^2 is the marginal variance of the temporal process. It can be shown that $\gamma(h) = \gamma(-h)$ and that for an AR(1) process, $\phi_1 = \rho(1)$.

The definition of the pacf is

$$\alpha(h) = \frac{\text{Cov}(a_t, a_{t-h} | a_{t-1}, \dots, a_{t-h+1})}{\sqrt{\text{V}(a_t | a_{t-1}, \dots, a_{t-h+1}) \times \text{V}(a_{t-h} | a_{t-1}, \dots, a_{t-h+1})}}, \quad t > 0, \quad h \geq 0.$$

For an AR(p) process, $\alpha(p) = \phi_p$ and $\alpha(h) = 0$ for $h > p$ (example 3.2.6 in Brockwell and Davis (2016)). The pacf only considers the direct correlation between two elements of a time series and from Equation (2.1) it is clear that only elements with lag $< p$ are directly correlated. The pacf is always in the interval $[-1, 1]$. Therefore the optimization problem is easier if one reparameterizes the AR(p) processes in terms of $\alpha(\cdot)$. Also, the posterior of the parameters will be better behaved.

Given ϕ -coefficients, the first p terms are required as initial values when simulating an AR process. The stationary distribution of the first p terms of an AR process is

$$a_{(-p+1):0} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}), \quad (2.3)$$

where $\mathbf{0}$ is a zero-vector of length p and

$$\mathbf{\Gamma} = \mathbf{\Gamma}(p) = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \gamma(2) & \gamma(1) & \dots & \gamma(p-3) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{bmatrix} \in \mathbb{R}^{p \times p}. \quad (2.4)$$

2.3 Bayesian hierarchical models

To understand Bayesian hierarchical models, one first needs to be familiar with Bayesian statistics. A quick introduction is given here, based on lecture notes from the course Bayesian Computation at Ecole Polytechnique Fdrale de Lausanne (EPFL) by Dehaene (2018). For a complete introduction to Bayesian statistics, the reader is referred to books such as "Bayesian Data Analysis" by Gleman et al. (2014).

When first introduced to statistics, one is taught the frequentist approach. The frequentist approach is based on the assumption that the population of interest follows some underlying model. The parameters of the underlying model, collected in the vector θ , are not random, but fixed and unobserved. The data is the only random quantity. Once the data is known, probability statements cannot be made. θ is estimated directly from data. For example, if the data, $\mathbf{y} = (y_1, \dots, y_n)$, is assumed to follow a normal distribution, $y_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \dots, n$, then $\theta = (\mu, \sigma^2)$ and μ is estimated by the sample mean and σ^2 by the sample variance.

An alternative approach is a Bayesian approach. In Bayesian statistics, the observed data, \mathbf{y} , is still assumed to follow some model with likelihood $\pi(\mathbf{y}|\theta)$. However, now, θ is a random quantity. One wants to incorporate a priori belief about θ into the model. This is done by assigning a prior distribution to θ , $\pi(\theta)$. Given the observations, \mathbf{y} , the model is updated using Bayes' theorem,

$$\pi(\theta|\mathbf{y}) = \pi(\theta) \frac{\pi(\mathbf{y}|\theta)}{\pi(\mathbf{y})}, \quad (2.8)$$

where $\pi(\mathbf{y})$ is a normalization constant, $\pi(\mathbf{y}) = \int_{-\infty}^{\infty} \pi(\theta)\pi(\mathbf{y}|\theta)d\theta$. The normalization constant ensures that $\int_{-\infty}^{\infty} \pi(\theta|\mathbf{y})d\theta = 1$. In Bayesian statistics, the parameters of the model, θ , are still random given the data, with probability density function $\pi(\theta|\mathbf{y})$. This is called the posterior. Thus, Equation 2.8 written in words is

$$\text{posterior} = \frac{\text{prior} \cdot \text{likelihood}}{\text{normalizing constant}}.$$

Bayesian hierarchical models are Bayesian models with several layers of latent structures. They are suitable when there is correlation within groups of the observations, as expected survey- or time-effect on mortality rates. "Hierarchical models can have enough parameters to fit the data well, while using a population distribution to structure some dependence into the parameters, thereby avoiding problems of overfitting" (p. 101, Gleman et al. (2014)). For the interested reader, Chapter 5 in the book by Gleman et al. (2014) explains Bayesian hierarchical models to greater detail and with several examples.

As given in the name, a Bayesian hierarchical model can be split into levels. Let y_i , $i = 1, \dots, n$, be the observed quantity or some transformation (e.g., logit) of it. The distribution of y_i is assumed conditional on latent variables, and the distribution of the latent variables are defined conditional on several parameters. In a three-stage model, those parameters are hyper-parameters that must be determined. In a hierarchical model with more levels, those parameters follow distributions that are conditional on a new set of parameters, and so on. Let \mathbf{x} denote the latent field and θ be the parameters. Three-stage hierarchical models have the following levels:

1. The observational level consists of the likelihood, $\mathbf{y}|\mathbf{x}, \boldsymbol{\theta} \sim \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$.
2. The latent level contains the distribution of the latent field, $\mathbf{x}|\boldsymbol{\theta} \sim \pi(\mathbf{x}|\boldsymbol{\theta})$.
3. The parameter level consists of the prior distributions of the parameters, $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

Let y_{ts} be the logit of the design-based estimated mortality rate for year t , and survey s . Assume that

$$y_{ts} \sim \mathcal{N}(\eta_{ts}, \sigma_{ts}^2), \quad \eta_{ts} = \mu + \xi t + a_t + \mathbf{z}_t^T \boldsymbol{\beta} + \delta_s, \quad t = 1, 2, \dots, T \quad s = 1, 2, 3, \quad (2.9)$$

where η_{ts} is the logit of the true mortality rate for year t and σ_{ts}^2 is the known estimated variance for year t from survey s . Further, μ is an intercept, ξ is a coefficient, a_t is an stochastic process, and \mathbf{z}_t is a vector of covariates with coefficients collected in $\boldsymbol{\beta}$. δ_s is an iid survey effect and data from three surveys, $s = 1, 2, 3$, are used. The latent field is thus $\mathbf{x} = (\mu, \xi, a_1, \dots, a_T, \boldsymbol{\beta}, \delta_1, \delta_2, \delta_3)$, where $\mathbf{a} = (a_1, a_2, \dots, a_T)$ is the temporal effect and T is the last year of interest. η_{ts} is the link between the observations and the latent field. $\boldsymbol{\theta} = (\boldsymbol{\rho}, \sigma_a^2, \sigma_\delta^2)$, where $\boldsymbol{\rho}$ is the coefficients from the temporal model, σ_a^2 is the marginal variance of the temporal model and σ_δ^2 is the variance of the survey-effect.

The models developed in this thesis belong to the class of latent Gaussian models (LGMs). LGMs is a large class of hierarchical models, where the prior distribution on the latent field must be Gaussian conditional on the parameters. In other words, every model component of the latent field must be independently Gaussian conditional on the model parameters. Using this, the model for mortality rates suggested in Equation (2.9), can be formulated as an LGM as given in Table 2.1. The high variance of μ , ξ and β_j , $j = 1, \dots, p$, indicates that the effects are considered fixed.

Observation level
$\pi(\mathbf{y} \boldsymbol{\eta}, \boldsymbol{\sigma}^2) = \prod_{i \in I} \pi(y_i \eta_i, \sigma_i^2)$
Latent level
$\pi(\mu) = \mathcal{N}(0, \infty)$
$\pi(\xi) = \mathcal{N}(0, 1000)$
$\pi(\mathbf{a} \boldsymbol{\rho}, \sigma_a^2) = \mathcal{N}(\boldsymbol{\rho}, \sigma_a^2)$
$\pi(\beta_j) = \mathcal{N}(0, 1000), j = 1, \dots, p$
$\pi(\boldsymbol{\delta} \sigma_\delta^2) = \mathcal{N}(0, \sigma_\delta^2)$
Parameter level
$\pi(\boldsymbol{\rho}), \pi(\sigma_a^2), \pi(\sigma_\delta^2)$

Table 2.1: Model from Equation (2.9) as a LGM. \mathbf{y} is the observations, $\boldsymbol{\eta}$ is the mean, and $\boldsymbol{\sigma}^2$ is the variance. I is the set of indices of the observations. μ is the intercept, ξ is a linear trend in time. \mathbf{a} , $\boldsymbol{\rho}$ and σ_a^2 are elements of the stochastic temporal effect. $\beta_j, j = 1, \dots, p$ are coefficients of additional p covariates z_j . $\boldsymbol{\delta}$ is the survey effect and σ_δ^2 is the variance of the survey effect.

Using the notation from Rue et al. (2016), LGMs can be written on the form

$$\eta_i = \mu + \sum_{j=1}^J \beta_j z_{ij} + \sum_{k=1}^K f_{k, j_k(i)}, \quad (2.10)$$

where μ is the intercept, z are fixed effects with covariates β_j and there are J fixed effects. Further, $\mathbf{f}_k = (f_{k,1}, \dots, f_{k,m_k})$, where m_k is the dimension of \mathbf{f}_k , are specific Gaussian processes and there are K such processes. For the model in Equation (2.9), the fixed effects are t and the covariates collected in z . The \mathbf{f}_k s are model components, in which element $j_k(i)$ contributes to the i th linear predictor. For the model in Equation (2.9), the model components are the stochastic temporal process, $\mathbf{f}_1 = \mathbf{a}$, and the survey effect, $\mathbf{f}_2 = \boldsymbol{\delta}$, so $K := 2$. $m_1 = T$, where T is the number of years one wants to estimate mortality rate and $m_2 = 3$, as data from three different surveys are used. This notation is useful to keep in mind when estimating the model from Equation 2.9 (Section 3.2).

LGMs are a class of flexible models that are used in a wide range of studies such as malaria studies (Bhatt et al. (2015)) and studies of traffic pollution and hospital admissions in London (Halonen et al. (2016)). For an extensive list over studies using LGMs, see the introduction of Rue et al. (2016) and Section 1.2 of Rue et al. (2009). LGMs allow for the combination of spatial and temporal effects and covariates. An LGM is a suitable choice for child mortality rates as the effects are assumed to be additive, and the resulting models are easily interpreted. However, in other situations where one wishes to model, for example, a non-linear dynamic system, or multiplicative or non-Gaussian effects, an LGM might not be suitable.

Once a Bayesian model is defined in terms of the likelihood and (possibly several stages of) priors, the posterior should be computed. However, it is often not possible to compute the posterior directly so an approximation method must be used. There is a wide range of methods to choose from, and the choice of method is not trivial. Some popular and relevant methods are presented in the next section.

2.4 Bayesian inference

The posteriors of the parameters of a Bayesian model can be used to answer several statistical questions about the model. However, for most Bayesian models, the posteriors cannot be computed analytically. Therefore, there exist several approximation methods. Popular approximations are the deterministic Laplace approximation and slower, more flexible Markov chain Monte Carlo (MCMC) approximation samples. The Laplace approximation is explained in for example Section 2.5 of Rue et al. (2016), while MCMC methods are explained in for example part three of Gleman et al. (2014). Software such as Stan (Carpenter et al. (2017)) can be used to do Bayesian inference, using MCMC methods to approximate the posterior. Recently, a method to obtain a deterministic approximation of the posterior of an LGM, the integrated nested Laplace approximation (INLA, Rue et al. (2009)) and the R-INLA software (Lindgren and Rue (2015)), were developed .

The Laplace approximation is based on Taylor expansions. The idea is that the normal distribution can approximate the unnormalized posterior. Let \mathbf{y} be the observed data and

$\boldsymbol{\theta}$ be a vector collecting the parameters of the model, then

$$\begin{aligned}\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) &\approx \exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T\mathbf{H}(\boldsymbol{\theta}-\boldsymbol{\mu})\right], \\ \boldsymbol{\mu} = \boldsymbol{\theta}^* &= \operatorname{argmax}_{\boldsymbol{\theta}}\left[\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})\right], \\ \mathbf{H} &= -\operatorname{Hessian}(\log \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}))|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*},\end{aligned}\tag{2.11}$$

where $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is the unnormalized posterior, $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) = \pi(\boldsymbol{\theta})\pi(\mathbf{y}|\boldsymbol{\theta})$. The Laplace approximation is easy to obtain, as only a maximum, $\boldsymbol{\theta}^*$, and a (inverse) Hessian with respect to $\boldsymbol{\theta}$, \mathbf{H} , is required. However, the Laplace approximation cannot capture multimodal-distributions and may fail to capture global properties of the posterior (Dehaene (2018)).

MCMC sampling methods are often a preferred alternative to the Laplace approximation because they can be made arbitrarily accurate. Values of θ are sampled from a Markov-chain and accepted or rejected according to some acceptance rule to create a sample approximating the posterior. However, MCMC algorithms have to generate large samples to produce highly accurate estimates. For LGMs MCMC methods are often painfully slow, see Section 1.4 of Rue et al. (2009). The software `RStan` can be used to approximate the posterior of any Bayesian models, including LGMs, using the MCMC No-U-Turn sampler (Hoffman and Gelman (2014)). However, if the model of interest is an LGM, there are a few disadvantages when using `RStan`. The user has to write specific code for all different model components and make other adjustments. Further, when using MCMC methods, obtaining rough estimates is relatively easy, but “an additional correct digit requires 100 times more computationally power” (Section 1.5, Rue et al. (2009)).

A more efficient way to estimate the posterior of LGMs is developed, namely the integrated nested Laplace approximation (INLA). INLA was introduced by Rue et al. (2009) and quickly became a popular method. INLA assumes two basic properties for the LGMs. First, it is assumed that even though the dimension of the latent field, $n = \dim(\mathbf{x})$, might be large ($10^2 - 10^5$), the latent field admits conditional independence properties. Thus the latent field is a GMRF with a sparse precision matrix, which makes it possible to use efficient numerical methods. Secondly, it is assumed that the number of hyperparameters m , $m = \dim(\boldsymbol{\theta})$, is low, “typically 2-5, but not exceeding 20” (Section 2.1, Rue et al. (2016)).

The goal is to compute an accurate approximation of the posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$, the marginal posteriors $\pi(\theta_j|\mathbf{y})$ for $j = 1, 2, \dots, m$, and the marginal posteriors of the latent field $\pi(x_i|\mathbf{y})$ for $i = 1, 2, \dots, n$. Let $\tilde{\pi}$ denote an approximated distribution. As stated in Martins et al. (2013), the approximated posterior marginals of interest returned by INLA have the following form

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}^{(k)}, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\mathbf{y})\Delta\boldsymbol{\theta}^{(k)}\tag{2.12}$$

and

$$\tilde{\pi}(\theta_j|\mathbf{y}) = \int I(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j},\tag{2.13}$$

where “ $\tilde{\pi}(x_i|\boldsymbol{\theta}^{(k)}, \mathbf{y})$ are the density values computed during a grid exploration on $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ ” (Martins et al. (2013), p. 6), $\Delta\boldsymbol{\theta}^{(k)}$ is a weight depending on integration design, I is an

interpolant constructed from $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ and $\boldsymbol{\theta}_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m)$. The INLA method can be divided into three main steps. Step 1: Propose an approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$, $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$. Step 2: Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ for $i = 1, 2, \dots, n$. Step 3: Explore $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ on a grid and use it to integrate out $\boldsymbol{\theta}$ in Equation (2.12) and $\boldsymbol{\theta}_{-j}$ in Equation (2.13).

From Bayes theorem, one gets that

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\boldsymbol{\theta}, \mathbf{x}, \mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} = \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x}, \mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}. \quad (2.14)$$

The numerator is known, but the denominator is not. Step one of INLA obtains an approximation of $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, using an Gaussian approximation as explained in Section 2.2 of Rue et al. (2009). Using $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ in place of $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ in Equation (2.14), $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is obtained.

There are three different ways of approximating $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$, all described in Section 3.2 of Rue et al. (2009). One option is to use a Gaussian approximation, namely the marginals of $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. The second option is a natural improvement of the Gaussian approximation, namely the Laplace approximation (LA). However, the LA is too computationally demanding. Therefore, two aspects of the LA is improved. Firstly, the optimization step is avoided by ‘‘approximating the modal configuration’’ (p. 12, Rue et al. (2009)),

$$\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta}) = \mathbb{E}_{\tilde{\pi}_G}(\mathbf{x}_{-i}|x_i).$$

Second, it is assumed that only points close to x_i have an effect on the marginal of x_i . Therefore only points in a region around x_i are considered. For more details, see Section 3.2.2 in Rue et al. (2009). The third option is called the simplified Laplace approximation (SLA). The SLA introduces a skewness correction in each marginal, see details in Section 3.2.3 in Rue et al. (2009).

$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is explored in order to obtain $\tilde{\pi}(\theta_j|\mathbf{y})$ and $\tilde{\pi}(x_i|\mathbf{y})$. Three steps are given in Section 3.1 of Rue et al. (2009). The first step is to locate the mode of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, $\boldsymbol{\theta}^*$, by optimizing $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$. The second step is to compute the negative Hessian, $\mathbf{H} > 0$, at $\boldsymbol{\theta}^*$. Then, let $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$ and let $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ be the eigen-decomposition of $\boldsymbol{\Sigma}$. Define $\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is a Gaussian density. The third step computes $\tilde{\pi}(\theta_j|\mathbf{y})$ and $\tilde{\pi}(x_i|\mathbf{y})$, using the given z-parameterization. As described in Section 3.1 of Rue et al. (2009), a grid is constructed and $\log(\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}))$ is computed for all points in the grid, and $\tilde{\pi}(x_i|\mathbf{y})$ can be computed using Equation (2.12). Finally, an interpolant is constructed using the points in the grid and $\tilde{\pi}(\theta_j|\mathbf{y})$ is computed using Equation (2.13). More details about the algorithms actually used are given in Martins et al. (2013).

INLA is implemented in the R-package INLA, which is used by among others SUMMER. Detailed information about INLA can be found on the website www.r-inla.org. INLA is easy to use, allows to user to include several different model components without having to write complicated code, and is fast. To change the temporal effect in the model presented in Equation (2.9), only minor adjustments have to be made. Furthermore, it is easy to estimate models with multivariate temporal processes.

2.5 Multivariate temporal models

Several state-of-the-art methods for estimating child mortality rates do so jointly for several countries. Section 1.1 discusses some disadvantages of some of those models, but it is reasonable to assume that trends in child mortality rates are similar in similar countries. Thus, using available information from more than one country should improve the estimates and forecasts. The model developed so far in this chapter can easily be extended to become a multivariate temporal model.

Assume that the mortality rates of k countries are correlated and that it is desired to model them together. For simplicity, assume that the only covariate of interest is a linear trend in time. The model given in Equation (2.9) can then be re-written as $y_{cts} \sim \mathcal{N}(\eta_{cts}, \sigma_{cts}^2)$, where $c = 1, 2, \dots, k$ is an index for the country, $t = 1, \dots, T$ is a time-index and T is the number of years of interest, $s = 1, \dots, S$ where S is the total number of surveys used and

$$\eta_{cts} = \mu_c + \xi_c t + a_{ct} + \delta_s. \quad (2.15)$$

The intercept and linear trend are country-specific, and the survey effect is unique for every survey from which data is used. The survey effect is not modeled as country-specific because the model should be scalable to consider several countries, and there are too few observations within a country to be able to estimate variance reliably. The temporal effect, a_{ct} , is correlated across countries. In other words, it is a multivariate temporal process.

A simple form of multivariate temporal models are called vector autoregressive (VAR) models and are described in Chapter 11 of Zivot and Wang (2006). Collect all a_{ct} , $c = 1, \dots, k$, in a k dimensional vector $\mathbf{a}_t = (a_{1,t}, a_{2,t}, \dots, a_{k,t})^T$, then a VAR(p) model can be written as

$$\mathbf{a}_t = \sum_{i=1}^p \Phi_i \mathbf{a}_{t-i} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \quad t = 1, 2, \dots,$$

where Φ_i , $i = 1, \dots, p$ are $k \times k$ matrices of coefficients and $\boldsymbol{\varepsilon}_t$, $t = 1, \dots, T$ is a $k \times 1$ vector of error terms. For this thesis Φ_i , $i = 1, 2, \dots$ are restricted to be on the form

$$\Phi_i = \begin{bmatrix} \phi_i & 0 & 0 & \dots & 0 \\ 0 & \phi_i & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \phi_i \end{bmatrix} \in \mathbb{R}^{k \times k}, \quad i = 1, 2, \dots, p,$$

and the covariance matrix of the error terms is on the form

$$\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix} \in \mathbb{R}^{k \times k},$$

where σ_ε^2 is marginal variance and ρ is a correlation coefficient, later referred to as ρ_{group} . The covariance structure described by $\boldsymbol{\Sigma}_\varepsilon$ is called exchangeable.

The marginal variance of each time series is still $\gamma(0) = \sigma_a^2$ and is identical for all countries. Let Γ be the covariance matrix of each time series as defined in Equation (2.4). Then the covariance matrix of the multivariate time series is

$$\Sigma = \Gamma \otimes \Sigma_\varepsilon,$$

where \otimes is the Kronecker product. Further, if $Q = \Sigma^{-1}$, $Q_\Gamma = \Gamma^{-1}$ and $Q_\varepsilon = \Sigma_\varepsilon^{-1}$ then

$$Q = Q_\Gamma \otimes Q_\varepsilon.$$

Note that even though Σ is a dense matrix, Q_Γ is sparse and Q_ε is in low dimension so Q is still sparse and computations involving Q can be made efficient.

As with the univariate temporal processes, multivariate random walks are just a special case of VAR. The multivariate RW(1) have Φ_1 equal to the identity matrix while for the multivariate RW(2) Φ_1 is a diagonal matrix with 2 on the diagonal and Φ_2 is a diagonal matrix with -1 on the diagonal.

The model in Equation (2.15) is still an LGM and can easily be fitted using INLA. As INLA fits LGMs in a matter of seconds, the user is enabled to run simulation studies with several different models and compare them. To be able to assess the performance of models and determine which model is most appropriate for estimation and forecast, some assessment criteria and scoring rules must be established. The next section outlines some popular choices that are used in this thesis.

2.6 Model assessment and scoring rules

The main motivations for temporal modeling of child mortality rates are explaining the historical and current mortality rates and predicting future mortality rates. To assess the explanatory strength of a model, diagnostics such as log marginal likelihood (MLIK), Watanabe-Akaike information criterion (WAIC), mean absolute error (MAE), and mean squared error (MSE) are of relevance. MLIK and WAIC can be computed for all fitted models and be used to compare how well the different models explain the observed data. MAE and MSE can also be used for model comparison, and the value of these measures gives direct information about a model's explanatory strength as well. In predictive modeling, the goal is to develop a model that can produce accurate predictions of future observations. It is interesting to compute the MAE, the MSE, and the continuous rank probability score (CRPS) for predictions to assess the predictive strength of a model.

In Bayesian models, the marginal likelihood is the probability of the data, given the model. The marginal likelihood is obtained by integrating out the parameters of the model. For a given model M , described by parameters are collected in θ , the marginal likelihood is

$$\pi(\mathbf{y}|M) = \int \pi(\mathbf{y}|\theta, M)\pi(\theta|M)d\theta,$$

for data \mathbf{y} . The model of best fit should be the model with the highest marginal likelihood, or equivalently the highest log marginal likelihood. INLA computes reliable estimates of the marginal likelihood (Hubin and Storvik (2016)).

It is also interesting to investigate the leave-one-one predictive density, estimated by the Watanabe-Akaike information criterion (Watanabe (2010)), also called the widely applicable information criterion. It is a generalization of the Akaike information criterion (AIC), which is assumed known to the reader. WAIC estimates the expected log pointwise leave-one-out predictive density and is computed as

$$\text{WAIC} = \sum_{i=1}^n \log(\pi(y_i|\mathbf{y})) - \sum_{i=1}^n V\left(\log(\pi(y_i|\boldsymbol{\theta}))\right) \quad (2.16)$$

(Vehtari and Gelman (2017)), where n is the number of datapoints. The variance can be estimated using the posterior of the log predictive density, see Section 2.2 of Vehtari and Gelman (2017). The second term in Equation (2.16) is the effective number of parameters, \hat{p}_{WAIC} , and can be regarded as a measure of the complexity of the model. Similarly to AIC, the model of best fit w.r.t. WAIC is the model with the lowest WAIC.

MAE and MSE are two well-known performance measures. Let the simulated mortality rate be y_i $i = 1, 2 \dots n$, while the predicted mortality rate is \hat{y}_i $i = 1, 2 \dots n$, and let there be n predicted data points for which the true simulated mortality rate is known. Then the MAE is simply

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

and the MSE is

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

The value of MAE/MSE gives the average prediction error and the best performing w.r.t MAE and MSE are the models with lowest MAE/MSE.

However, in this thesis, the observations themselves are noisy with varying variances, and the predictions are also probabilistic. That is not taken into account by the MAE and MSE. CRPS, discussed in Section 4.2 in Gneiting and Raftery (2007), considers not only the point observations and point predictions but also the associated uncertainty. CRPS is a proper scoring rule, meaning that the expected score for an observation drawn from the distribution F is minimized if the probabilistic forecast is F , rather than $G \neq F$ (Section 1, Gneiting and Raftery (2007)). In contrast, MAE and MSE are not proper scoring rules, meaning that given a large number of observations and a set of models including the true model, the model selected based on the lowest MAE/MSE might not be the true model.

Given a probabilistic forecast with cumulative distribution function $F(y)$, the CRPS is defined as

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 dy, \quad (2.17)$$

where x is the true value and $\mathbb{1}\{\cdot\} = 1$ if the statement inside the brackets is true and zero otherwise. With the definition given in Equation (2.17), the CRPS is always positive and a CRPS close to zero indicates that the approximated distribution is close to the true distribution. A predicted value with low bias and high variance would have a higher CRPS value than a predicted value with a higher bias but much lower variance.

When the predictive distribution is Gaussian, one obtains

$$\text{CRPS}(\mathcal{N}(\mu, \sigma^2), x) = -\sigma \left[\frac{1}{\sqrt{\pi}} - 2\phi\left(\frac{x - \mu}{\sigma}\right) - \frac{x - \mu}{\sigma} \left(2\Phi\left(\frac{x - \mu}{\sigma}\right) - 1 \right) \right]. \quad (2.18)$$

Here, ϕ is the probability density function and Φ the cumulative probability function of the normal distribution. If the true value is not known, which is the case for child mortality rates, the CRPS can be calculated as

$$\text{CRPS}(F_{\text{approximated}}, F_{\text{true}}) = \int_{-\infty}^{\infty} (F_{\text{approximated}}(y) - F_{\text{true}}(y))^2 dy. \quad (2.19)$$

If the true and forecasted value are point-values (i.e. no uncertainty), the CRPS reduces to the MAE.

Chapter 3

Methods

3.1 Producing yearly design-based mortality rate estimates

To access the DHS data, one has to apply the DHS program's website¹. The data is coded according to the DHS recode manuals, which are updated at regular intervals. The three Kenya surveys used in this thesis are coded according to different recode manuals. KDHS03 follows DHS (2008), KDHS08-09 follows DHS (2012) and KDHS14 follows DHS (2013). The surveys from Uganda and Tanzania used in Chapter 6 also follow these three recode manuals. Even though the surveys follow different recode manuals, the variables of interest are coded in almost the same way in all data sets, and are presented in Table 3.1. *Century month code* is given as the number of months since January 1900.

v002	household number (secondary sampling unit)
v005	sample weight times 10 ⁶
v008	century month code of date of interview
v021	cluster id (primary sampling unit)
v023	stratification
v024	place of residence (region)
v025	type of residence (urban/rural)
b3	century month code for the date of birth of the child
b6	age at death of the child

Table 3.1: Relevant responses from the DHS recode manuals.

In Table 3.1 there are three responses that contain information about the stratification: v023, v024 and v025. After a closer examination of the data, it is found that for some surveys v023 correctly defines the strata, while for others only the combination of v024 and v025 gives the correct stratification. v023 is used as stratification variable for the

¹<https://dhsprogram.com/Data/>

surveys following DHS (2013) (KDHS14), while v024 and v025 is used for the surveys following DHS (2008) (KDHS03) and DHS (2012) (KDHS08-09).

The neonatal mortality rate is the probability of dying within the first month of life for live births. The definition of neonatal mortality rate varies. The UN considers the deaths within the first 28 days of life. Page 8.5 and on-wards in Croft et al. (2018) outlines an alternative to calculating simply the proportion of eligible children who died within the specified time frame. Other sources, such as MEASURE (2019), make a distinction between the first seven days and the next 14 days. They claim that the risk of dying is not constant during the first month of life and that it should be reflected in the calculation. However, the data used in this thesis is not accurate enough to be used to classify the age of death at such a fine scale. This thesis defines NMR according to Definition 3.1.1. The first 30 days of life is used to be consistent with the methods used by the DHS program when producing their reports (e.g., Kenya National Bureau of Statistics et al. (2015)).

Definition 3.1.1. The neonatal mortality rate is the proportion of children who died “at age 0-30 days (also includes deaths reported at age zero month)” (page 8.5, Croft et al. (2018)), given that the children were born 30 days or more before data collection.

To calculate the neonatal mortality rate, all children who died within 30 days after birth have to be flagged. In the KDHS surveys, the date of birth is given in month and year. To ensure that children who have lived less than 30 days are not included in the analysis, all children who are born in the same month as the interview, or in the month before the interview, are removed from the dataset. Observations for which the child was born at the end of one year and die in the next year are counted as an observation in the birth year.

NMRs are calculated independently for each year and each survey. The design of a survey can easily be encoded in a `survey.design` object, using the R package `survey`, and the information about the design included in the dataset. Once a design-object is obtained, the function `svyratio()` is applied to obtain the NMR. An example of how to do this is given in Listing 3.1. In Listing 3.1 the variable `data` is a dataframe with columns `neonatal_death`, `ones`, `time`, `v002`, `v005`, `v021`, `v023`, `v024` and `v025`, with the final six given in Table 3.1. `neonatal_death` is one if the child died within the first 30 days of life and zero otherwise, while `ones` is simply a vector of ones. `time` is the year the observation belongs to and the variable `which.time` is the year for which one wishes to estimate the NMR.

Listing 3.1: Creating a `survey.design` object and computing NMR for one year.

```
data$rweights=data$v005/1000000
design_object=svydesign(ids=~v021+v002, weights=~rweights, strata=~
  v024+v025, data=data)
tmp = subset(design_object, (time == which.time))
logit_NMR=svyratio(~neonatal_death, ~ones, tmp)
```

The arguments of the `svydesign()` function are as follows: `ids` is a formula defining the stages of clusters, `weights` are the sample weights of the observations, and `strata` is a formula defining the stratification. Listing 3.1 defines a `survey.design` object for KDHS03 or KDHS08-09. For KDHS14 `v023` defines the strata. The arguments of `svyratio()` are a formula for the numerator, a formula for the denominator and a

survey.design object. For more information about the arguments of `svydesign()` and `svyratio()`, see the documentation².

The function `svyratio()` returns the estimated value and the estimated variance of NMR. It is desirable to obtain the mean and variance of $\text{logit}(\text{NMR})$. When transforming data using a function g , $y = g(\tilde{y})$, the expected value and variance of y is approximated using the Delta method, which is based on Taylor expansions evaluated at the expected value of y , $g(\tilde{y}) = g(\text{E}(y)) + (y - \text{E}(y))g'(\text{E}(y))$. The expected value is approximated simply as $\text{E}(y) \approx g(\text{E}(\tilde{y})) = g(\mu)$, where μ is the expected value of \tilde{y} . The approximation of the variance becomes

$$V(y) \approx (g'(\mu))^2 V(\tilde{y}).$$

The Delta method is used to compute the approximate expected value and variance of $\text{logit}(\text{NMR})$ from the value and variance of NMR. The NMR and $\text{logit}(\text{NMR})$ with corresponding 95% credible intervals are presented in Figure 3.1.

When calculating the NMR, it is assumed that the risk of dying is constant during the first 30 days of life. However, the same assumption does not hold for the first five years of life. To illustrate this point, the piece-wise constant discrete hazard function for under-5 mortality in Kenya in 2005, calculated using KDHS14, is displayed in Figure 3.2. In Figure 3.2 the first five years of life is divided into months $[0, 1)$, $[1, 12)$, $[12, 24)$, $[24, 36)$, $[36, 48)$, $[48, 60)$ and it shows that the risk of dying changes during the first five years of life. Consequently, the U5MR is calculated using discrete time survival analysis as described in the article by Mercer et al. (2015) and its supplementary material. The same method is used by others as well, for example, Wakefield et al. (2018).

The discrete hazard function is defined as $h(x) = \text{Pr}(\text{dying before } x + 1 \mid \text{lived until } x)$. The U5MR is the probability of not dying in any of the first 60 months of life, and can thus be calculated as $1 - \prod_{a=0}^{59} (1 - h(a))$. Computing the U5MR in this way requires large amounts of data, as the probability of dying in each of the first 59 months of life must be calculated. To avoid this problem, a piece-wise constant discrete hazard function is assumed, as illustrated in Figure 3.2. Let the piece-wise constant discrete hazard be defined as ${}_n q_x = \text{Pr}(\text{dying before } x + n \mid \text{lived until } x)$. Following Mercer et al. (2015), the period from 0 – 5 years is split into $J = 6$ intervals, which are months $[0, 1)$, $[1, 12)$, $[12, 24)$, $[24, 36)$, $[36, 48)$, $[48, 60)$. The U5MR is then calculated as

$${}_{59}q_0 = 1 - \prod_{j=1}^J (1 - {}_{n_j}q_{x_j}), \quad (3.1)$$

where $(x_1, \dots, x_6) = (0, 1, 12, 24, 36, 48)$ and $(n_1, \dots, n_6) = (1, 11, 12, 12, 12, 12)$. The hazard can be estimated using logistic regression for each of the J intervals, by using J factors for age intervals. Using logistic regression to estimate a hazard function is a standard method, described by for example Efron (1988).

Before estimating the hazard using logistic regression, the data have to be prepared. Each row in the dataset contains information about one child, including when it was born and when it died. When calculating the probability of dying in each age-interval one have to consider the total number of child-months at risk and the total number of deaths within

²<http://r-survey.r-forge.r-project.org/survey/>

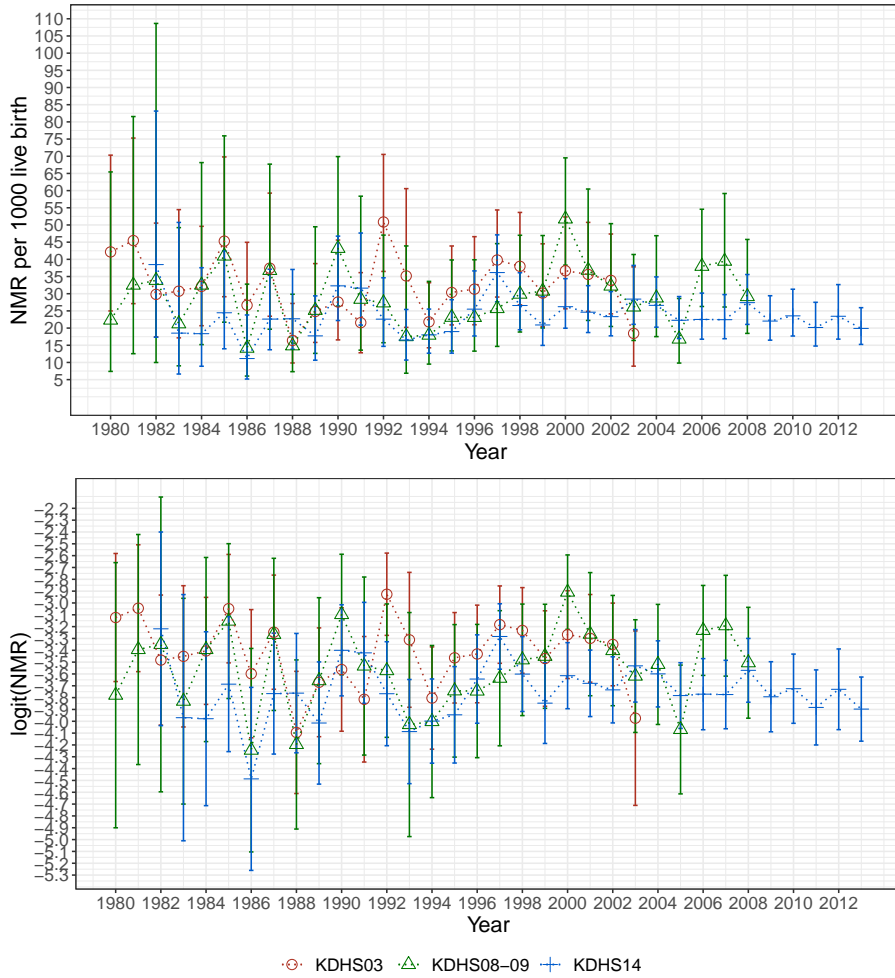


Figure 3.1: Design-based estimates of NMR per 1000 live birth and $\text{logit}(\text{NMR})$ with 95% credible interval, calculated separately for each year from 1980 to 2013, for each KDHS survey.

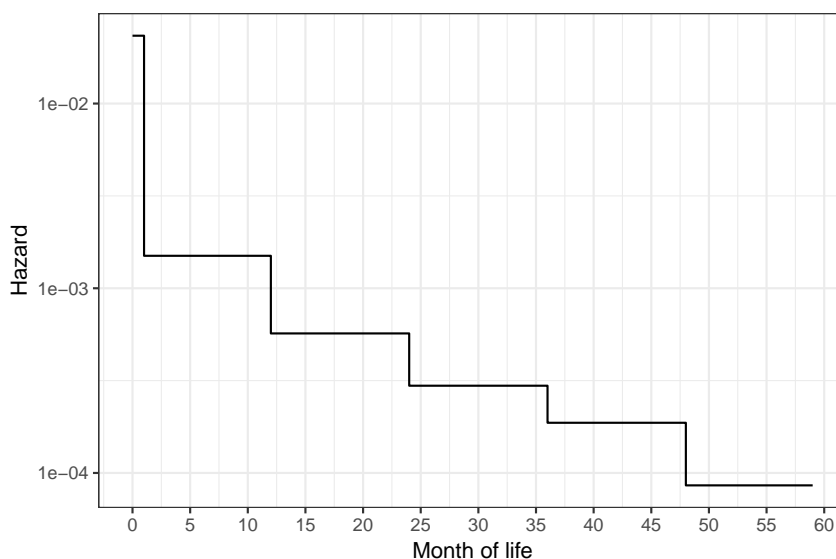


Figure 3.2: Piece-wise constant discrete hazard function for the first five years of life in Kenya in 2005, estimated using data from KDHS14. The first five years of life is divided into months $[0, 1)$, $[1, 12)$, $[12, 24)$, $[24, 36)$, $[36, 48)$, $[48, 60)$.

each interval. Therefore, let each child be represented at up to 60 rows, one for each month up to and including the month of death or the 59th month of life (the first month of life is counted as month 0). Each row has an outcome of 0 if the child did not die that month and 1 if it did, so each row represents one child-month. When estimating the hazard function, logistic regression is performed with child-months as the number of trials. In this thesis, yearly U5MR is calculated, so only child months belonging to year i , $i = 1980 \dots 2013$, is considered when calculating the U5MR of year i .

The function `svyglm()`, a function that can estimate generalized mixed models while accounting for the survey design, is used to perform the regression. `svyglm()` requires a `survey.design` object. An example of how to use `svyglm()` is given in Listing 3.2. The variable `data` is a dataframe with columns `age`, `died`, `time`, `v002`, `v005`, `v021`, `v023`, `v024` and `v025`, with the final six given in Table 3.1. Note that Listing 3.2 defines a `survey.design` object corresponding to KDHS03 or KDHS08-09. The strata are defined by `v023` in KDHS14. `age` is a factor indicating the age interval and `time` states the year the observation was made. `died=1` if the child died that month and 0 otherwise. The variable `which.time` is the year for which one wishes to estimate the U5MR. The argument `family` in `svyglm()` defines the distribution used in the model. The quasi-binomial distribution differs from the binomial distribution only as the dispersion parameter is not fixed at one. Thus, the quasi-binomial distribution can model overdispersion. For more on overdispersion and quasi-likelihood, see Carruthers et al. (2008), page 9 and onwards.

Listing 3.2: Estimating U5MR.

```
data$rweights=data$v005/1000000
design_object = svydesign(ids =~v021+v002,
  strata = ~v024 +v025, weights = ~rweights, data = data)
tmp = subset(design_object, (time == which.time))
glm.ob = svyglm(died ~ (-1) + factor(age),
  design = tmp, family = stats::quasibinomial, maxit = 50)
```

Using the fact that the coefficients, β , obtained by the logistic regression is $\beta_j = \text{logit}({}_j q_x)$ for $x \in [x_j, x_j + n_j)$, ${}_n q_x$ can be estimated by ${}_n \hat{q}_{x_j} = 1 - (1 - \text{expit}(\hat{\beta}_j))^{n_j}$, $j = 1, 2, 3, 4, 5, 6$. The estimated ${}_n q_x$ s are then inserted into Equation 3.1 to obtain the U5MR. The corresponding variance is obtained using the Delta-method, using the variance of β_j $j = 1, \dots, 6$, $\beta_j = \text{logit}({}_j q_x)$ and Equation (3.1). The covariance matrix of β is estimated by `svyglm()`, and since the probability of dying in each time interval is independently estimated, only the variance of β_j $j = 1, \dots, 6$ is considered. Since U5MR is estimated directly, the Delta-method is applied once again to obtain the expected value and variance of `logit(U5MR)`.

The obtained estimates for U5MR and `logit(U5MR)` in Kenya in 1980-2013 are presented in Figure 3.3. Separate estimates are computed for each year for each of the three KDHS surveys. For more details the reader is referred to the article by Mercer et al. (2015), the corresponding supplementary material and the package `SUMMER` in R. Be aware that the functions implemented in `SUMMER` cannot be directly applied to calculate yearly U5MR for Kenya, as they compute estimates on a regional level for five-year periods.

3.1 Producing yearly design-based mortality rate estimates

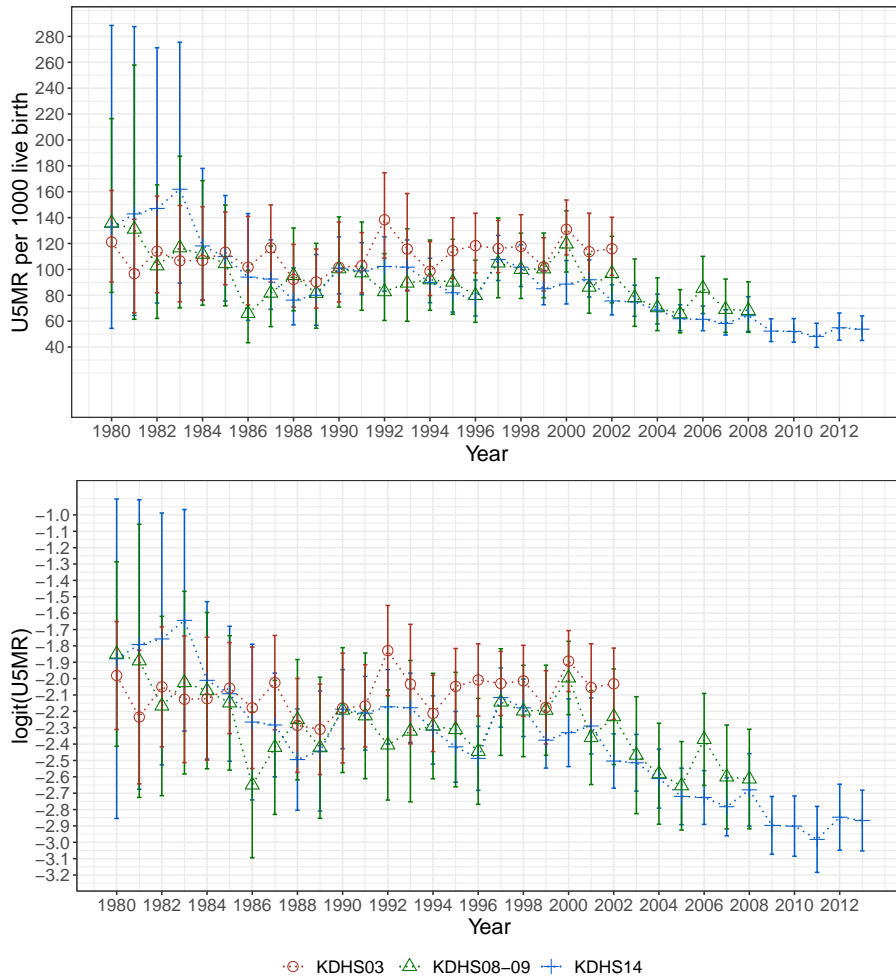


Figure 3.3: Design-based estimates of U5MR and $\text{logit}(U5MR)$ with 95% credible interval, calculated separately for each year from 1980 to 2013, and each KDHS survey.

3.2 Estimating models using `inla()`

Recall that the logit of the design-based estimate of the mortality rates is denoted by y_{ts} , $t = 1, 2, \dots, s = 1, 2, 3$. Further, $y_{ts} \sim \mathcal{N}(\eta_{ts}, \sigma_{ts}^2)$ is assumed by Mercer et al. (2015). The model given in Equation (2.9), $\eta_{ts} = \mu + \xi t + \mathbf{z}_t^T \boldsymbol{\beta} + a_t + \delta_s$, is chosen in this thesis, where μ is an intercept, ξt is a linear drift, \mathbf{z}_t^T is a vector of time dependent covariates with coefficients $\boldsymbol{\beta}$, a_t is a temporal stochastic process and δ_s is a survey effect. Four different temporal processes are explored, AR(1), AR(2), RW(1) and RW(2). INLA quickly provides a deterministic approximation of the posterior of the LGM. The models are fitted using the function `inla()` from the R-package `INLA`. To simplify presentation, the only covariate considered in the rest of this section is the linear trend ξt . Other covariates can easily be added in the same way as the linear trend.

To estimate the model using `inla()` and data from more than one DHS survey, the data should be arranged in a dataframe with columns `y` and `var`, `idxs`, `idx` and `idx1`. `y` contains the logit of the estimated mortality rate (y_{ts}) and `var` contains the corresponding estimated variance (σ_{ts}^2), `idxs` contains the survey-indices (s), and the last two contain time indices (t). `idx` is used for the temporal stochastic process, while `idx1` is used for the linear trend. The data is arranged as illustrated in Table 3.2. The first part of rows contains data from KDHS03, the next from KDHS08-09 and the third part from KDHS14. The fourth part of rows defines the predictions.

It is desirable to fit the model for the period for which there is data (1980-2013) and survey-independent forecasts are needed until 2030. Therefore, as seen in Table 3.2, `idx` is (1980, ..., 2013, 1980, ..., 2013, 1980, ..., 2013, 1980, ..., 2030). `idx1` is used to model the linear trend, and to avoid numerical problems `idx1` contains time-indices that are transformed by subtracting 1980, giving (0 : 33, 0 : 33, 0 : 33, 0 : 50). `idxs` is a vector, (1, ..., 1, 2, ..., 2, 3, ..., 3, NA, ..., NA), where `inla()` interprets any “NA” (not a number in R) as “missing information”. The final elements of `idxs` are NAs as the predictions should not be assigned to a specific survey.

As seen in Table 3.2, the position of y_{ts} is in column `y`, in the row where `idx=t` and `idxs=s`. Likewise for the design-based variance in the `var`-column. As seen in Figure 3.1 and Figure 3.3, design-based estimates are not obtained from every survey for every year between 1980-2013. For the combinations of year and survey for which a design-based estimate is not obtained, y -values are represented by a NA and variances are set to one. The same goes for the rows where the `idxs`-column contains NA. The object returned by `inla()` will contain predictions for all missing y -values. To test the predictive strength of the models, it is desired to fit models using only estimated mortality rates until a specific year t_s , $t_s < 2013$. To fit those models, all y -values that corresponds to `idx > t_s` are set to NA and the corresponding variance is set to 1.

Once the input is defined, the model can be fitted. An example of how the models in this thesis are fitted is given in Listing 3.3. The first argument of the `inla()` function is a formula, given on the format of Equation (2.10). Note that the intercept μ is automatically added. The function `f()` defines a model component and the code written in Listing 3.3 defines components as given in Table 2.1. The argument `extraconstr` in the RW(2) model adds the weighted sum-to-zero constraint, which is needed to allow for a linear trend. The sum-to-zero constraint is automatically in place for RW(1) and RW(2), so no additional arguments are needed for the RW(1) model. The second argument of `inla()`

y	var	idxt	idxt1	idxs
$y_{1980,1}$	$\sigma_{1980,1}^2$	1980	0	1
\vdots	\vdots	\vdots	\vdots	
$y_{2013,1}$	$\sigma_{2013,1}^2$	2013	33	2
$y_{1980,2}$	$\sigma_{1980,2}^2$	1980	0	
\vdots	\vdots	\vdots	\vdots	3
$y_{2013,2}$	$\sigma_{2013,2}^2$	2013	33	
$y_{1980,3}$	$\sigma_{1980,3}^2$	1980	0	NA
\vdots	\vdots	\vdots	\vdots	
$y_{2013,3}$	$\sigma_{2013,3}^2$	2013	33	
NA	1	1980	0	
		\vdots	\vdots	
		2030	55	

Table 3.2: Table illustrating how the data is arranged in a dataframe before being used as an input in `inla()`, when using data from three different surveys from the same country. Data is available for 1980-2013 and predictions are made until 2030. NA is interpreted as missing information and `inla()` returns predictions for y -values given as NA.

is the name of the dataframe, and the third argument defines the scale, which is the inverse of the variance. The other arguments define whether or not marginals and statistics should be computed.

Listing 3.3: Fitting an `inla` model

```
#Use one of the following, depending on desired temporal model
formula=y ~ idxt1 + f(idxt, model="ar", order=1, values=min(input$
  idxt):max(input$idxt)) + f(idxs, model="iid")
formula=y ~ idxt1 + f(idxt, model="ar", order=2, values=min(input$
  idxt):max(input$idxt)) + f(idxs, model="iid")
formula=y ~ idxt1 + f(idxt, model="rw1", values=min(input$idxt):
  max(input$idxt)) + f(idxs, model="iid")
formula=y ~ idxt1 + f(idxt, model="rw2", values=min(input$idxt):
  max(input$idxt), extraconstr = list(A=rbind(1,(0:(end_year-min
  (input$idxt))/(end_year-min(input$idxt))), e=matrix(0,nrow =
  2,ncol=1))) + f(idxs, model="iid")

model=inla(formula=formula, data=input, scale=1/var, control.family
  = list(fixed=TRUE, initial=0), control.predictor = list(
  compute = TRUE), control.compute = list(waic=TRUE, mlik=TRUE))
```

To fit the multivariate time model given in Equation (2.15) simple adjustments have to be made to the input and the formula in the function call. The columns should be extended so that the first rows contain data from the first country, then the next part of rows contain the second country, and so on. As country-specific linear trends are wanted, the column `idxt1` should contain only NAs in the additional rows, and for each additional country

y	var	idxt	idxt1	idxs	idxt2	idxt3	group	group1
				1	NA	NA	1	1
				2	NA	NA	1	1
				3	NA	NA	1	1
NA	1			NA	NA	NA	1	1
			NA	4		NA	2	2
			NA	5		NA	2	2
			NA	6		NA	2	2
NA	1		NA	NA		NA	2	2
			NA	7	NA		3	3
			NA	8	NA		3	3
			NA	9	NA		3	3
NA	1		NA	NA	NA		3	3

Table 3.3: Table illustrating how data is arranged in a dataframe when estimating models using data from nine independent surveys from three different countries (three surveys from each country). Each row represents several rows in the dataframe. The upper left corner marked in blue corresponds to Table 3.2 and all empty cells are filled with information as shown in Table 3.2.

that is considered a new time-index column should be added. Call the new columns for $\text{idxt}i$, $i = 2, \dots, k$, column $\text{idxt}i$ is all NAs except for the rows corresponding to country i , in which the scaled time indices are. When adding covariates that are to be estimated jointly for the different countries, only one additional column per covariate is needed (for example, one column containing GDP-data). Two more columns must be added, **group** and **group1**, both containing $\text{factor}(i)$, where i corresponds to the relevant country. **group** is used for the VAR model, while **group1** allows for country-specific intercepts. See Table 3.3 for an illustration of how the complete dataframe looks like when data from a total of nine surveys from three different countries is used. An example of how the formula must be updated when estimating a model with information about three countries and temporal effect RW(1) is given in Listing 3.4. In Listing 3.4 -1 removes the overall intercept.

Listing 3.4: Fitting an multivariate time inla model

```
formula= y ~ -1 +group1+ idxt1 + idxt2 + idxt3+ f(idxt, model="
  rw1", values=max(input$idxt):min(input$idxt), group=group,
  control.group =list(model="exchangeable"))+f(idx, model="iid"
  )
```

The priors used are the `inla()` defaults. The priors for the fixed effects are normal distributions with high variances, as given in Table 2.1. Detailed information about the priors used on the parameters of the model ($\pi(\boldsymbol{\rho})$, $\pi(\sigma_a^2)$, $\pi(\sigma_\delta^2)$) can be found by executing the command `inla.doc("model")`, with "model" equal to "ar", "rw1", "rw2", or "iid". Key information is given in the following paragraphs.

The priors on parameters of AR(p) models are defined for $\theta_1, \alpha(1), \dots, \alpha(p)$, where $\theta_1 = \log(1/\sigma_a^2)$ and $\alpha(\cdot)$ is the partial autocorrelation function. The prior on θ_1 is `pc.prec`, a penalized complexity (PC) prior (Sørbye and Rue (2017)) for precision. For details about the distribution, see the document obtained by the command `inla.doc("pc.prec")`.

The hyperparameters are $(u, \alpha) = (3, 0.01)$. The interpretation of (u, α) is $\text{Prob}(\sigma_a > u) = \alpha$, $u > 0$, $0 < \alpha < 1$. The prior on the $\alpha(\cdot)$ s are `pc.cor0`, a PC prior for correlation. More information about `pc.cor0` can be found in the pdf accessed by the command `inla.doc("pc.cor0")`. The hyperparameters are $(u, \alpha) = (0.5, 0.5)$ for $\alpha(1)$ and $(u, \alpha) = (0.5, 0.4)$ for $\alpha(2)$. The hyperparameters are `s.t` $\text{Prob}(|\alpha(\cdot)| > u) = \alpha$, $0 \leq u < 1$, $0 < \alpha$. The PC distributions are used because they penalize complex temporal models and prefer simpler ones.

The parameter of RW(1) and RW(2) models, σ_a^2 is defined on $\theta_1 = \log(1/\sigma_a^2)$. The default prior in INLA on θ_1 is the log gamma distribution (`inla.doc("loggamma")`), with hyper parameters $(a, b) = (1, 5 \times 10^{-05})$. $a > 0$ is the shape parameter and $b > 0$ is the rate parameter. The survey effect is $\overset{iid}{\sim} \mathcal{N}(0, \tau^{-1})$. The prior is defined on $\log(\tau)$ and the default is the log gamma distribution, with hyper parameters $(a, b) = (1, 5 \times 10^{-05})$.

For multivariate models, there is also a prior on ρ_{group} , or more precisely on $\theta = \log((1 + \rho_{\text{group}})/(1 - \rho_{\text{group}}))$. Information about the prior can be found by executing `inla.models()`, and then search for “exchangeable” in the resulting output. The prior is $\theta \sim \mathcal{N}(0, 0.20)$. Thus, the 95% credible interval for θ is $[-0.88, 0.88]$ and $[-0.41, 0.41]$ for ρ_{group} .

A call to the function `inla()` returns a model object with several attributes. In the next section, relevant attributes for extracting posterior inference and information about useful functions in the INLA package is given.

3.3 Extracting posterior inference

To help the reader understand the object returned by `inla()`, Table 3.4 displays relevant attributes. Fitted values are “obtained by transforming the linear predictors by the inverse of the link function” (Rue and Martino (2019)). Note that fitted values and linear predictor, in this case, are the same values, as the identity link function has been used. The subset `[is.na(input$idxs)]` of the fitted values are what is later referred to as “predictions”, as they are the fitted values that are not associated with a survey. As seen from Table 3.4, the WAIC, and MLIK can be obtained directly from an `inla` object. MSE, MAE, and CRPS have to be computed using the attributes in Table 3.4.

<code>\$summary.fitted.values\$mean</code>	mean of fitted values
<code>\$summary.fitted.values\$`0.025quant`</code>	2.5% quantile of fitted values
<code>\$summary.fitted.values\$`0.975quant`</code>	97.5% quantile of fitted values
<code>\$waic\$waic</code>	WAIC
<code>\$mlik[1]</code>	marginal log likelihood
<code>\$summary.linear.predictor\$mean</code>	mean of linear predictor
<code>\$summary.linear.predictor\$sd</code>	standard deviation of linear predictor
<code>\$marginals.linear.predictor</code>	marginals of the linear predictor

Table 3.4: Relevant attributes for model analysis of an `inla` object.

The returned `inla` object also contains information about the parameters of the model. Information about the hyperparameters are given in `$summary.hyperpar` while infor-

mation about the fixed effects are given in `$summary.fixed`. As an example, some relevant attributes from models estimated with AR(1) and AR(2) as temporal effect are given in Table 3.5 and Table 3.6, respectively. The ϕ coefficients and autocorrelation function of the AR(2) process can be obtained using the functions `inla.ar.pacf2phi()` and `inla.ar.pacf2acf()`.

<code>\$summary.hyperpar\$`0.5quant`[1]</code>	$\gamma(0)^{-1}$
<code>\$summary.hyperpar\$`0.5quant`[2]</code>	ϕ_1
<code>\$summary.fixed\$mean[2]</code>	ξ
<code>\$summary.fixed\$mean[1]</code>	μ

Table 3.5: Relevant attributes of an `inla` object estimated with AR(1) as temporal effect.

<code>\$summary.hyperpar\$`0.5quant`[1]</code>	$\gamma(0)^{-1}$
<code>\$summary.hyperpar\$`0.5quant`[2]</code>	$\alpha(1)$
<code>\$summary.hyperpar\$`0.5quant`[3]</code>	$\alpha(2)$
<code>\$summary.fixed\$mean[2]</code>	ξ
<code>\$summary.fixed\$mean[1]</code>	μ

Table 3.6: Relevant attributes of an `inla` object estimated with AR(2) as temporal effect.

3.4 Including covariates

It is reasonable to assume that mortality rates are associated with, for example, the size of the population, money spent on health interventions, natural disasters, and temperature. Covariates that are correlated with the response can improve a model's predictive strength. Therefore one should explore the effect of including covariates for which there exists reliable data in the mortality rates models. When using a model with covariates to produce forecasts, forecasted values of the covariates must be available. This section provides information about population size and gross domestic product (GDP), two covariates that are used in Chapter 5 to illustrate the impact of covariates.

It is impossible to know future values of population size and GDP, partly because it is not possible to know how human behavior will affect these quantities. The climate change research community have produced forecasts for population, GDP, and urbanization for five different scenarios. The scenarios are based on five different sets of assumptions, called Shared Socioeconomic Pathways (SSP). Riahi et al. (2017) gives an overview of the SSPs. Table 2 in the article by Riahi et al. (2017) gives an explanation of the five SSPs and the headings from that table is re-given in Table 3.7. For more details about the SSPs, please see O'Neill et al. (2017) and the SSP database at the IIASA website³.

Historical and forecasted GDP and population size of many countries are available at the IIASA website. There are two competing models for the development of GDP, and it is recommended that both are used and compared. Henceforth they will be referred to

³<https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=10>

SSP1	Taking the Green Road (Low challenges to mitigation and adaptation)
SSP2	Middle of the Road (Medium challenges to mitigation and adaptation)
SSP3	Regional Rivalry. A Rocky Road (High challenges to mitigation and adaptation)
SSP4	Inequality. A Road Divided (Low challenges to mitigation, high challenges to adaptation)
SSP5	Fossil-fueled Development. Taking the Highway (High challenges to mitigation, low challenges to adaptation)

Table 3.7: Summary of Table 2 in the article by Riahi et al. (2017), explaining the SSPs.

as GDP1 and GDP2. The historical GDP and population numbers are available for every fifth year from 1980 to 2010 plus the year 2008. Predictions for each of the five scenarios are available for every fifth year from 2010 to 2100. As mortality estimates are computed yearly, it is desirable to get the yearly population and GDP data. That is obtained using cubic spline interpolation. It is reasonable to assume that the relative change in covariates is more indicative of changes in mortality rates than the absolute change. Therefore, log transformations of the covariates are also considered.

Models with covariates are estimated as described in Section 3.2, with minor modifications. The input has to contain one column for each covariate, and the formula must be updated to include the name of the columns with covariates.

Simulation Study

The goal of this simulation study is to determine which temporal processes that are most suitable in models for NMR and U5MR for estimation and forecasting.

4.1 Details of the simulation study

Simulated NMR and U5MR are used to explore the performance of models with temporal effect AR(1), AR(2), RW(1), and RW(2). These temporal effects are chosen because they are simple and keep the modeling process transparent. In this chapter, the model given in Equation (2.9) is considered without any covariates. The coefficients and parameters used when simulating yearly NMR and U5MR are estimated using data from DHS surveys undertaken in Kenya in 2003, 2008-2009 and 2014. As random walks are non-stationary with exploding variances, NMR and U5MR are simulated using AR(1) and AR(2) models. The models are estimated as explained in Section 3.2. The parameters of the four fitted models are given in Table 4.1.

To simulate NMR and U5MR, the initial values of the temporal effect must be obtained. As explained in Section 2.2, the initial p values of an AR(p) process can be simulated using their joint distribution, $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$. The co-variance matrix, $\mathbf{\Gamma}$, is computed using the pacf values in the returned `inla` object. Once the required initial values, a_t , $t = -p + 1, \dots, 0$, are simulated, a_t for $t > 0$ can be simulated using Equation (2.1), with

Rate	Model	ϕ_1	ϕ_2	σ_a
NMR	AR(1)	0.14		0.13
NMR	AR(2)	0.14	0.04	0.13
U5MR	AR(1)	0.97		0.27
U5MR	AR(2)	0.97	-0.03	0.29

Table 4.1: Estimated parameters of models (median). ϕ_1 and ϕ_2 are the estimated coefficients of the AR process, while σ_a is the marginal standard deviation of the AR process.

$p = 1$ for AR(1) and $p = 2$ for AR(2). Unique initial values should be simulated for every desired simulated dataset and unique innovations should be simulated for every simulated dataset and for every time step. An `inla` object does not contain the residual variance, σ_ε^2 , only the (inverse of the) marginal variance σ_a^2 . Therefore Equation (2.5) must be used to obtain σ_ε^2 .

It is desirable to reproduce the uncertainty of the survey estimates in the simulations. Therefore, models are fitted to “observed” simulated NMR and U5MR, where the observation noise for year t , ω_t , is distributed $\mathcal{N}(0, \sigma_{t,3}^2)$, for years $t < 2014$ and $\omega_t = 0$ for $t \geq 2014$. The variance from the 2014 survey is preferred over the variance from the two other surveys as the KDHS14 is the most recent survey and has the largest number of observations. As evident in Figure 3.3 and Figure 3.1, KDHS14 gives estimates for U5MR for 1980-2013 and for NMR for 1982-2013. Therefore, when simulating data, U5MR is simulated for 1980-2030 while NMR is simulated for 1982-2030. To ensure reproducibility, `set.seed(222)` is used when simulating the initial values and when simulating from AR(1)/AR(2).

In addition to investigate how well the models work if the current trend continues, it is also interesting to explore scenarios where there are significant positive or negative changes in mortality rates. Such scenarios are simulated by adding a quadratic effect $\psi(t - 2013)^2$, where t is the year and $\psi = 0$ if $t \leq 2013$ but $\psi \neq 0$ if $t > 2013$, to the simulated `logit(NMR)` and `logit(U5MR)`. Negative values of ψ lead to a decrease in simulated mortality rate, thus a “better health” scenario, while positive values of ψ do the opposite, a “worse health” scenario. For convenience, henceforth S0 will refer to the “no change in trends” scenario, SBH to “scenario better health” and SWH to “scenario worse health”. As it is desired to make predictions until 2030, $|\psi|$ is set to $1/17^2$, leading to an increase or decrease of 1 on the logit scale when $t = 2030$. It could have been interesting to investigate other values of $|\psi|$. However, that is not within the scope of this thesis.

The model assessments and scoring rules explained in Section 2.6 are used to evaluate the performance of the models. WAIC, MLIK, MAE, root MSE (RMSE), and CRPS are used to evaluate the models’ explanatory strength. MAE, RMSE, and CRPS are also calculated to evaluate the predictive strength of the models. When computing MAE, RMSE, and CRPS, the fitted values are compared to the simulated values. RMSE is presented instead of MSE so that the values are comparable with MAE. MAE and RMSE are calculated both at logit-scale and using the mortality rates directly. CRPS is calculated, using Equation (2.18), only at logit-scale as it is believed that absolute changes in probability are not comparable. For example, a reduction of 0.1 percentage points is different if the reduction is from 1% to 0.9% compared to 0.1% to 0.0%. When determining which models perform best the focus is on prediction and in particular the CRPS scores. Most weight is given to scenario S0.

In the following two sections, analysis of 1000 simulated datasets is presented. All models are fitted using data for $t < 2014$, thereafter predictions are made until $t = 2030$. Statistics explaining the explanatory strength of the models are calculated for years $t < 2014$, while statistics evaluating the predictive strength of the models are calculated for years $t \in [2014, 2030]$. In Section 4.2 the datasets are simulated from an AR(1) model, whereas in Section 4.3 the datasets are simulated from an AR(2) model. All models that are fitted to the simulated datasets are estimated without a survey-effect.

4.2 Autoregressive process of order 1

In this section, datasets are simulated using an AR(1) model with coefficients as presented in Table 4.1. Logit(NMR) and NMR per 1000 live birth are presented in Figure 4.1. Corresponding figures for U5MR can be found in Figure 4.2. From the plots in Figure 4.1 one can see that RW(2) is the only model for which all three scenarios are well within the credible interval. Figure 4.2 shows that the credible intervals for U5MR created by the AR(1), AR(2) and RW(1) models are reasonable. The credible intervals of the RW(2) models for NMR and U5MR are so large that the predictions are not informative. This is expected and coherent with the theory presented in Section 2.2. The AR(1) and RW(1) models look very similar, while the AR(2) model is less smooth and better at capturing the year-to-year variation for NMR.

Model	WAIC	-MLIK	MAE $\times 10^{-01}$	RMSE $\times 10^{-01}$	CRPS $\times 10^{-01}$
AR(1)	5.31	11.96	1.86	2.44	1.59
AR(2)	-1.60	13.00	1.34	1.88	1.07
RW(1)	5.66	13.88	1.87	2.45	1.60
RW(2)	6.28	25.72	1.80	2.35	1.53

Table 4.2: Mean explanatory statistics over 1000 NMRs simulated from an AR(1) model. All statistics are computed using the logit transformation of the NMR. The values marked with green are most favorable in each column.

Model	WAIC	-MLIK	MAE $\times 10^{-01}$	RMSE $\times 10^{-01}$	CRPS $\times 10^{-01}$
AR(1)	-24.99	0.78	1.23	1.85	1.18
AR(2)	-31.24	0.17	1.02	1.64	0.78
RW(1)	-27.16	1.33	1.20	1.80	1.04
RW(2)	-28.04	11.55	1.19	1.74	0.91

Table 4.3: Mean explanatory statistics over 1000 U5MRs simulated from an AR(1) model. All statistics are computed using the logit transformation of the U5MR. The values marked with green are most favorable in each column.

From Table 4.2 one can see that the best temporal effect for NMR with respect to WAIC is AR(2), with the other three performing worse and relatively similar. The best temporal effect in terms of MLIK is AR(1), and the worst is RW(2). The negative marginal log-likelihood of the AR(2) and RW(1) models are much closer to that of the AR(1) models than the RW(2) models. That indicates that RW(2) models are not good fits for the datasets. In terms of MAE, RMSE and CRPS AR(2) performs best, with RW(2) on second place followed by AR(1) and RW(1). That indicates that even though RW(2) models have the least favorable MLIK scores, RW(2) models can produce smooth fits of the data. Overall, the models with AR(2) as temporal effect have the highest explanatory strength, as the AR(2) models score best with respect to four out of five criteria.

Table 4.3 shows similar results for U5MR. The AR(2) models have the lowest average for all assessment criteria. The RW(2) models come in second with respect to WAIC but

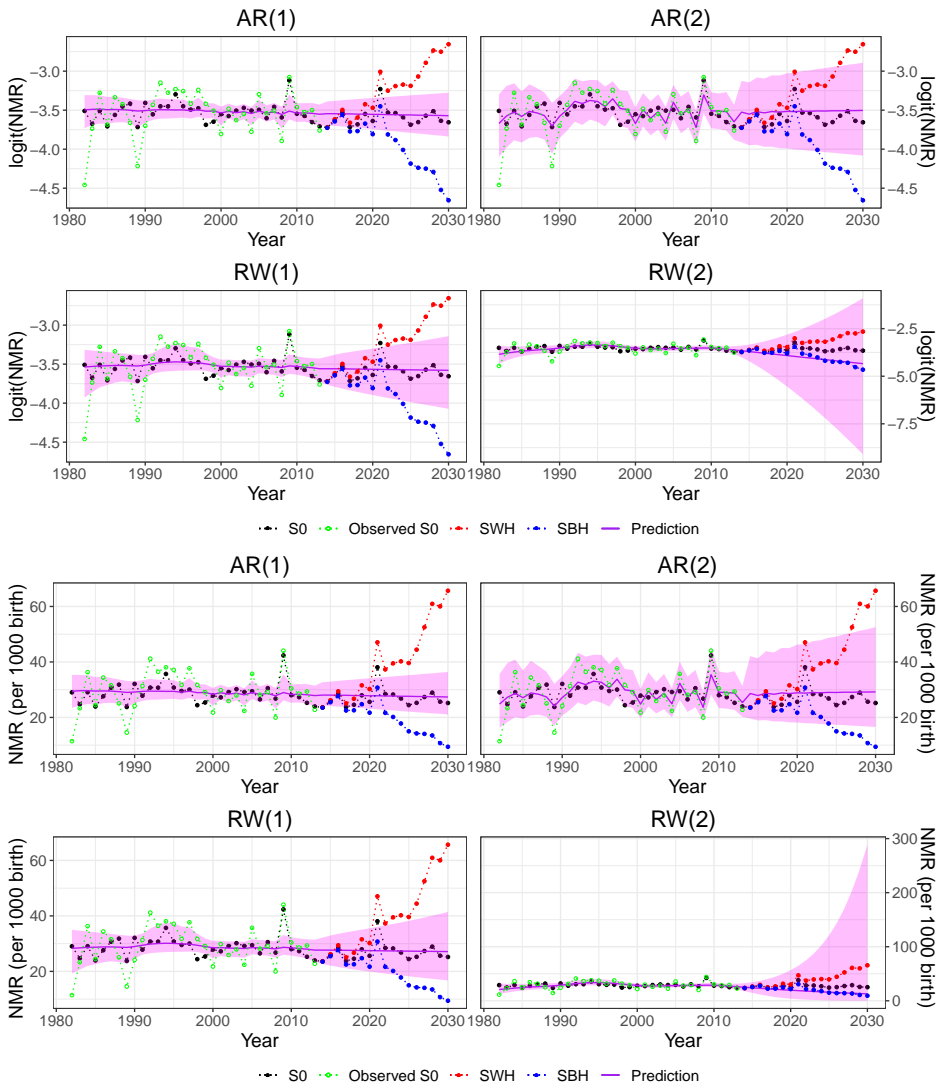


Figure 4.1: logit(NMR) and NMR per 1000 live birth, simulated by an AR(1) model. S_0 and observations of S_0 (with noise) are plotted as well as projections for SBH and SWH with $|\gamma| = 1/17^2$. The solid purple line is the fitted model/predictions, computed using data for $t < 2014$, with a 95% credible interval.

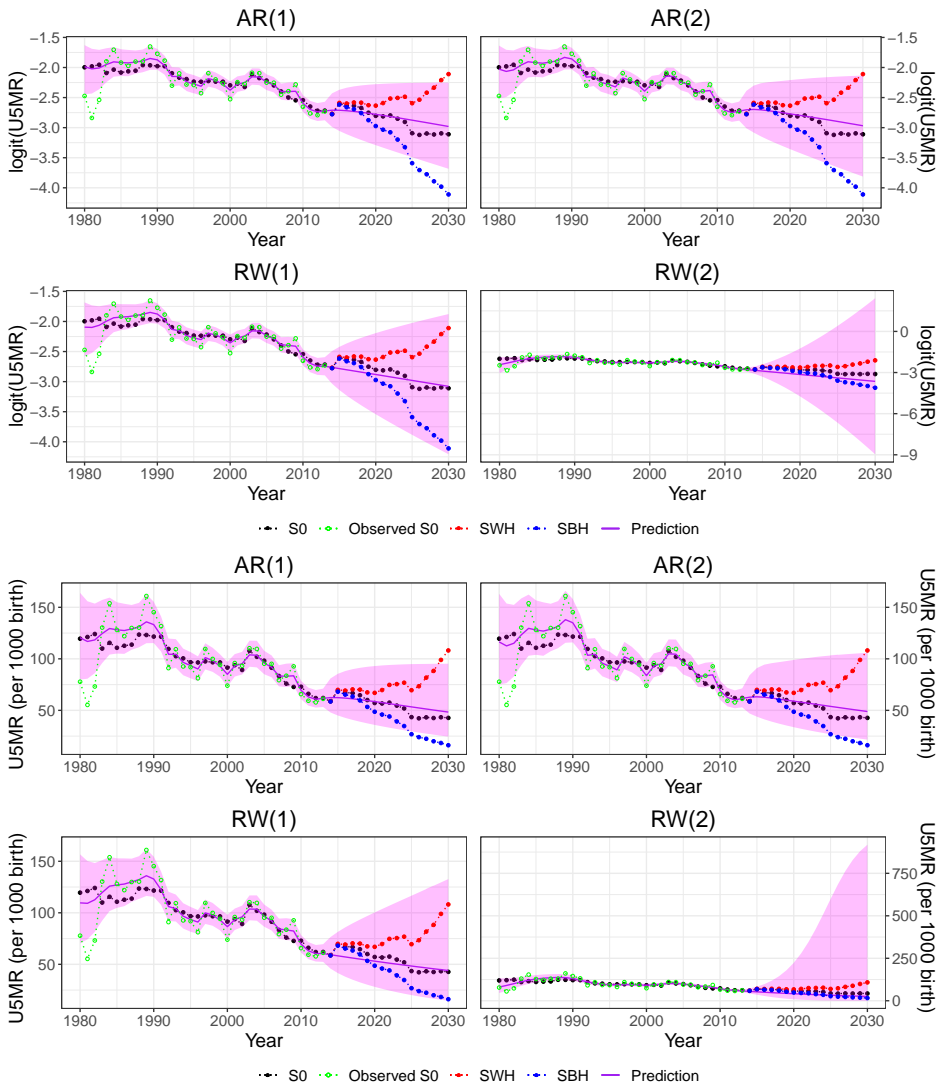


Figure 4.2: $\text{logit}(U5MR)$ and $U5MR$ per 1000 live birth, simulated by an AR(1) model. S_0 and observations of S_0 (with noise) are plotted as well as projections for SBH and SWH with $|\gamma| = 1/17^2$. The solid purple line is the fitted model/predictions, computed using data for $t < 2014$, with a 95% credible interval.

have once again a far higher negative marginal log-likelihood score than the other three model types. RW(2) models also have the second lowest MAE, RMSE and CRPS, again revealing that RW(2) models can produce smooth fits of observed data.

Next, it is desirable to investigate the models' predictive strengths. Plots of CRPS are given in Figure 4.3 and Figure 4.4 for $\text{logit}(\text{NMR})$ and $\text{logit}(\text{U5MR})$, respectively. CRPS and RMSE for 2, 7, 12 and 17 steps (corresponding to year 2015, 2020, 2025 and 2030) can be found in Table 4.4 and Table 4.5.

From Figure 4.3 and Table 4.4 one can see that RW(2) models for NMR are outperformed by the three other model types. However, RW(2) models have the lowest average CRPS score for the final prediction steps in the alternative scenarios. The differences in CRPS for RW(2) models and the other types of models are smaller for scenario SBH and SWH than for scenario S0. Both in terms of CRPS and RMSE AR(1), AR(2) and RW(1) models perform very similarly. From Table 4.4 one can see that the AR(2) models most often have the lowest average CRPS and a competitive RMSE for NMR, while AR(1) models most often have the lowest RMSE. For U5MR, Figure 4.4 and Table 4.5 show that AR(2) and RW(1) models have similar scores and that they perform better than AR(1) and RW(2). For scenario SBH and SWH RW(2) models once again have the lowest average CRPS for the final prediction steps, due to the high uncertainty in the RW(2) predictions.

Plots similar to Figure 4.3 and Figure 4.4 are given for MAE and RMSE for both logit of mortality rates and mortality rates in Appendix A. The plots in the appendix are consistent with the results presented in this section. They show that in general RW(2) models perform worse than the other three model types, which performs very similar.

In conclusion, Table 4.2 and Table 4.3 reveal that models with RW(2) as temporal effect are capable of smoothing observed data, however, the RW(2) models do not perform well with respect to marginal log likelihood. The AR(2) models consistently have the lowest WAIC, MAE, RMSE, and CRPS. For NMR, AR(2) models have average negative MLIK close to the lowest value, and AR(2) models have the lowest average negative MLIK for U5MR. Therefore one can argue that AR(2) models have the highest explanatory strength. Based on analysis of predictions one can conclude that AR(2) models perform best for NMR with respect to CRPS while AR(2) and RW(1) models perform well for U5MR.

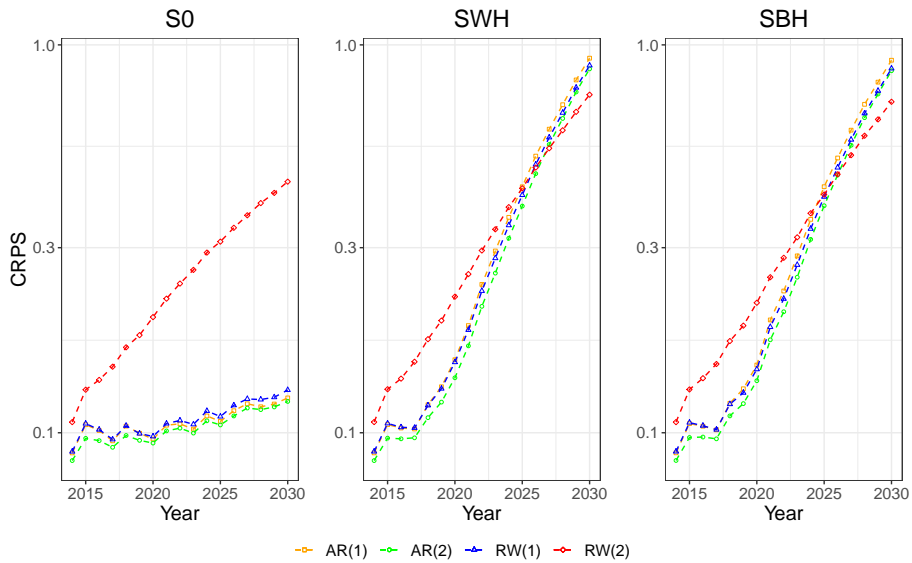


Figure 4.3: CRPS of NMR simulated by AR(1).

Model	CRPS $\times 10^{-01}$				RMSE $\times 10^{-01}$			
	2015	2020	2025	2030	2015	2020	2025	2030
Model	S0							
AR(1)	1.05	0.96	1.07	1.23	1.67	1.61	1.81	2.09
AR(2)	0.97	0.94	1.05	1.20	1.68	1.63	1.81	2.09
RW(1)	1.06	0.98	1.10	1.29	1.72	1.67	1.89	2.19
RW(2)	1.29	1.99	3.11	4.44	2.30	3.67	5.42	7.22
Model	SBH							
AR(1)	1.06	1.49	4.31	9.13	1.67	2.32	5.30	10.20
AR(2)	0.97	1.36	3.85	8.58	1.69	2.34	5.31	10.20
RW(1)	1.06	1.46	4.06	8.69	1.73	2.35	5.32	10.20
RW(2)	1.29	2.16	4.11	7.13	2.30	3.98	7.23	12.10
Model	SWH							
AR(1)	1.05	1.54	4.31	9.24	1.67	2.36	5.30	10.30
AR(2)	0.97	1.39	3.84	8.67	1.68	2.36	5.29	10.30
RW(1)	1.06	1.52	4.10	8.85	1.72	2.41	5.34	10.30
RW(2)	1.29	2.24	4.27	7.44	2.31	4.11	7.49	12.60

Table 4.4: 2, 7, 12 and 17 step ahead prediction CRPS and RMSE of NMRs simulated by AR(1), for scenario S0, SBH and SWH. Last year with data is 2013, so the prediction steps corresponds to year 2015, 2020, 2025 and 2030. The values marked with green are the lowest (most favorable) in each column, for each scenario. CRPS and RMSE are calculated at logit-scale.

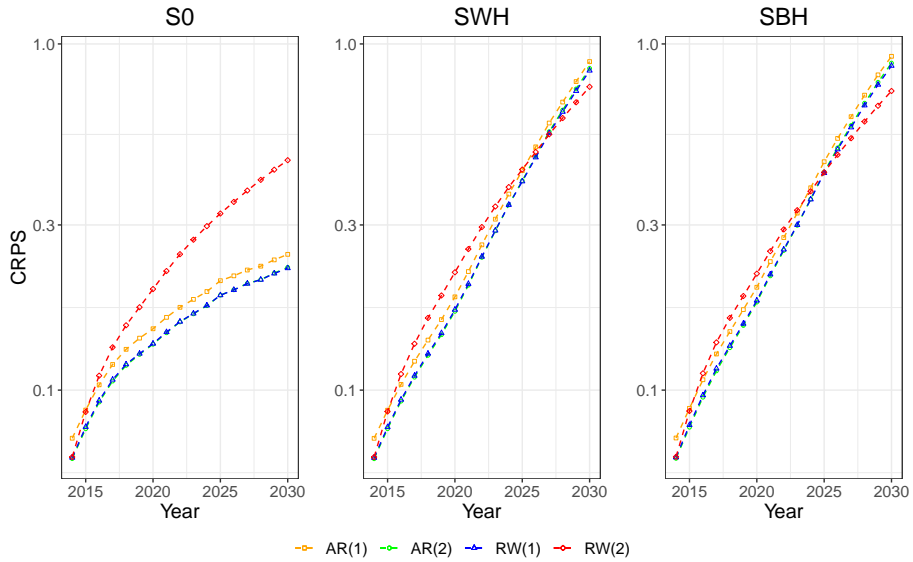


Figure 4.4: CRPS of U5MR simulated by AR(1).

Model	CRPS $\times 10^{-01}$				RMSE $\times 10^{-01}$			
	2015	2020	2025	2030	2015	2020	2025	2030
Model	S0							
AR(1)	0.88	1.50	2.07	2.47	1.43	2.40	3.23	3.88
AR(2)	0.77	1.36	1.88	2.26	1.35	2.33	3.15	3.79
RW(1)	0.78	1.36	1.88	2.25	1.32	2.28	3.11	3.76
RW(2)	0.86	1.96	3.24	4.61	1.55	3.48	5.51	7.52
Model	SBH							
AR(1)	0.89	1.98	4.57	9.21	1.44	3.01	6.07	10.90
AR(2)	0.78	1.80	4.26	8.80	1.36	2.94	6.02	10.80
RW(1)	0.79	1.81	4.24	8.65	1.34	2.90	5.99	10.80
RW(2)	0.87	2.17	4.26	7.31	1.56	3.86	7.37	12.40
Model	SWH							
AR(1)	0.88	1.86	4.31	8.89	1.43	2.87	5.81	10.60
AR(2)	0.77	1.69	4.03	8.50	1.35	2.82	5.77	10.50
RW(1)	0.78	1.71	4.01	8.38	1.32	2.78	5.76	10.60
RW(2)	0.87	2.19	4.34	7.52	1.56	3.89	7.49	12.60

Table 4.5: 2, 7, 12 and 17 step ahead prediction CRPS and RMSE of U5MRs simulated by AR(1), for scenario S0, SBH, and SWH. Last year with data is 2013, so the prediction steps correspond to the years 2015, 2020, 2025 and 2030. The values marked with green are the lowest (most favorable) in each column, for each scenario. CRPS and RMSE are calculated at logit-scale.

4.3 Autoregressive process of order 2

In this section, datasets are simulated from a model with AR(2) as the temporal stochastic process. The model coefficients are presented in Table 4.1. AR(1), AR(2), RW(1) and RW(2) models for one simulated NMR dataset and one simulated U5MR dataset are presented in Figure 4.5 and Figure 4.6, respectively. From Figure 4.5 and Figure 4.6 one can see that RW(2) is the only stochastic process for which all three scenarios are within the 95% credible interval for NMR, and that the credible intervals of the RW(2) models are huge. The other three models are more similar, but once again the AR(2) models are less smooth and capture more of the simulated data and observation noise within the credible intervals than the AR(1) and RW(1) models for NMR.

From Table 4.6 and Table 4.7 one can see the same tendencies as in Table 4.2 and Table 4.3. It is again evident that RW(2) models have less favorable average marginal log-likelihood score compared to the other models, but that they are capable of smoothing observed data. The AR(2) models have the lowest average WAIC, MAE, RMSE, and CRPS. For NMR, the AR(2) models have a competitive marginal log likelihood and for U5MR AR(2) models have the lowest marginal log likelihood. Therefore one can conclude that AR(2) models have the overall best explanatory strength.

Model	WAIC	-MLIK	MAE $\times 10^{-01}$	RMSE $\times 10^{-01}$	CRPS $\times 10^{-01}$
AR(1)	4.16	11.32	1.85	2.44	1.47
AR(2)	-2.14	12.64	1.35	1.89	1.01
RW(1)	4.36	13.23	1.86	2.44	1.48
RW(2)	4.85	25.09	1.78	2.33	1.42

Table 4.6: Mean explanatory statistics over 1000 NMRs simulated from an AR(2) model. All statistics are computed using the logit transformation of the NMR. The values marked with green are the most favorable in each column.

Model	WAIC	-MLIK	MAE $\times 10^{-01}$	RMSE $\times 10^{-01}$	CRPS $\times 10^{-01}$
AR(1)	-24.44	1.09	1.25	1.94	1.25
AR(2)	-30.70	0.49	1.04	1.73	0.89
RW(1)	-26.59	1.64	1.23	1.89	1.14
RW(2)	-27.38	11.88	1.21	1.82	1.00

Table 4.7: Mean explanatory statistics over 1000 U5MRs simulated from an AR(2) model. All statistics are computed using the logit transformation of the U5MR. The values marked with green are the most favorable in each column. CRPS and RMSE are calculated at logit-scale.

With respect to predictive strength, the conclusions one can draw based on Figure 4.7, Figure 4.8, Table 4.8, and Table 4.9 are similar to those based on the corresponding figures and tables in Section 4.2. RW(2) is outperformed by the other three models with respect to the prediction scores, although RW(2) models have the lowest CRPS score for scenario SBH and SWH for the final prediction steps. The other three models score very similarly. It is possible to argue that AR(2) and RW(1) models are better at predicting

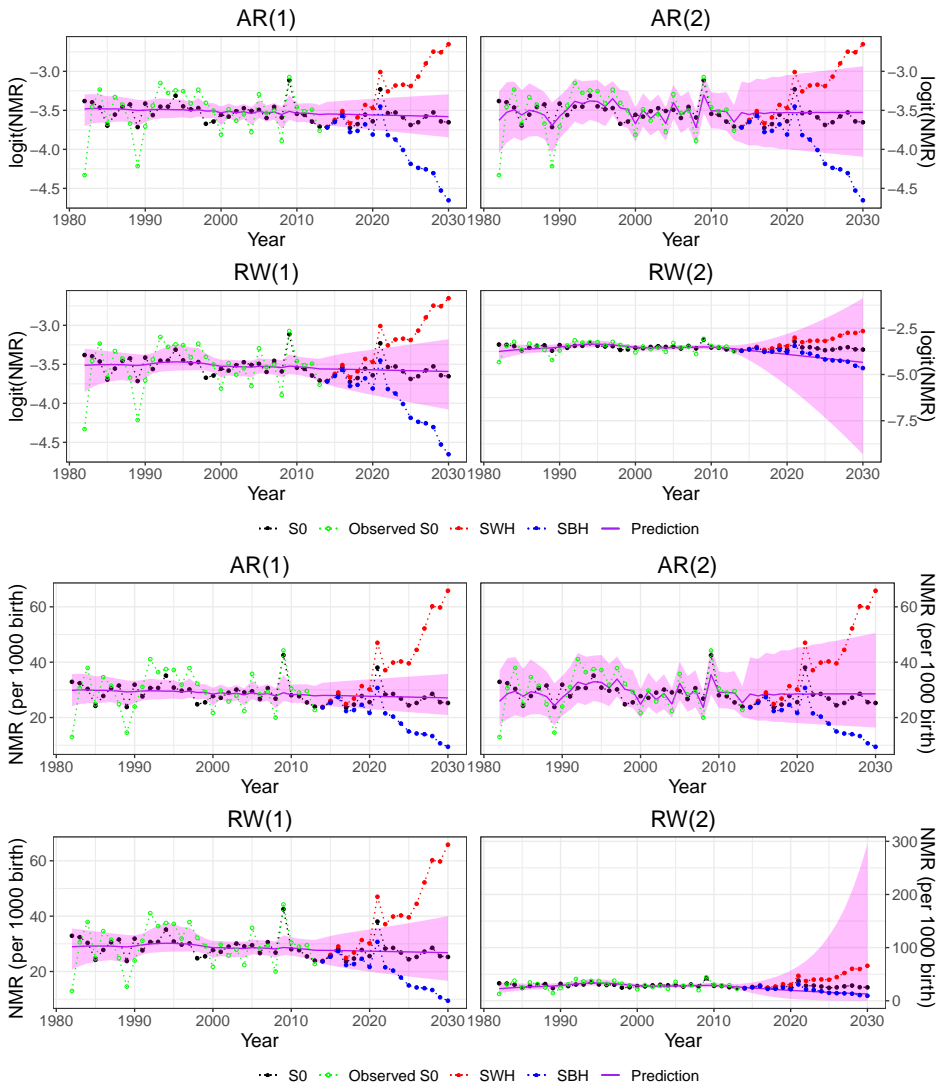


Figure 4.5: logit(NMR) and NMR per 1000 live birth, simulated by an AR(2) model. S_0 and observations of S_0 (with noise) are plotted as well as projections for SBH and SWH with $|\gamma| = 1/17^2$. The solid purple line is the fitted model/predictions, computed using data for $t < 2014$, with a 95% credible interval.

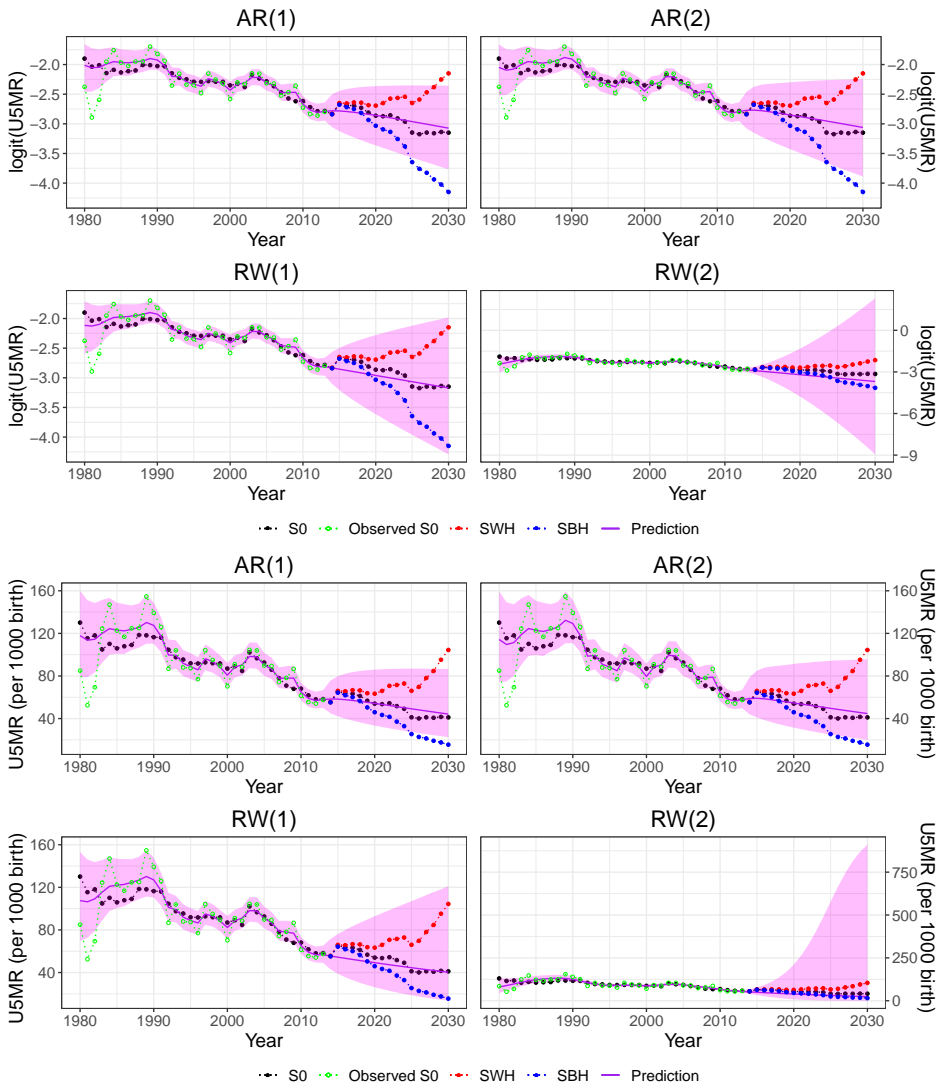


Figure 4.6: $\text{logit}(U5MR)$ and $U5MR$ per 1000 live birth, simulated by an AR(2) model. S_0 and observations of S_0 (with noise) are plotted as well as projections for SBH and SWH with $|\gamma| = 1/17^2$. The solid purple line is the fitted model/predictions, computed using data for $t < 2014$, with a 95% credible interval.

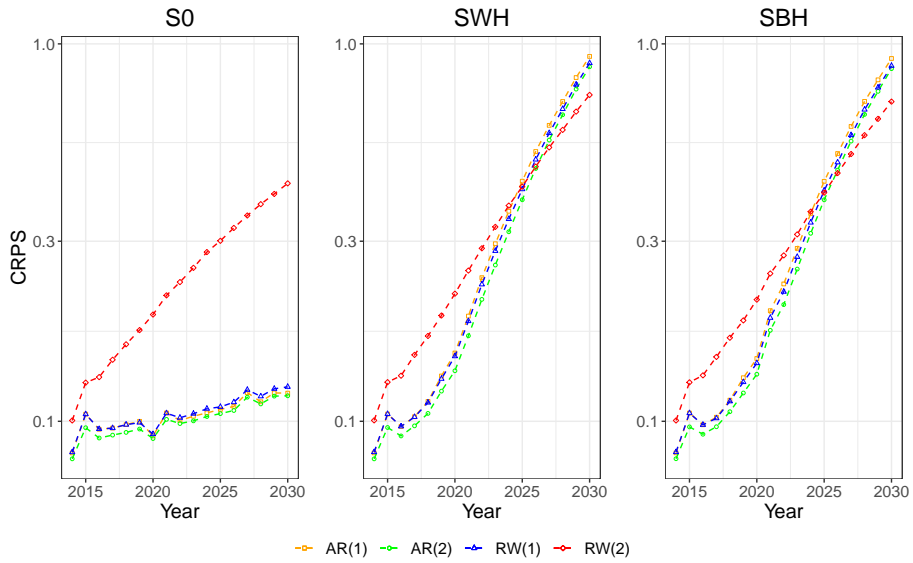


Figure 4.7: Average CRPS for NMR simulations with AR(2).

Model	CRPS $\times 10^{-01}$				RMSE $\times 10^{-01}$			
	2015	2020	2025	2030	2015	2020	2025	2030
Model	S0							
AR(1)	1.04	0.92	1.07	1.19	1.66	1.53	1.81	2.03
AR(2)	0.96	0.90	1.05	1.17	1.67	1.55	1.82	2.03
RW(1)	1.04	0.93	1.09	1.23	1.70	1.58	1.88	2.11
RW(2)	1.27	1.92	3.01	4.27	2.25	3.55	5.25	6.96
Model	SBH							
AR(1)	1.05	1.47	4.32	9.15	1.67	2.27	5.30	10.10
AR(2)	0.97	1.33	3.86	8.60	1.68	2.28	5.31	10.20
RW(1)	1.05	1.43	4.08	8.75	1.71	2.29	5.31	10.10
RW(2)	1.27	2.10	4.03	7.03	2.26	3.87	7.11	11.90
Model	SWH							
AR(1)	1.05	1.52	4.33	9.26	1.66	2.31	5.30	10.30
AR(2)	0.96	1.36	3.86	8.70	1.67	2.31	5.30	10.30
RW(1)	1.05	1.49	4.12	8.89	1.70	2.35	5.33	10.30
RW(2)	1.27	2.18	4.19	7.33	2.26	3.99	7.36	12.40

Table 4.8: 2, 7, 12 and 17 step ahead prediction CRPS and RMSE of NMRs simulated by AR(2), for scenario S0, SBH, and SWH. Last year with data is 2013, so the prediction steps correspond to the years 2015, 2020, 2025 and 2030. The values marked with green are the lowest (most favorable) in each column, for each scenario. CRPS and RMSE are calculated at logit-scale.

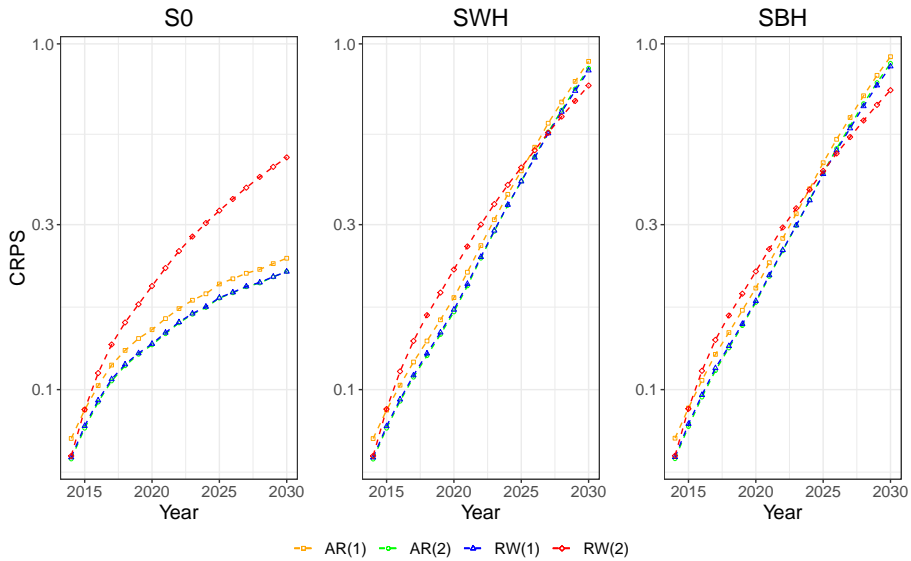


Figure 4.8: CRPS of U5MR simulated by AR(2).

Model	CRPS $\times 10^{-01}$				RMSE $\times 10^{-01}$			
	2015	2020	2025	2030	2015	2020	2025	2030
Model	S0							
AR(1)	0.87	1.49	2.02	2.40	1.43	2.37	3.16	3.74
AR(2)	0.77	1.35	1.84	2.20	1.35	2.30	3.08	3.66
RW(1)	0.79	1.36	1.84	2.19	1.33	2.27	3.05	3.65
RW(2)	0.88	1.99	3.29	4.70	1.57	3.54	5.60	7.63
Model	SBH							
AR(1)	0.89	1.97	4.53	9.18	1.44	2.98	6.02	10.80
AR(2)	0.78	1.79	4.24	8.78	1.36	2.92	5.98	10.80
RW(1)	0.80	1.80	4.20	8.61	1.34	2.88	5.94	10.80
RW(2)	0.88	2.20	4.29	7.35	1.58	3.91	7.42	12.40
Model	SWH							
AR(1)	0.87	1.85	4.29	8.90	1.43	2.85	5.78	10.50
AR(2)	0.77	1.68	4.01	8.50	1.35	2.80	5.74	10.50
RW(1)	0.78	1.70	4.00	8.38	1.33	2.78	5.74	10.50
RW(2)	0.88	2.23	4.39	7.58	1.58	3.95	7.58	12.80

Table 4.9: 2, 7, 12 and 17 step ahead prediction CRPS and RMSE of U5MRs simulated by AR(2), for scenario S0, SBH, and SWH. Last year with data is 2013, so the prediction steps correspond to the years 2015, 2020, 2025 and 2030. The values marked with green are the lowest (most favorable) in each column, for each scenario. CRPS and RMSE are calculated at logit-scale.

than AR(1) models, as the AR(2) and RW(1) models overall have marginally lower CRPS scores for U5MR. Plots similar to Figure 4.7 and Figure 4.8 are given for MAE and RMSE in Appendix A. The plots in the appendix are consistent with the results presented in this section. In general, the RW(2) models are outperformed by the other three models. In some plots, one can also see that the AR(1) models perform slightly worse than AR(2) and RW(1) models.

The analysis of datasets simulated from a model with AR(2) as temporal effect thus leads to the same conclusion as the analysis of the datasets simulated by a model with AR(1) as temporal effect; AR(2) models have the greatest explanatory strength and produce best predictions for NMR with respect to CRPS while AR(2) and RW(1) models perform well for U5MR.

4.4 Discussion

RW(2) models have the lowest average CRPS scores for many steps ahead predictions in alternative scenarios. However, the CRPS scores for the alternative scenarios are not considered more important than the CRPS score for scenario S0 because they are simply an addition of a quadratic trend in time. Provided availability of expert knowledge, it would have been interesting to develop more realistic scenarios that, for example, incorporate the effect of potential government interventions. If a model is to be evaluated based on its ability to produce forecast in alternative futures, some covariate should be included in the model allowing it to account for the changes. The RW(2) models have low CRPS scores for far-ahead forecasts in alternative scenarios not because they were able to predict the change in trend in mortality rates, but because they have large credible intervals.

The results presented in this chapter clearly show that different model assessment scores prefer different models. For example in Table 4.2 and Table 4.6 AR(1) models most often have the lowest RMSE while AR(2) models most often have the lowest CRPS score. In addition, the difference in performance of the different models is relatively small. Because of this, it is not possible to determine an absolute best model, and the choice of criteria used in analysis affects the conclusions.

The assessed models are of different complexity. RW(1) and RW(2) have only one parameter, but the study has shown that even though they are simple, RW(1) models can produce competitive forecasts. AR(1) has two parameters, and AR(2) has three. If there is not enough data to estimate ϕ_2 in the AR(2) models, the AR(2) models may produce worse forecasts than AR(1) models. In this situation, the simulation study has shown that there is enough data for the added complexity of AR(2) models to improve on the AR(1) models.

The overall conclusion from the simulation study is that AR(1), AR(2) and RW(1) are preferred over RW(2) as the stochastic temporal process in an LGM for mortality rates. RW(2) models interpolate data well, but they are not able to extrapolate data with meaningful credible intervals. AR(2) models have the greatest explanatory strength, and AR(2) models perform well with respect to prediction scores. For U5MR and one can argue that AR(2) and RW(1) models produce better predictions than AR(1) models. Mortality rates in Kenya are, therefore, analyzed using LGMs with stochastic temporal effects AR(2) and RW(1) in Chapter 5.

Univariate Modeling of Child Mortality Rates in Kenya

In this chapter, historical and forecasted population size (POP) and gross domestic product (GDP) of Kenya are explored as possible covariates in models aiming to produce long-term forecasts of child mortality rates. Models with temporal effect AR(2) and RW(1) are applied to yearly NMR and U5MR in Kenya from 1980-2013, and their predictive strengths are compared. The overall best models to predict child mortality rates in Kenya are used to evaluate Kenya's progress towards the SDGs and MDGs.

5.1 Model selection

The SSP database has published forecasts for POP and GDP under multiple scenarios for several countries, including Kenya. The population data and their log transformations are shown in Figure 5.1. The two different projections for GDP and their log transformations are presented in Figure 5.2 and Figure 5.3. Population size and GDP are expected to continue to grow, and a more rapid growth in GDP is projected in the future.

To explore the correlation between mortality rates and the potential covariates, cross plots are presented in Appendix B. Figure B.1 and Figure B.2 show POP, GDP, $\log(\text{POP})$ and $\log(\text{GDP})$ for years 1980-2010 plotted against $\text{logit}(\text{NMR})$ and $\text{logit}(\text{U5MR})$, respectively. Figure B.3 and Figure B.4 show the ratio and ratio of log of GDP and POP for years 1980-2010 plotted against $\text{logit}(\text{NMR})$ and $\text{logit}(\text{U5MR})$. Figure B.1 and Figure B.3 do not indicate a clear correlation between NMR and any of the possible covariates. However, this is not that surprising as Figure 5.1, Figure 5.2 and Figure 5.3 show clear trends whereas Figure 3.1 show no clear trends. From Figure B.2 and Figure B.4 it appears that U5MR is correlated with the covariates.

Impact of covariates in models for mortality rates should be tested for every combination of covariates available. As the two GDP models are two competing models, it is not meaningful to include GDP1 and GDP2 in the same model. Since it is possible that the

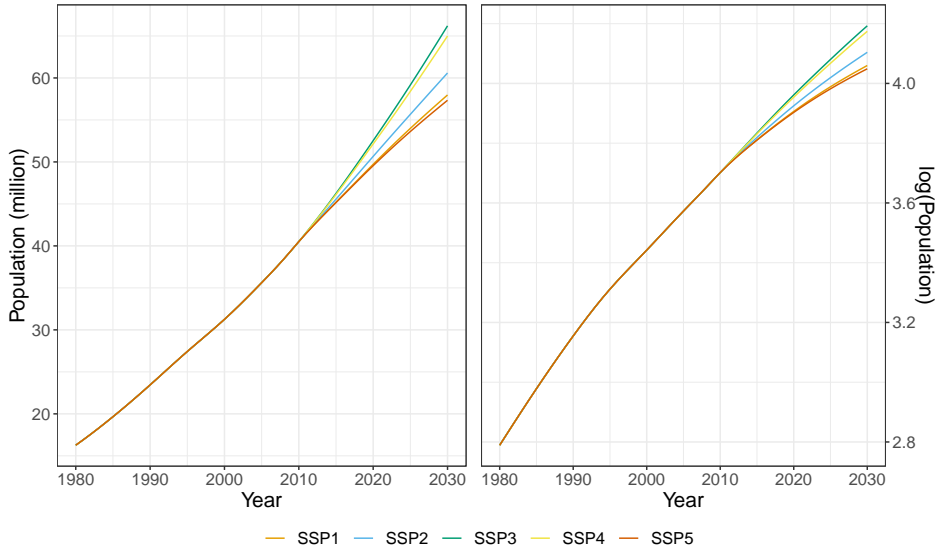


Figure 5.1: Population and log(population) data and projections for Kenya for five different scenarios. Historical data until year 2010.

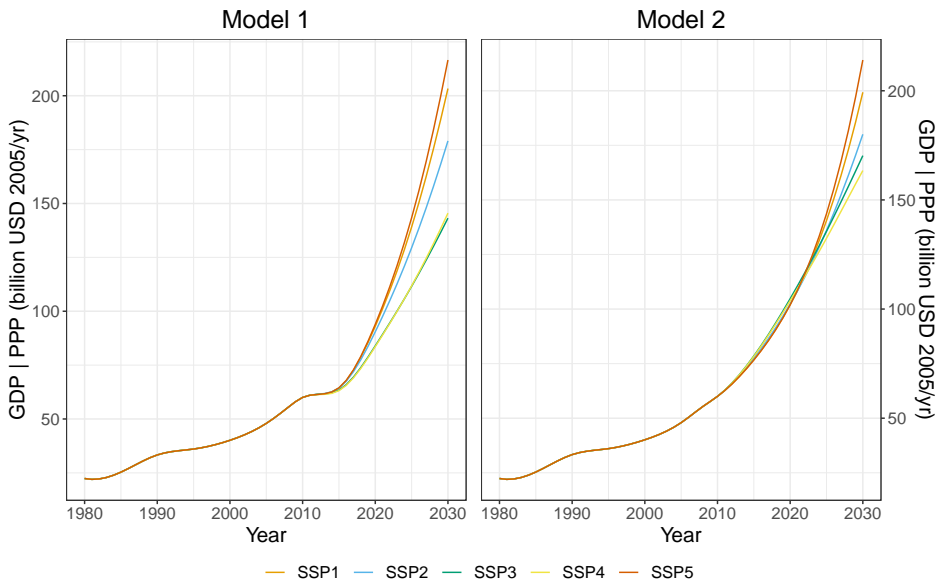


Figure 5.2: GDP data and GDP projections for Kenya for five different scenarios, using two different models. Historical data until year 2010.

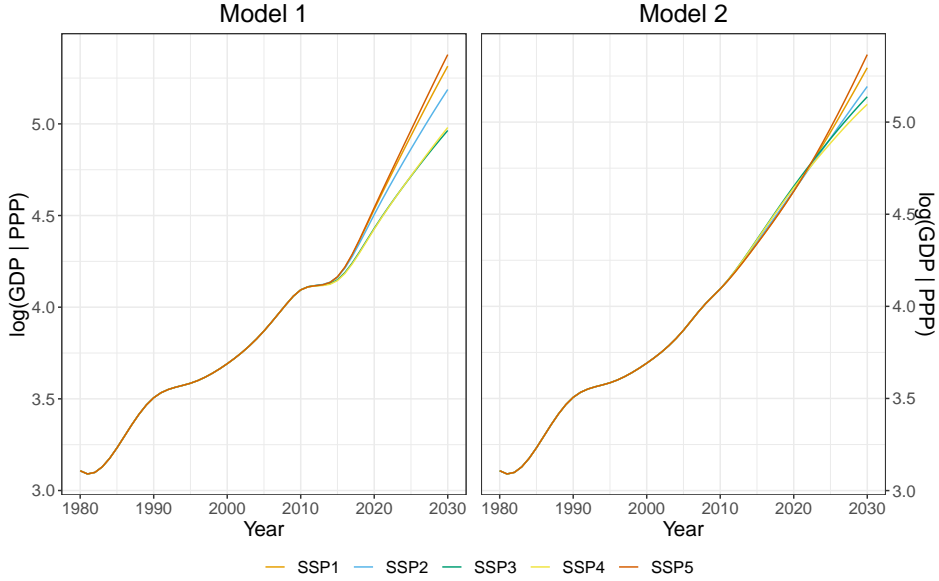


Figure 5.3: $\log(\text{GDP})$ data and $\log(\text{GDP})$ projections for Kenya for five different scenarios, using two different models. Historical data until year 2010.

effect of POP and GDP are confounded in the linear trend, models with and without linear trend are explored. Recall that y_{ts} denotes the logit of the estimated mortality rate for year t , and survey s and that

$$y_{ts} \sim \mathcal{N}(\eta_{ts}, \sigma_{ts}^2), \quad \eta_{ts} = \mu + \xi t + a_t + \mathbf{z}_t^T \boldsymbol{\beta} + \delta_s, \quad t = 1, 2, \dots, \quad s = 1, 2, 3.$$

Here η_{ts} is the logit of the true mortality rate for year t and σ_{ts}^2 is the estimated variance for year t from survey s . μ is an intercept, ξ is an estimated coefficient, a_t is a stochastic process, and \mathbf{z}_t is a vector of covariates with coefficients collected in $\boldsymbol{\beta}$. δ_s is a survey effect and data from three different surveys are used. An overview of the different combinations of covariates and corresponding model names are given in Table 5.1. The 22 different models are fitted with temporal effect AR(2) and RW(1), for NMR and U5MR.

It is interesting to investigate the coefficients of the covariates and their credible intervals. The estimated coefficients are presented in Appendix C, in Table C.1 for NMR and Table C.2 for U5MR. From Table C.1, one can see that none of the coefficients are significant. Based on the previously shown plots of yearly NMR in Kenya (Figure 3.1), plots of simulated NMR (Figure 4.1 and Figure 4.5), and the NMR plots in Appendix B, this is not surprising results; there is no clear trend in change in NMR between 1980 and 2013, and it is, therefore, difficult to extract signals. Table C.2 reveals that for U5MR, several of the coefficients are significant. Using AR(2) as temporal effect makes the coefficients in model 2, 3, 4, 7, 8, 9, and 12 significant. Also, the coefficient of GDP1 in model 19 is significant. RW(1) makes the coefficients in model 3, 4, 8, and $\log(\text{GDP}1)$ in model 16 significant. As one could expect, the signs of the significant coefficients are negative.

Name	Covariates	Name	Covariates
1	-	12	t
2	$\log(\text{POP})$	13	$t, \log(\text{POP})$
3	$\log(\text{GDP1})$	14	$t, \log(\text{GDP1})$
4	$\log(\text{GDP2})$	15	$t, \log(\text{GDP2})$
5	$\log(\text{POP}), \log(\text{GDP1})$	16	$t, \log(\text{POP}), \log(\text{GDP1})$
6	$\log(\text{POP}), \log(\text{GDP2})$	17	$t, \log(\text{POP}), \log(\text{GDP2})$
7	POP	18	t, POP
8	GDP1	19	$t, \text{GDP1}$
9	GDP2	20	$t, \text{GDP2}$
10	POP, GDP1	21	$t, \text{POP}, \text{GDP1}$
11	POP, GDP2	22	$t, \text{POP}, \text{GDP2}$

Table 5.1: Overview of model names and corresponding covariates. t is a linear trend, POP is population size, GDP1 is GDP according to model 1 and GDP2 is GDP according to model 2.

The WAIC of each model and the CRPS and MSE of one-step, two-step, and three-step predictions are computed to evaluate the models' explanatory and predictive strengths. All values are calculated on logit-scale. Since the predictions are to be compared with the design-based estimates from KDHS14, a small modification is made to the input fed to `inla()`; the fourth part of the survey-index column (**idxs**) does not contain NAs, but the index corresponding to KDHS14 (3). When calculating CRPS it is assumed that both the logit of the estimated mortality rates and the logit of the predicted mortality rates follow a normal distribution, therefore Equation (2.19) is used. The calculated MSE considers the design-based estimate of mortality rate from KDHS14 and the corresponding mean prediction made by the models.

Stepwise CRPS and MSE are computed by estimating models using data until year 2005, . . . , 2012, and then computing one-, two-, and three-steps predictions. This makes it possible to compute six 3-step ahead, seven two-steps ahead, and eight one-step ahead CRPS scores and take the average. As the goal is to make predictions until 2030 when the last year of observations is 2013, 17-step ahead predictions are what one is really interested in. However, only data up until and including 1996 could have been used to estimate models if one were to produce a 17-step ahead prediction score. Thus, three-step ahead predictions are considered as a compromise between what is practical to compute and the desire for long-term forecasts.

The WAIC and average CRPS scores for NMR are presented in Table 5.2 while the MSE scores are presented in Table 5.3. The final row in each of the tables is the median difference between the AR(2) and RW(1) models. A negative median indicates that the AR(2) models most often perform best, while a positive value indicates that RW(1) models most often perform best. The median differences presented in the last row of Table 5.2 show that in general, the models fitted with temporal effect AR(2) have lower WAIC and CRPS scores than the corresponding models with temporal effect RW(1). From the median differences in Table 5.3 one can see that models with temporal process AR(2) most often have the lowest three-step ahead prediction MSE, whereas RW(1) processes most often have the lowest one and two-step ahead prediction MSE. Overall, model 15 with temporal

effect AR(2) performs best with respect to prediction as it has the lowest MSE and CRPS score for two- and three-step predictions. Consequently, it is decided to study model 15 further. However, model 14 is the equivalence of model 15, using GDP model 1. It is not known which of the two GDP projections that are most accurate, and the SSP database recommends that both projections are used. In that way, the results can be compared, and the model's sensitivity to GDP projections evaluated. Therefore model 14 is also considered when moving forward. For comparison, model 12 (only linear trend) fitted with AR(2) is also studied further.

Different conclusions can be drawn for U5MR based on Table 5.4 and Table 5.5. The median values presented in the two tables show that models with temporal effect RW(1) in general have lower CRPS and MSE scores for one-, two-, and three-step predictions, but that models with AR(2) have the lowest WAIC. Thus, the models best describing the data are not necessarily the best at producing forecasts. Three models stand out in both tables; model 8, 10, and 16. Model 10 has the lowest CRPS and MSE scores for models with temporal effect AR(2) and lowest one-step ahead prediction MSE and CRPS score with temporal effect RW(1). Model 16 is almost consistently in second place. Model 8 with temporal effect RW(1) has the overall lowest three-step ahead prediction CRPS score and has competitive scores otherwise. The goal of this thesis is to produce forecasts far into the future with reasonable credible intervals. Model 8 with temporal effect RW(1) is therefore chosen for analysis of Kenya's progress towards the SDGs and MDGs for U5MR. Model 8 and model 9 have the same covariate, for two alternative futures. Therefore model 9 also has to be considered. For comparison, model 12 is also interesting.

Based on this analysis it is chosen to evaluate Kenya's progress towards the SDGs and MDGs using model 14, 15 and 12 with temporal effect AR(2) for NMR and model 8, 9 and 12 with temporal effect RW(1) for U5MR.

Model	WAIC		One-step CRPS $\times 10^{-02}$		Two-step CRPS $\times 10^{-02}$		Three-step CRPS $\times 10^{-02}$	
	AR(2)	RW(1)	AR(2)	RW(1)	AR(2)	RW(1)	AR(2)	RW(1)
1	19.07	30.62	2.16	3.85	2.17	3.73	2.37	3.89
2	20.32	31.58	2.13	2.97	2.22	2.87	2.71	3.93
3	20.35	31.43	1.92	2.68	2.03	2.56	2.63	3.56
4	20.36	31.34	1.91	2.64	2.02	2.50	2.63	3.52
5	20.94	31.44	1.51	2.38	1.77	2.35	2.71	3.45
6	20.49	30.33	1.49	2.20	1.67	2.07	2.64	3.18
7	20.33	31.50	2.26	2.96	2.41	2.90	2.92	4.11
8	20.38	31.42	2.00	2.56	2.16	2.50	2.94	3.80
9	20.42	31.29	1.99	2.50	2.16	2.42	2.96	3.76
10	20.78	31.17	1.74	2.25	2.55	2.80	4.66	5.12
11	20.30	29.83	1.81	2.19	2.51	2.50	4.65	4.84
12	20.32	31.74	2.24	3.02	2.37	2.96	2.83	4.08
13	21.24	33.02	3.72	3.52	4.33	3.71	5.43	5.28
14	20.80	31.86	1.72	2.61	1.76	2.56	2.27	3.49
15	20.24	30.29	1.55	2.43	1.49	2.17	1.98	2.98
16	21.91	33.14	2.23	2.33	2.55	2.33	3.75	3.94
17	21.29	32.38	2.06	2.16	2.22	2.00	3.40	3.62
18	21.40	31.29	2.76	2.09	3.52	2.07	5.75	4.57
19	20.84	30.79	1.79	2.22	2.78	2.96	5.31	5.87
20	20.45	29.90	1.87	2.20	2.78	2.76	5.35	5.69
21	22.00	33.36	2.44	2.18	3.77	3.12	6.53	6.61
22	21.20	32.38	2.14	2.13	2.97	2.70	5.52	6.11
Median difference	-11.05		-0.54		-0.29		-0.66	

Table 5.2: WAIC and step-wise prediction CRPS $\times 10^{-02}$ for NMR, at logit scale. The median difference is the median of the difference between the CRPS score for the same models fitted by AR(2) and RW(1). The values marked with green are the lowest values in the column, while the yellow values are the second lowest.

Model	One-step MSE $\times 10^{-02}$		Two-step MSE $\times 10^{-02}$		Three-step MSE $\times 10^{-02}$	
	AR(2)	RW(1)	AR(2)	RW(1)	AR(2)	RW(1)
1	1.22	1.34	1.25	1.41	1.39	1.64
2	1.24	1.23	1.31	1.32	1.64	2.05
3	1.11	1.13	1.18	1.21	1.59	2.02
4	1.11	1.13	1.18	1.20	1.60	2.02
5	0.85	0.90	1.01	1.05	1.66	1.92
6	0.84	0.88	0.94	0.96	1.60	1.85
7	1.33	1.27	1.43	1.38	1.80	2.22
8	1.16	1.14	1.26	1.24	1.80	2.31
9	1.16	1.13	1.26	1.23	1.82	2.32
10	0.98	0.97	1.52	1.49	3.24	3.41
11	1.02	1.00	1.46	1.41	3.18	3.34
12	1.31	1.27	1.40	1.38	1.74	2.15
13	2.32	1.69	2.78	1.95	3.67	3.11
14	0.98	0.98	1.01	1.07	1.36	1.75
15	0.88	0.90	0.84	0.91	1.17	1.54
16	1.28	1.10	1.39	1.21	2.21	2.47
17	1.16	1.03	1.16	1.05	1.95	2.31
18	1.65	1.06	2.03	1.11	3.73	3.10
19	0.99	0.99	1.65	1.66	3.84	4.15
20	1.04	1.05	1.62	1.63	3.81	4.13
21	1.34	1.20	2.07	1.92	4.35	5.08
22	1.11	1.13	1.44	1.62	3.44	4.73
Median difference	0.01		0.01		-0.37	

Table 5.3: Step-wise prediction MSE $\times 10^{-02}$ for NMR, at logit scale. The median difference is the median of the difference between the MSE for the same models fitted by AR(2) and RW(1). The values marked with green are the lowest values in the column, while the yellow values are the second lowest.

Model	WAIC		One-step CRPS $\times 10^{-02}$		Two-step CRPS $\times 10^{-02}$		Three-step CRPS $\times 10^{-02}$	
	AR(2)	RW(1)	AR(2)	RW(1)	AR(2)	RW(1)	AR(2)	RW(1)
1	-84.28	-84.42	4.79	3.65	7.71	4.98	12.16	7.43
2	-84.02	-83.97	3.63	2.82	4.91	3.10	7.52	4.61
3	-84.04	-83.79	2.98	2.23	3.60	2.06	5.59	3.24
4	-83.35	-82.99	3.15	2.48	3.63	2.24	5.45	3.31
5	-82.63	-82.14	2.43	1.97	2.68	1.72	4.25	2.91
6	-82.13	-81.63	2.78	2.41	2.81	2.05	4.13	3.09
7	-83.80	-83.55	3.16	2.48	3.67	2.26	5.53	3.38
8	-83.48	-82.93	2.18	1.79	1.91	1.28	3.13	2.48
9	-82.32	-81.50	2.66	2.48	2.36	2.07	3.44	3.21
10	-82.20	-80.13	1.58	1.68	1.18	1.48	2.56	3.26
11	-81.90	-81.41	2.82	2.81	2.60	2.75	3.98	4.48
12	-83.97	-83.78	3.42	2.66	4.38	2.72	6.62	4.05
13	-83.09	-82.92	2.71	2.45	2.84	2.20	4.50	3.43
14	-83.42	-83.23	2.71	2.14	3.23	1.96	5.19	3.21
15	-82.82	-82.53	2.92	2.43	3.27	2.17	5.02	3.29
16	-80.56	-75.83	1.73	1.78	1.37	1.37	2.61	2.64
17	-81.13	-79.50	2.44	2.52	2.11	2.16	3.29	3.35
18	-82.39	-82.05	2.72	2.82	2.65	3.00	4.88	5.73
19	-82.16	-77.94	1.63	1.81	1.40	1.83	3.17	4.00
20	-81.80	-81.25	2.99	3.03	3.09	3.25	4.95	5.43
21	-81.77	-79.59	2.40	2.95	3.14	4.52	6.83	9.24
22	-81.92	-81.35	3.45	3.98	4.55	5.82	8.28	10.75
Median difference	-0.49		0.32		0.63		0.85	

Table 5.4: WAIC and step-wise prediction CRPS $\times 10^{-02}$ for U5MR, at logit scale. The median difference is the median of the difference between the CRPS score for the same models fitted by AR(2) and RW(1). The values marked with green are the lowest values in the column, while the yellow values are the second lowest.

Model	One-step MSE $\times 10^{-02}$		Two-step MSE $\times 10^{-02}$		Three-step MSE $\times 10^{-02}$	
	AR(2)	RW(1)	AR(2)	RW(1)	AR(2)	RW(1)
1	1.97	1.38	3.72	2.14	6.98	3.67
2	1.44	1.04	2.20	1.26	3.89	2.12
3	1.15	0.80	1.54	0.80	2.73	1.41
4	1.23	0.90	1.54	0.87	2.64	1.43
5	0.93	0.71	1.09	0.66	1.97	1.27
6	1.09	0.89	1.15	0.80	1.88	1.33
7	1.23	0.91	1.57	0.89	2.70	1.47
8	0.82	0.64	0.73	0.48	1.35	1.08
9	1.04	0.95	0.93	0.85	1.48	1.45
10	0.61	0.63	0.38	0.59	1.04	1.58
11	1.20	1.15	1.12	1.21	1.86	2.25
12	1.35	0.98	1.92	1.09	3.34	1.82
13	1.06	0.92	1.17	0.87	2.10	1.51
14	1.06	0.78	1.37	0.76	2.52	1.41
15	1.15	0.90	1.38	0.85	2.41	1.43
16	0.66	0.64	0.48	0.52	1.08	1.19
17	0.98	0.96	0.83	0.88	1.43	1.53
18	1.14	1.16	1.07	1.31	2.29	2.95
19	0.64	0.68	0.49	0.76	1.42	2.04
20	1.30	1.26	1.39	1.47	2.51	2.85
21	1.01	1.23	1.35	2.08	3.86	5.36
22	1.53	1.69	2.17	2.76	4.90	6.33
Median difference	0.16		0.27		0.41	

Table 5.5: Step-wise prediction MSE $\times 10^{-02}$ for USMR, at logit scale. The median difference is the median of the difference between the MSE for the same models fitted by AR(2) and RW(1). The values marked with green are the lowest values in the column, while the yellow values are the second lowest.

5.2 Estimation and forecasting

At this point the chosen models are used to forecast mortality rates in Kenya until 2030. Forecasts of NMR are presented in Figure 5.4, Figure 5.5 and Figure 5.6 for model 12, 15, and 14, respectively. Forecasts for U5MR are presented in Figure 5.7, Figure 5.8 and Figure 5.9 for model 12, 8, and 9, respectively. In the figures, SSP0 refers to "no scenario", which is always the case for year < 2011 and for models that do not include any covariates.

From Figure 5.4, Figure 5.5 and Figure 5.6 one can see that the models with covariates produce forecasts with wider credible intervals. This could be due to the relatively large credible intervals of the coefficients, especially for that of $\log(\text{GDP})$ in both model 14 and model 15 (see Table C.1). Several different linear combinations of $\log(\text{GDP})$ and t might explain the observed data well, leading to the uncertain estimates of the coefficients of the covariates. If the relative trend of the covariates change outside the interval for which one has observations, predictions become even more uncertain. The same is not true for U5MR, based on Figure 5.7, Figure 5.8 and Figure 5.9. The models with GDP as covariate, model 8 and model 9, have tighter credible intervals than the model with only a linear trend, model 12. For both NMR and U5MR one can see that the models with covariates produce forecasts whose mean decrease more rapidly compared to the forecasts made by model 12.

Recall that SDG target 3.2 is to reduce NMR to "at least as low as 12 per 1000 live births" (UN (2019a)) and U5MR to "at least as low as 25 per 1000 live birth" (UN (2019a)) by 2030, while MDG target 4A was to reduce U5MR by two thirds between 1990 and 2015. As it is difficult to assess the progress towards the SDGs and MDGs based on the figures, Table 5.6 and Table 5.7 presents the estimated/forecasted mortality rates for key years, namely 1990, 2015 and 2030. When comparing the values in Table 5.6 and Table 5.7 one can see that in 2030 the median predicted U5MR is lower than the median predicted NMR, which is not realistic.

Using the median values for NMR and U5MR presented in Table 5.6 and Table 5.7 one gets that SDG target 3.2 is reached for NMR only in scenario SSP5 using model 15 (rounding to the nearest whole number), and for U5MR in all scenarios using model 8 or 9. When considering SSP0 or SSP2 (middle of the road) the projected NMR per 1000 live birth in 2030 is 26.50 for model 12, 15.51 for model 14 and 14.93 for model 15. For U5MR, the corresponding median projected values for 2030 are 39.06 using model 12, 4.86 using model 8 and 9.18 using model 9. Thus it is clear that the median predictions are sensitive to model choice. However, compared to the size of the credible intervals, the differences in predictions for the different models and scenarios are small for both NMR and U5MR.

As the forecasts are probabilistic, it is more interesting to look at the probability of reaching SDG target 3.2. That is done by using the marginal posterior of the linear predictor. The probabilities of reaching SDG target 3.2 in scenario SS0 or SSP2 are presented in Table 5.8. Table 5.8 further highlights the fact that the models with covariates give very different projections than the models without covariates. The probability of reaching SDG target 3.2 for NMR is 0.19% for model 12, but 29.28% and 29.63% for model 8 and 9. Thus, the probability of reaching SDG target 3.2 for NMR is not that sensitive to the choice of GDP model. The probability of reaching SDG target 3.2 for U5MR is more sensitive

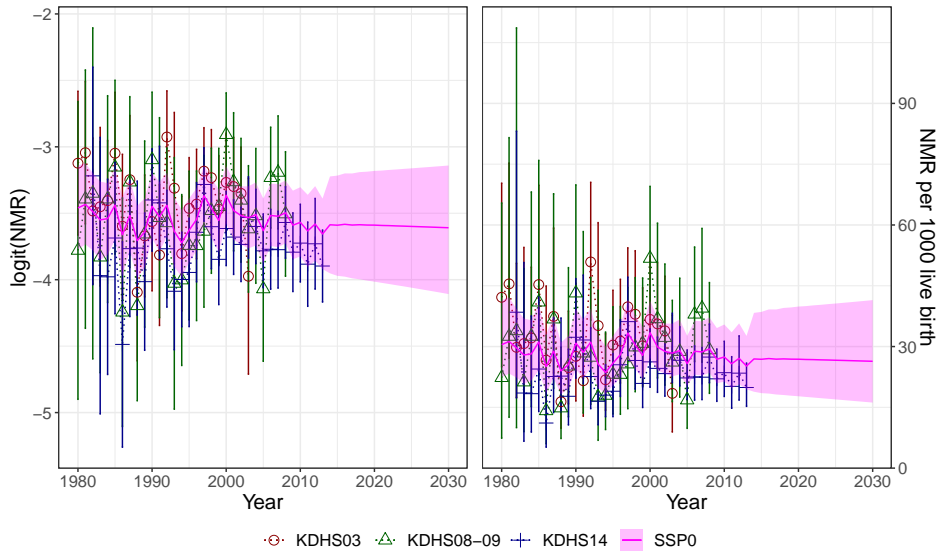


Figure 5.4: NMR for Kenya fitted using model 12 (linear trend, no covariates), using AR(2) as temporal effect and estimated NMR from 1980-2013.

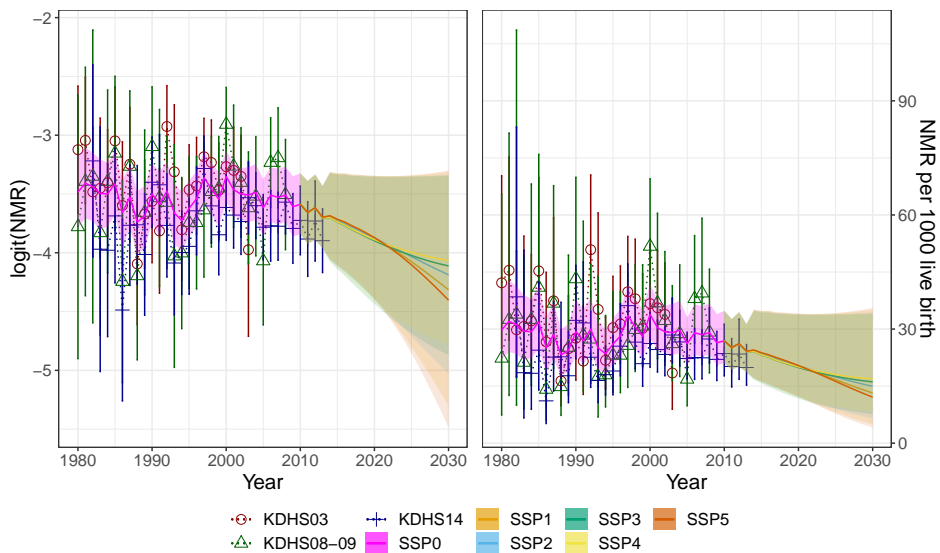


Figure 5.5: NMR for Kenya fitted using model 15 (t and $\log(\text{GDP}2)$ as covariates), using AR(2) as temporal effect and estimated NMR from 1980-2013.

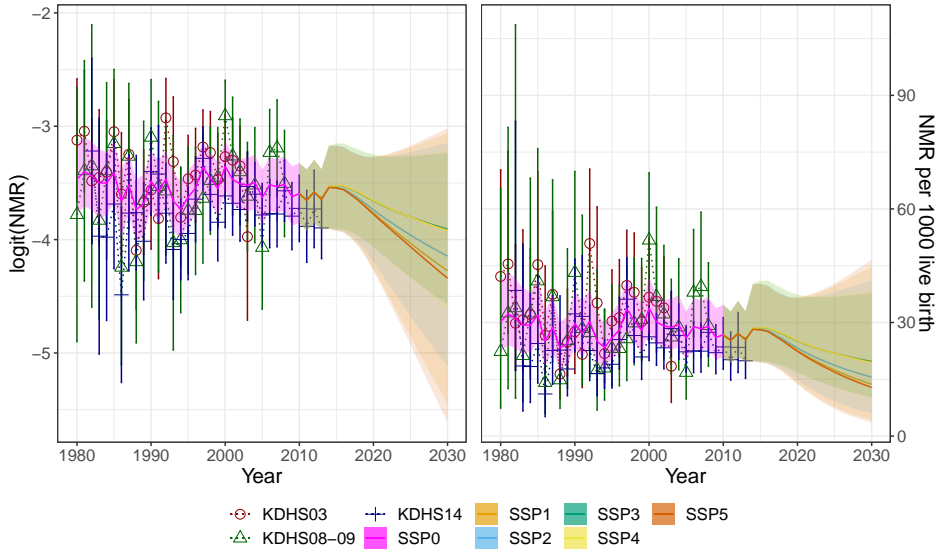


Figure 5.6: NMR for Kenya fitted using model 14 (t and $\log(\text{GDP1})$ as covariates), using AR(2) as temporal effect and estimated NMR from 1980-2013.

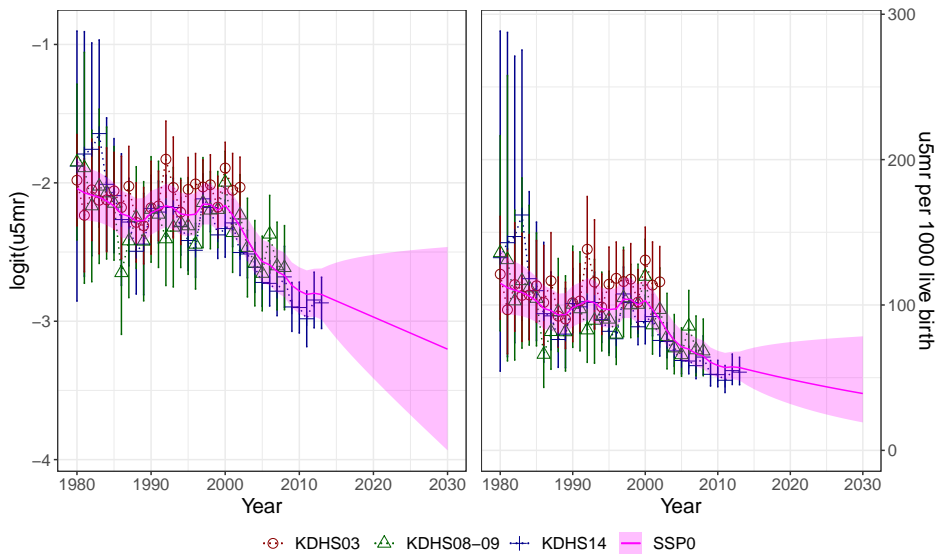


Figure 5.7: U5MR for Kenya fitted using model 12 (linear trend, no covariates), using RW(1) as temporal effect and estimated U5MR from 1980-2013.

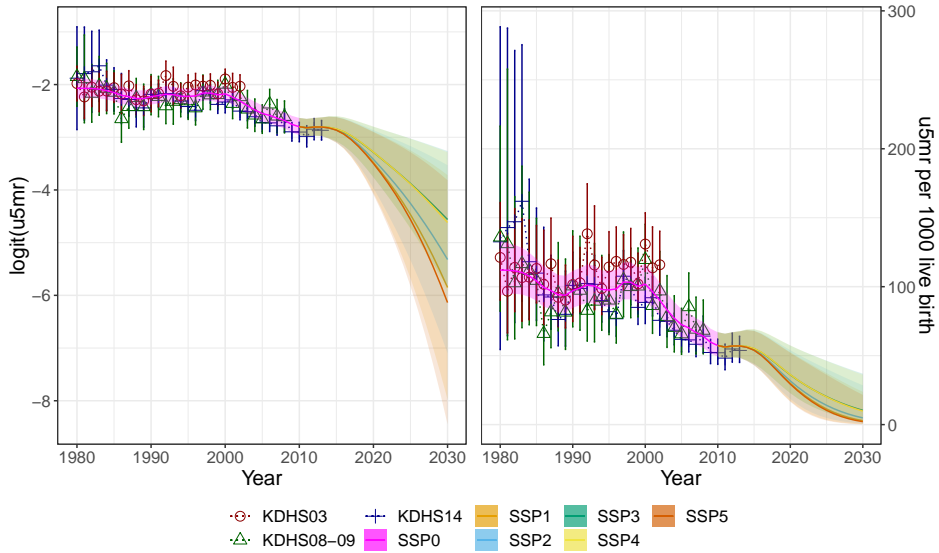


Figure 5.8: U5MR for Kenya fitted using model 8 (GDP1 as covariate), using RW(1) as temporal effect and U5MR rates from 1980-2013.

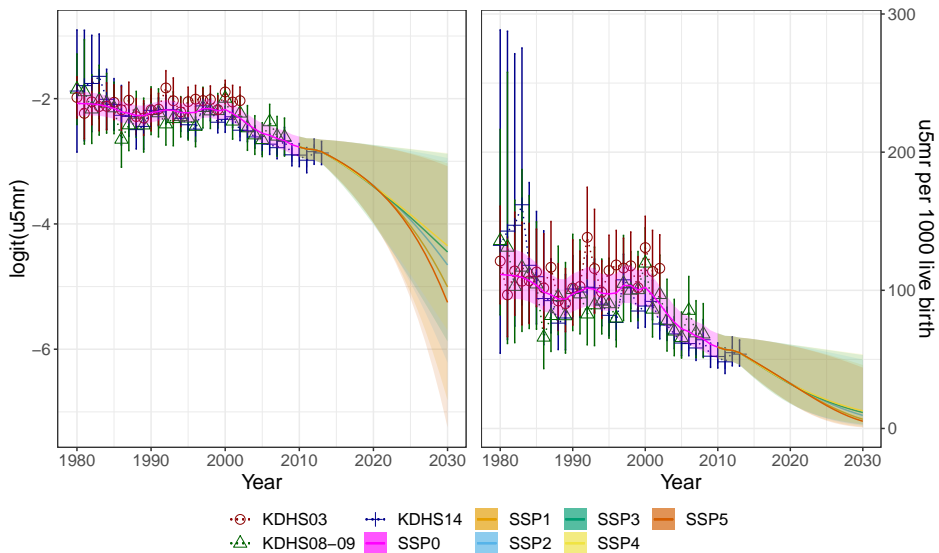


Figure 5.9: U5MR for Kenya fitted using model 9 (GDP2 as covariate), using RW(1) as temporal effect and U5MR rates from 1980-2013.

Model 12 (liner trend,no covariates)			
Scenario	1990	2015	2030
SSP0	30.77 [24.38, 39.32]	26.90 [18.51, 38.55]	26.50 [16.17, 41.46]

Model 14 (t , log(GDP1))			
Scenario	1990	2015	2030
SSP1	29.67 [23.33, 38.32]	28.12 [19.46, 40.22]	13.63 [4.29, 44.82]
SSP2	29.63 [23.31, 38.25]	28.18 [19.56, 40.21]	15.51 [5.96, 41.19]
SSP3	29.62 [23.30, 38.23]	28.58 [19.74, 40.91]	19.68 [10.27, 37.81]
SSP4	29.68 [23.34, 38.33]	28.64 [19.64, 41.23]	19.40 [9.84, 38.42]
SSP5	29.67 [23.33, 38.33]	28.10 [19.44, 40.20]	12.77 [3.64, 46.60]

Model 15(t , log(GDP2))			
Scenario	1990	2015	2030
SSP1	29.28 [23.03, 37.79]	23.73 [16.19, 34.75]	13.15 [4.97, 35.09]
SSP2	29.24 [23.04, 37.74]	23.53 [16.04, 34.54]	14.93 [6.51, 34.35]
SSP3	29.25 [23.02, 37.78]	23.21 [15.70, 34.33]	16.09 [7.63, 34.01]
SSP4	29.27 [23.06, 37.81]	23.26 [15.73, 34.41]	16.87 [8.35, 34.12]
SSP5	29.28 [23.03, 37.79]	23.76 [16.22, 34.77]	12.06 [4.13, 35.59]

Table 5.6: Median estimated and forecasted NMR in Kenya in 1990, 2015 and 2030, with 95% credible intervals.

Model 12 (linear trend, no covariates)			
Scenario	1990	2015	2030
SSP0	98.20 [85.13, 113.24]	54.44 [42.12, 70.69]	39.06 [19.21, 78.44]

Model 8 (GDP1)			
Scenario	1990	2015	2030
SSP1	97.09 [84.65, 111.43]	53.91 [43.12, 67.99]	2.86 [0.34, 23.72]
SSP2	97.09 [84.65, 111.43]	54.05 [43.24, 68.15]	4.86 [0.83, 28.48]
SSP3	97.08 [84.65, 111.42]	54.97 [44.09, 69.11]	10.33 [2.87, 36.92]
SSP4	97.08 [84.65, 111.42]	55.18 [44.27, 69.31]	9.77 [2.62, 36.07]
SSP5	97.09 [84.65, 111.42]	53.86 [43.07, 67.94]	2.15 [0.21, 21.55]

Model 9 (GDP2)			
Scenario	1990	2015	2030
SSP1	97.36 [84.70, 111.99]	48.41 [37.19, 64.65]	6.46 [1.09, 45.93]
SSP2	97.38 [84.70, 112.02]	48.23 [36.96, 64.63]	9.18 [1.99, 49.84]
SSP3	97.40 [84.71, 112.07]	47.84 [36.46, 64.55]	11.27 [2.80, 53.15]
SSP4	97.39 [84.70, 112.06]	47.90 [36.55, 64.56]	12.47 [3.34, 53.77]
SSP5	97.36 [84.70, 111.99]	48.46 [37.25, 64.68]	5.04 [0.71, 44.02]

Table 5.7: Median estimated and forecasted U5MR in Kenya in 1990, 2015 and 2030, with 95% credible intervals.

Rate	Model 12	Model 14/8	Model 15/9
NMR	0.19 %	29.28%	29.63%
U5MR	9.82%	96.73 %	89.35%

Table 5.8: Probability of reaching SDG 3 in scenario SSP0 (model 12) or SSP2 (model 8 and 9 for U5MR, and model 14 and 15 for NMR).

to the choice of GDP model, with a probability of 96.73% for model 8 and 89.35% for model 9. For comparison, the probability of reaching SDG target 3.2 for U5MR is 9.82% using model 12.

To properly assess if Kenya reached MDG target 4A, the joint posterior distribution of the linear predictor for the year 1990 and the year 2015 should be used. However, in this thesis, the discussion is limited to an evaluation of median values, presented in Table 5.7. When rounding to the nearest child, a 2/3 reduction between 1990 and 2015 in U5MR equals a U5MR of 33 per 1000 live birth for model 12 and a U5MR of 32 per 1000 live birth using model 8 and 9. Thus, MDG target 4A was not reached regardless of scenario and model. Model 12 gives a U5MR of 54.44 per 1000 live birth in 2015, while model 8 gives 54.05 and model 9 gives 48.23, both for scenario SSP2.

The analysis concludes that the probability of reaching SDG target 3.2 is heavily dependent on the choice of model. If using model 12, the probability of reaching SDG target 3.2 is 0.19% for NMR and 9.82% for U5MR. The probability of reaching the target increases when using models with covariates. If GDP model 1 is accurate, the probability of reaching SDG 3 in scenario SSP2 is 29.28% for NMR and 96.73% for U5MR. If GDP model 2 is accurate, the probabilities change to 29.63% for NMR and 89.35% for U5MR. Further, the analysis shows that Kenya did not reach MDG target 4A. For NMR including covariates generally increased the size of credible intervals of predictions, whereas the opposite is true for U5MR. Modeling NMR is difficult as there is no clear trend in the change in NMR between 1980-2013.

For a complete analysis, the effect of other covariates than those explored here could be considered. Besides using information about Kenya, it is also possible to include information from countries that are similar to Kenya or who are likely to influence Kenya. An example of two such countries is Uganda and Tanzania, two of Kenya's neighbors. In the following chapter, models that simultaneously estimate child mortality rates in those three countries are analyzed and compared to the best-performing models from this chapter.

Joint Modeling of Child Mortality

This chapter explores the effect of including information from two neighboring countries, Uganda and Tanzania, in models for NMR and U5MR in Kenya. Section 6.1 presents the estimated mortality rates for Uganda and Tanzania and the effect of joint modeling on forecasts is investigated. Section 6.1 aims to determine if more reliable forecasts for one country can be produced by using data from neighboring countries as well, compared to only using data from the country of interest. The goal of Section 6.2 is to examine if estimates of mortality rates for a country for which recent data is not available can be improved by borrowing strength from neighboring countries for which more recent data is available.

6.1 Effect of borrowing strength from neighboring countries on forecast

Data from three separate DHS surveys undertaken in Uganda in 2006, 2011, and 2016 is used to produce design-based estimates of NMR and U5MR in Uganda. The surveys are henceforth referred to as UDGS06, UDHS11, and UDHS16. The survey data is coded in a similar way to the KDHS data, so the mortality rates are computed as explained in Section 3.1. Design-based estimates of NMR and U5MR in Tanzania are produced in the same way, using data from DHS surveys undertaken in 2004-2005, 2010 and 2015-2016. The surveys are henceforth referred to as TDGS05, TDHS10, and TDHS16.

The NMR in Uganda from 1980-2015 is given in Figure 6.1 and the U5MR is given in Figure 6.2. The estimated NMR and U5MR for Tanzania are presented in Figure 6.3 and Figure 6.4, respectively. By comparing Figure 3.1, Figure 6.1 and Figure 6.3 one can see that the behaviour of NMR is quite similar in all three countries. The same can be said for U5MR by comparing Figure 3.3, Figure 6.2 and Figure 6.4. Thus there are reasons to believe that models of child mortality rates in Kenya might improve by including information from Uganda and Tanzania.

In this section, multivariate time models as explained in Section 2.5 with latent struc-

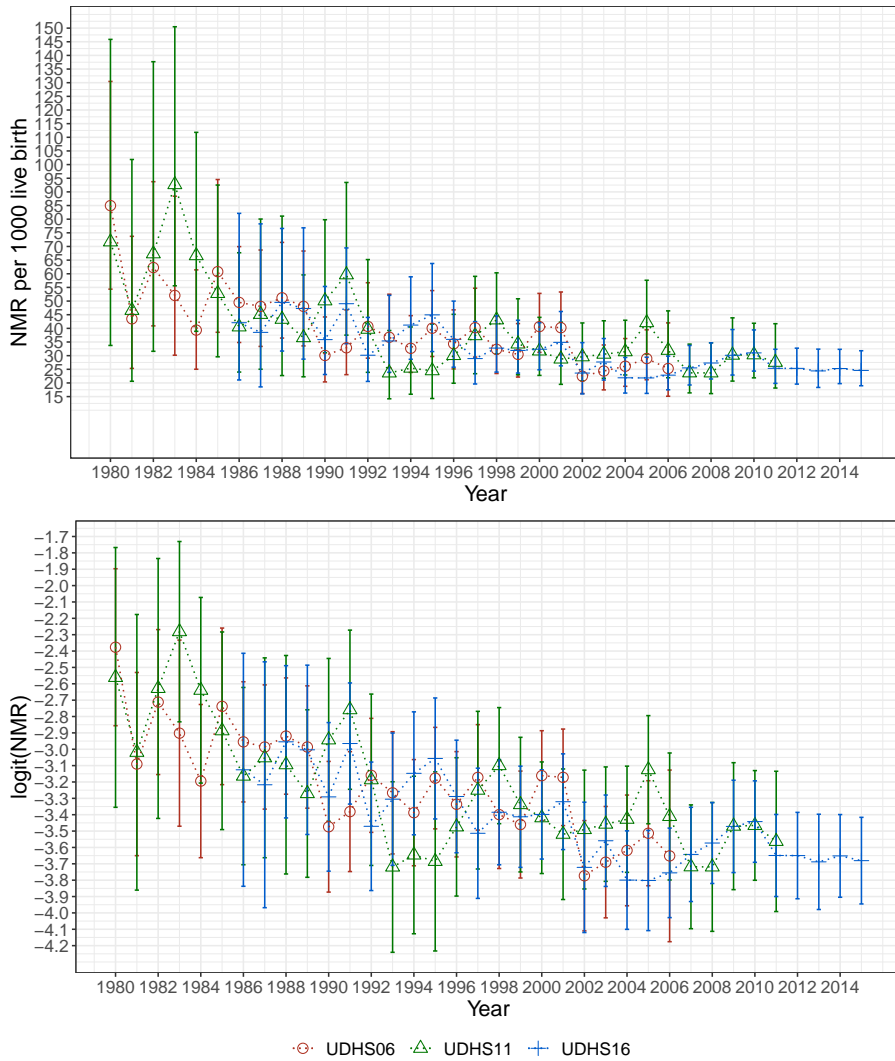


Figure 6.1: Design-based NMR per 1000 live birth and $\text{logit}(\text{NMR})$ in Uganda with 95% credible interval, calculated separately for each year from 1980 to 2015, using data from three independent UDHS surveys.

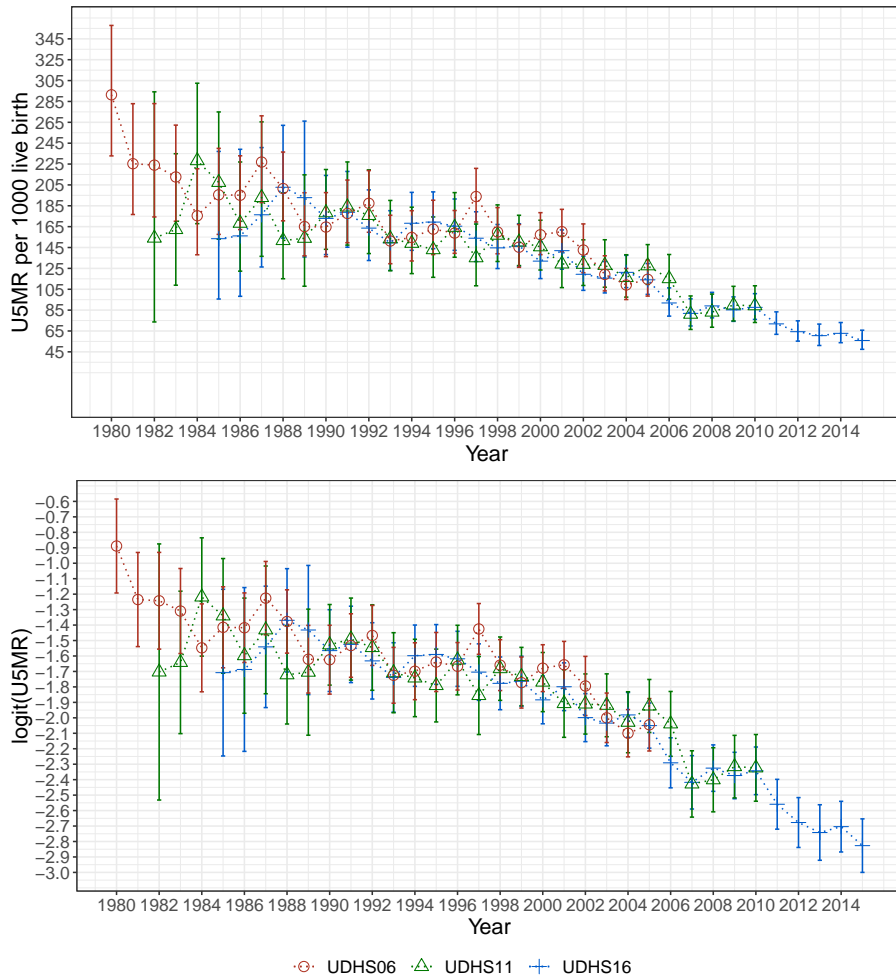


Figure 6.2: Design-based U5MR and logit(U5MR) in Uganda with 95% credible interval, calculated separately for each year from 1980 to 2015, using data from three independent UDHS surveys.

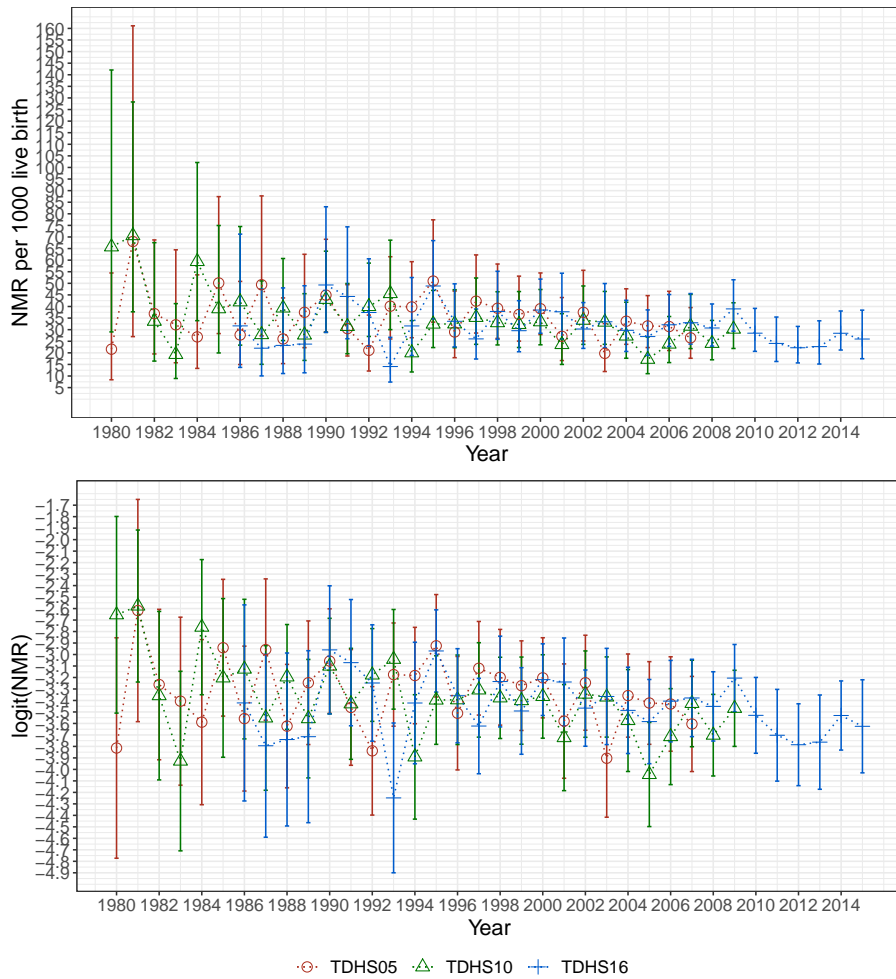


Figure 6.3: Design-based NMR per 1000 live birth and $\text{logit}(\text{NMR})$ in Tanzania with 95% credible interval, calculated separately for each year from 1980 to 2014, using data from three independent UDHS surveys.

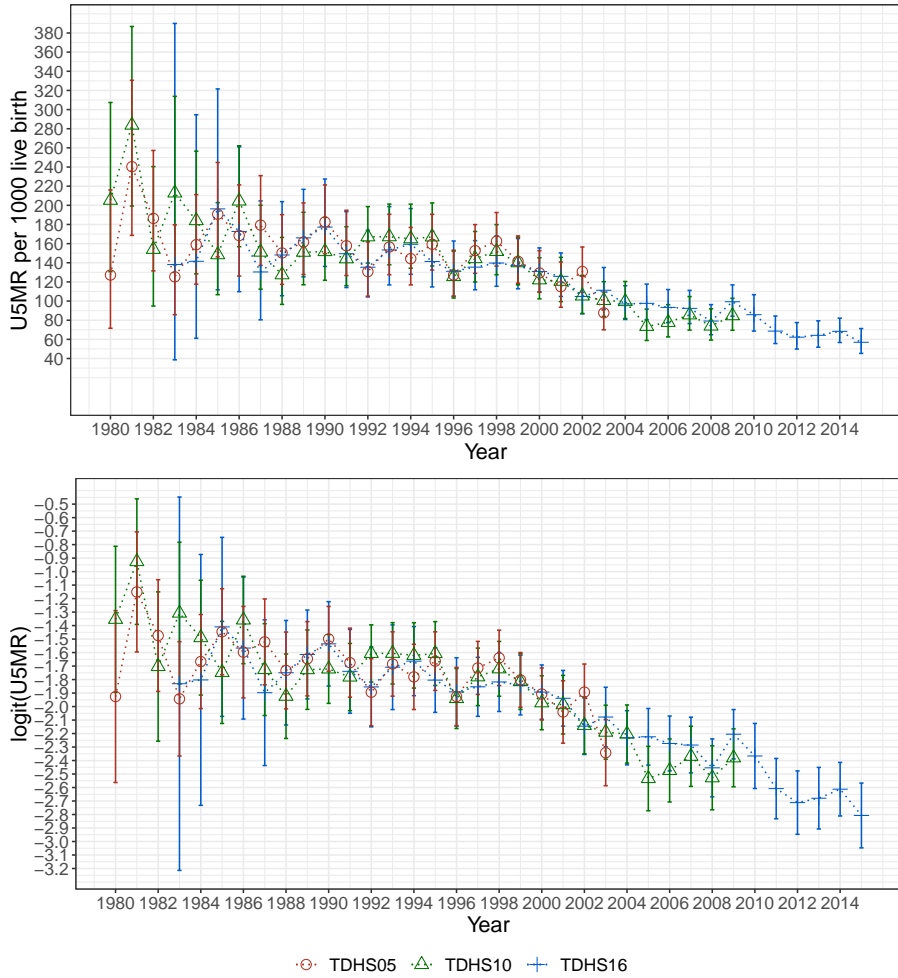


Figure 6.4: Design-based U5MR and $\text{logit}(\text{U5MR})$ in Tanzania with 95% credible interval, calculated separately for each year from 1980 to 2014, using data from three independent UDHS surveys.

ture defined by Equation (2.15) are explored. Recall that the logit of the design-based estimate of the mortality rate for country c and year t estimated using data from survey s is denoted y_{cts} , is fixed and known. Further, $y_{cts} \sim \mathcal{N}(\eta_{cts}, \sigma_{cts}^2)$, where $c = 1, 2, \dots, k$, k is the number of countries considered, $t = 1, 2, \dots$, $s = 1, 2, \dots, S$, where S is the total number of surveys used and

$$\eta_{cts} = \mu_c + \xi_c t + a_{ct} + \delta_s + \mathbf{z}_{ct}^T \boldsymbol{\beta}.$$

μ_c is a country-specific intercept, ξ_c is a country-specific linear trend in time, a_{ct} is a vector autoregressive temporal effect and δ_s is a survey effect. The last term is needed when additional covariates are considered; \mathbf{z}_{ct} is a vector of country specific, time-dependent covariates with estimated coefficients collected in $\boldsymbol{\beta}$. The model is estimated as explained in Section 3.2 using data from Kenya, Uganda, and Tanzania. For NMR models equivalent to model 12, 14 and 15 from Chapter 5 are explored, and they are from this point onward called model 12M, 14M, and 15M, respectively. For U5MR, models equivalent to model 12, 8 and 9 from Chapter 5 are explored, and referred to as model 12M, 8M, and 9M. Using the results from Chapter 5, NMR is fitted using a model with temporal effect AR(2) while U5MR is fitted using a model with temporal effect RW(1).

Even though the focus of this thesis is mortality rates in Kenya, the estimated models can also be used to estimate and predict mortality rates in Uganda and Tanzania. Estimated and forecasted NMR for the three countries using model 15M are presented in Figure 6.5, while corresponding plots for U5MR estimated by model 8M is given in Figure 6.6. For years after 2010, SSP2 data is used for GDP. Please note that although estimated mortality rates for Kenya is only available until 2013, estimated mortality rates for 2014 and 2015 for Uganda and Tanzania are also available.

It is desirable to compare estimations and forecasts of mortality rates in Kenya as produced by model 15 and 15M for NMR and model 8 and 8M for U5MR. Estimated and forecasted NMR is presented in Figure 6.7, while U5MR is presented in Figure 6.8. From Figure 6.7 and Figure 6.8 one can see that the multivariate time models have tighter credible intervals. The fitted/predicted values appear to be more smooth when fitted using the multivariate time model. The predicted NMR is increasing using model 15M and decreasing using model 15. The predicted U5MR is lower for model 8M compared to model 8. For the interested reader, plots of model 14M and 15M for NMR and 8M and 9M for U5MR for Kenya for all five scenarios are given in Appendix D. The plots presented in Appendix D show no major differences between model 14M and 15M or model 8M and 9M, other than that the predictions produced by model 14M are a bit higher than those of model 15M.

To explore the difference in the univariate and multivariate models further, the estimated parameters of the best models from Chapter 5, model 15 with AR(2) for NMR and model 8 with RW(1) for U5MR, are compared to the estimated parameters of the equivalent multivariate models (model 15M and model 8M). The estimated parameters for NMR are presented in Table 6.1, while the parameters for the U5MR models are presented in Table 6.2. The covariates (including linear trend) are transformed using \exp , so the effect is a multiplicative effect on odds-ratio on an increase of one unit of the covariate. Thus, an estimated effect less than one decreases the mortality rate, while an effect greater than one increases it.

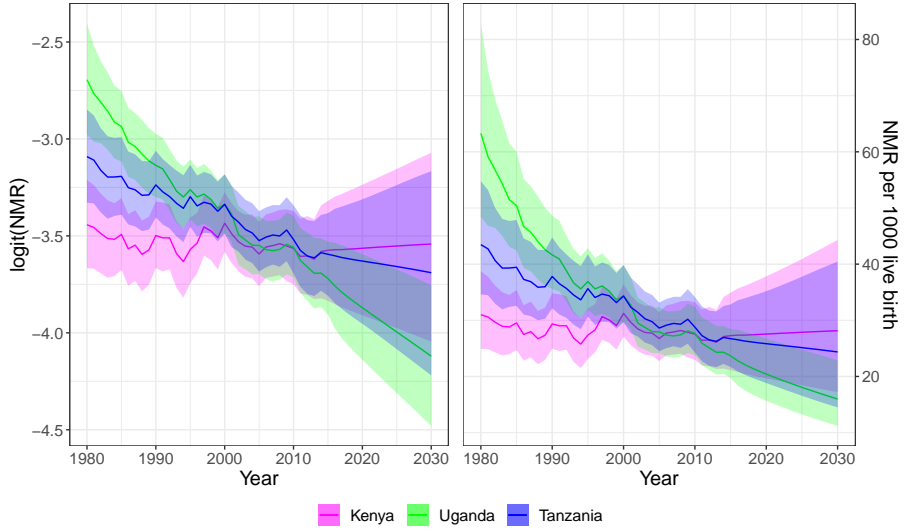


Figure 6.5: Estimated and predicted NMR and logit(NMR) in Kenya, Uganda and Tanzania with 95% credible intervals, estimated model 15M with temporal effect AR(2).

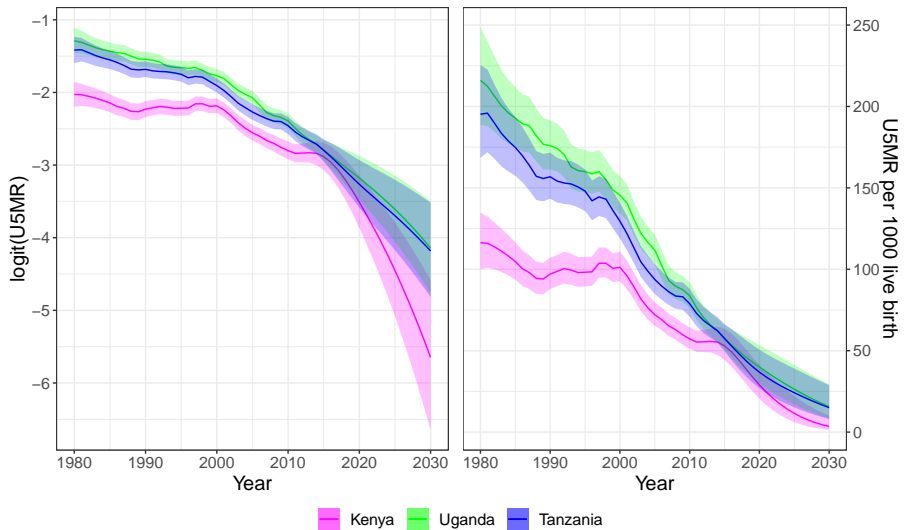


Figure 6.6: Estimated and predicted U5MR and logit(U5MR) in Kenya, Uganda and Tanzania with 95% credible intervals, estimated using model 8M with temporal effect RW(1).

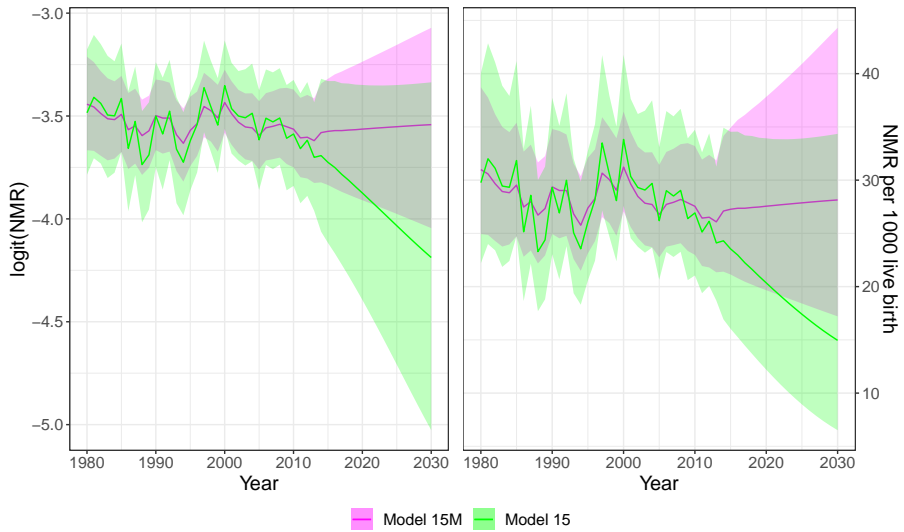


Figure 6.7: Estimated and predicted NMR and logit(NMR) in Kenya with 95% credible interval, estimated using a univariate time model (model 15) and a multivariate time model (model 15M), with temporal effect AR(2).

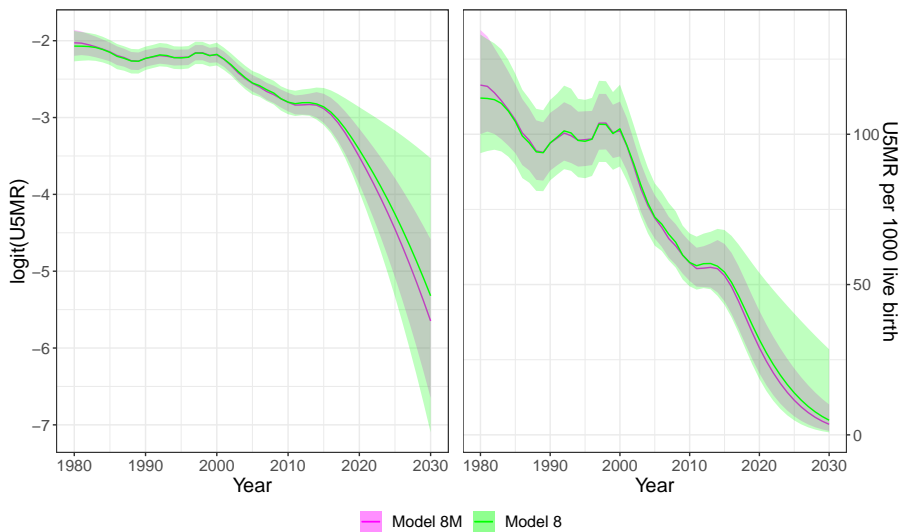


Figure 6.8: Estimated and predicted U5MR and logit(U5MR) in Kenya with 95% credible interval, estimated using a univariate time model (model 8) and a multivariate time model (model 8M), with temporal effect RW(1).

From Table 6.1 one can see that the median of the intercept, μ , in model 15 is 0.30, whereas the country-specific intercept for Kenya in model 15M, μ_1 , is much lower, -4.51 . The 95% credible interval for μ is wide, $[-4.51, 4.90]$, so both a significant negative and positive effect is possible. The credible interval for μ_1 is much narrower, and μ_1 is significantly different from zero. The estimated median and credible interval for $\exp(\xi)$ and $\exp(\xi_1)$ in Table 6.1 are similar. The credible intervals are relatively narrow and the mean and median values are $\exp(\xi) = 1.04$ and $\exp(\xi_1) = 0.99$. Both credible intervals include 1, so an increase in time can lead to a small decrease or a small increase in NMR. This is supported by previously shown plots of NMR (Figure 3.1), where NMR does not seem to follow a linear trend in time. The mean and median value of $\exp(\beta)$ in Table 6.1 is higher for model 15M than for model 15. For model 15M, increasing $\log(\text{GDP2})$ increases NMR, which can explain the increase in NMR in the predictions in Figure 6.7. The credible interval is tightest for model 15, but both credible intervals include 1. It is therefore difficult to determine the effect of $\log(\text{GDP2})$ on NMR.

Table 6.1 further shows that the standard deviation of the AR(2) process is smaller for model 15M than for model 15. The interval $[\exp(-2\sigma_a), \exp(2\sigma_a)]$ gives the approximate 95% credible interval for the time series' effect on the odds-ratio. For model 15 that interval is approximately $[0.79, 1.21]$ and for model 15M it is approximately $[0.85, 1.17]$. The AR coefficients for both model 15 and model 15M have large credible intervals, which include 0. The credible intervals are somewhat narrower for model 15 than for model 15M. The medians of the AR coefficients are positive, indicating a positive correlation in time. The credible intervals given for σ_s for both models in Table 6.1 do not contain zero and the credible interval for σ_s is tightest for model 15M. In Table 6.1 ρ_{group} has a median of 0.46, with 95% credible interval $[-0.20, 0.92]$. Thus the effect is not significantly different from zero. When comparing the 95% credible interval to that of the default prior given in Section 3.2 ($[-0.41, 0.41]$) one can see that the interval has shifted upwards, indicating that there is a positive correlation.

Table 6.2 shows that equivalent parameters in model 8 and model 8M are almost identical. The median of μ and μ_1 are close in value, but μ_1 have a tighter 95% credible interval. The median estimates for $\exp(\beta)$ are identical for the two models at 0.98, but once again model 8M has a tighter credible interval. The median of the standard deviation of the random walk and the survey effect are almost identical for the two models, but model 8M still have tighter credible intervals. $[\exp(-2\sigma_a), \exp(2\sigma_a)] = [0.90, 1.11]$ for both models. The correlation between countries presented in Table 6.2 has median $\rho_{\text{group}} = 0.42$, with 95% credible interval $[-0.04, 0.77]$. Thus the correlation is not significantly different from zero, but it is more skewed towards positive correlation than the default prior.

To properly assess the new models' capability of predicting mortality rates, average one-, two-, and three-step prediction CRPS scores and MSE are computed. As in Chapter 5, the predictions are computed by estimating models using data until and including year 2005, 2006, . . . , 2012 (for all three countries) and then computing one, two and three steps predictions for KDHS14. When calculating CRPS, it is assumed that both the estimated mortality rates and the predicted mortality rates follow a normal distribution, so Equation (2.19) is used. As in Chapter 5, SSP2 is used for POP, GDP1 and GDP2 for year > 2010 . The resulting average CRPS scores and MSE are presented in Table 6.3 and Table 6.5, for NRM and U5MR, respectively. For comparison, the scores for the corresponding

Parameter	0.025Q	Median	0.975Q
Model 15			
μ	-4.51	0.30	4.90
$\exp(\xi)$	0.99	1.04	1.09
$\exp(\beta)$	0.07	0.29	1.38
σ_a	0.07	0.12	0.20
ϕ_1	-0.42	0.27	0.56
ϕ_2	-0.38	0.14	0.37
σ_s	0.05	0.11	0.27
Model 15M			
μ_1	-6.38	-4.51	-2.65
$\exp(\xi_1)$	0.97	0.99	1.01
$\exp(\beta)$	0.77	1.39	2.52
σ_a	0.05	0.08	0.14
ϕ_1	-0.69	0.71	0.97
ϕ_2	-0.34	0.24	0.55
σ_s	0.04	0.07	0.15
ρ_{group}	-0.20	0.46	0.92

Table 6.1: Estimated 0.025 quantile, median and 0.975 quantile of parameters of the model 15 and 15M for NMR, estimated using all available data. μ is the overall intercept, μ_1 is the country specific intercept for Kenya. Similarly, ξ and ξ_1 are the estimated linear trends for Kenya. β is the coefficient of the covariate, $\log(\text{GDP2})$. σ_a is the standard deviation of the time series, while ϕ_1 and ϕ_2 are the estimated coefficients in the autoregressive process. σ_s is the standard deviation of the survey effect. ρ_{group} is the estimated correlation between temporal process in Kenya, Uganda and Tanzania.

Parameter	0.025Q	Median	0.975Q
Model 8			
μ	-2.18	-1.49	-0.80
$\exp(\beta)$	0.96	0.98	0.99
σ_a	0.03	0.05	0.09
σ_s	0.4	0.08	0.19
Model 8M			
μ_1	-1.87	-1.46	-1.08
$\exp(\beta)$	0.97	0.98	0.99
σ_a	0.03	0.05	0.07
σ_s	0.03	0.05	0.10
ρ_{group}	-0.04	0.42	0.77

Table 6.2: Estimated 0.025 quantile, median and 0.975 quantile of parameters of the model 8 and 8M for U5MR, estimated using all available data. μ is the overall intercept, μ_1 is the country specific intercept for Kenya. β is the coefficient of the covariate, GDP1 . σ_a is the standard deviation of the time series and σ_s is the standard deviation of the survey effect. ρ_{group} is the estimated correlation between temporal process in Kenya, Uganda and Tanzania.

Model	One-step		Two-step		Three-step	
	CRPS	MSE	CRPS	MSE	CRPS	MSE
12	2.24	1.31	2.37	1.40	2.83	1.74
12M	4.11	1.89	4.36	2.08	5.20	2.61
14	1.72	0.98	1.76	1.01	2.27	1.36
14M	4.17	1.93	4.39	2.12	5.19	2.62
15	1.55	0.88	1.49	0.84	1.98	1.17
15M	4.48	2.11	4.89	2.41	5.74	2.97

Table 6.3: One, two and three step prediction CRPS $\times 10^{-02}$ and MSE $\times 10^{-02}$ for NMR in Kenya, using models with temporal effect AR(2). The values marked in green are the most favourable in each column.

Survey	15	15M
KDHS03	0.11	0.09
KDHS08-09	0.04	0.04
KDHS14	-0.15	-0.14

Table 6.4: Survey effects for Kenya estimated by model 15 and 15M using data from 1980-2013.

univariate models (presented in Chapter 5) are also included in the tables.

From Table 6.3, one can see that the average CRPS scores and the MSE are larger for the multivariate models than for the univariate models. To further investigate the reason behind this, the three-step predicted NMR with credible interval for 2008-2013 is presented in Figure 6.9. Figure 6.9 shows that the credible interval of model 15M does not appear to be tighter than that of model 15, but it has shifted a bit upwards. The mean predictions have also shifted upwards and are thus further away from the KDHS14 estimates. To check if this is due to the estimated survey effects, the survey effects of model 15 and 15M estimated using data from 1980-2013 are presented in Table 6.4. As seen in Table 6.4, the differences between the estimated survey effects are small. Furthermore, there are no apparent systematic differences between the survey effects in Kenya, Uganda, and Tanzania.

Model	One-step		Two-step		Three-step	
	CRPS	MSE	CRPS	MSE	CRPS	MSE
12	2.66	0.98	2.72	1.09	4.05	1.82
12M	3.51	1.14	3.31	1.21	3.36	1.35
8	1.79	0.64	1.28	0.48	2.48	1.08
8M	1.93	0.67	1.46	0.55	1.65	0.69
9	2.48	0.95	2.07	0.85	3.21	1.45
9M	3.78	1.43	3.85	1.66	4.02	1.78

Table 6.5: One-, two-, and three-step prediction CRPS $\times 10^{-02}$ and MSE $\times 10^{-02}$ for U5MR in Kenya, using models with temporal effect RW(1). The values marked in green are the most favourable in each column.

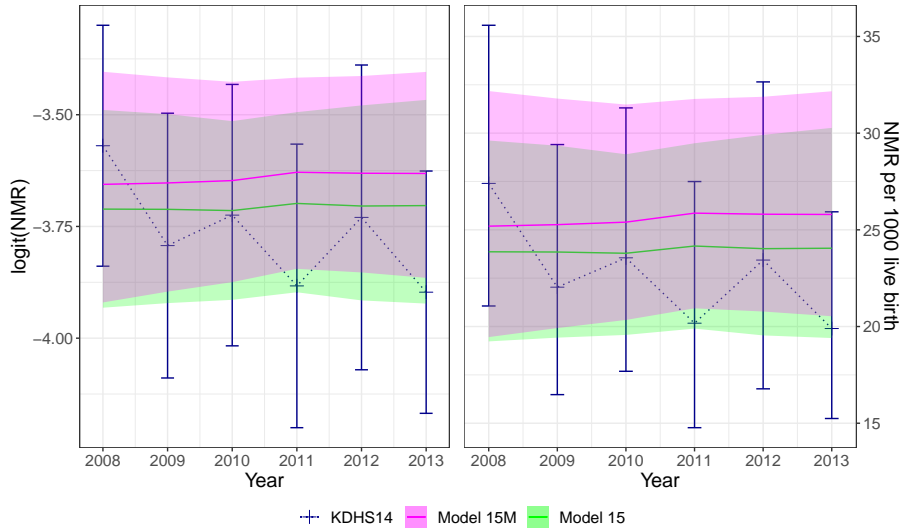


Figure 6.9: Three-steps predictions for NMR for 2008-2013 produced by model 15 and 15M, with 95% credible intervals, and KDHS14 NMR estimates.

Rate	Model 12M	Model 14M/8M	Model 15M/9M
NMR	0.05%	0.07%	0.05%
U5MR	9.13%	99.97%	99.96%

Table 6.6: Probability of reaching SDG 3 for Kenya.

Table 6.5 shows that for U5MR, the multivariate models 12M and 9M have lower average three step prediction MSE and CRPS scores than their univariate counterparts, while the opposite is true for one- and two-step predictions and for model 9M. To further assess the three-step predictions, the three-step predictions and 95% credible intervals produced using model 8 and 8M are plotted with the KDHS14 estimates in Figure 6.10. From Figure 6.10 one can see that model 8 and model 8M have predicted mean closest to the estimated mean three times each, and that the credible interval of model 8M is better at capturing the uncertainty in the KDGS14 estimates.

Finally it is interesting to closer study the estimates/projections for 1990, 2015 and 2030 and investigate the progress towards the SDGs and MDGs, using the multivariate time models. As in Section 5.2, progress towards the SDGs is assessed by calculating the probability of reaching SDG target 3.2, whereas progress towards MDG target 4A is assessed by studying median predicted U5MR. The probabilities of reaching SDG target 3.2 is presented in Table 6.6, and the final estimates/predictions with credible intervals for 1990, 2015 and 2030 are presented in Table 6.7 and Table 6.8.

From Table 6.6 one gets that the probability of reaching SDG target 3.2 using model 12M is 0.05% for NMR and 9.13% for U5MR. That is a small reduction compared to the results for model 12 (the univariate model) presented in Table 5.8. The probability

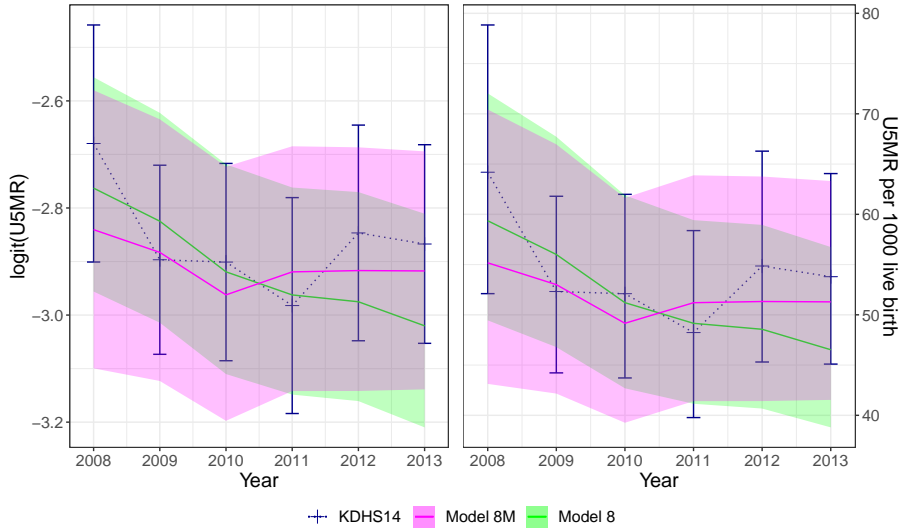


Figure 6.10: Three-steps predictions for U5MR for 2008-2013 produced by model 8 and 8M, with 95% credible intervals, and KDHS14 U5MR estimates.

Model 12M (liner trend, no covariates)			
Scenario	1990	2015	2030
SSP0	29.08 [24.83, 34.40]	26.76 [20.92, 33.90]	25.61 [17.08, 36.26]

Model 14M (t , $\log(\text{GDP1})$)			
Scenario	1990	2015	2030
SSP1	29.49 [25.12, 34.91]	26.26 [20.49, 33.28]	31.10 [18.05, 50.78]
SSP2	29.51 [25.11, 34.98]	26.35 [20.55, 33.42]	30.16 [18.15, 47.51]
SSP3	29.50 [25.12, 34.94]	26.22 [20.43, 33.26]	28.07 [18.01, 41.35]
SSP4	29.50 [25.12, 34.94]	26.19 [20.40, 33.24]	28.23 [18.04, 41.78]
SSP5	29.49 [25.12, 34.91]	26.26 [20.49, 33.28]	31.72 [17.96, 53.07]

Model 15M(t , $\log(\text{GDP2})$)			
Scenario	1990	2015	2030
SSP1	29.27 [24.92, 34.70]	27.14 [21.03, 34.69]	28.65 [16.68, 46.97]
SSP2	29.27 [24.92, 34.70]	27.17 [21.04, 34.74]	28.07 [16.97, 44.15]
SSP3	29.27 [24.92, 34.70]	27.23 [21.06, 34.87]	27.72 [17.08, 42.68]
SSP4	29.28 [24.92, 34.70]	27.22 [21.06, 34.85]	27.51 [17.13, 41.82]
SSP5	29.27 [24.92, 34.70]	27.13 [21.02, 34.68]	29.08 [16.45, 49.10]

Table 6.7: Median estimated and forecasted NMR in Kenya in 1990, 2015 and 2030, with 95% credible intervals.

Model 12M (linear trend, no covariates)			
Scenario	1990	2015	2030
SSP0	98.66 [89.32, 108.74]	51.21 [42.81, 61.94]	35.48 [20.94, 59.85]

Model 8M (GDP1)			
Scenario	1990	2015	2030
SSP1	97.23 [88.00, 107.34]	53.87 [45.09, 64.71]	2.60 [0.86, 8.74]
SSP2	97.07 [87.90, 107.12]	52.92 [44.41, 63.49]	3.47 [1.30, 10.09]
SSP3	97.02 [87.86, 107.06]	53.55 [44.95, 64.24]	7.69 [3.61, 17.20]
SSP4	97.01 [87.85, 107.05]	53.76 [45.12, 64.48]	7.19 [3.31, 16.42]
SSP5	97.23 [88.00, 107.34]	53.82 [45.05, 64.66]	1.94 [0.59, 7.24]

Model 9M (GDP2)			
Scenario	1990	2015	2030
SSP1	97.17 [87.80, 107.47]	44.19 [36.27, 54.67]	3.01 [0.98, 11.07]
SSP2	97.20 [87.83, 107.51]	44.13 [36.21, 54.61]	4.89 [1.85, 14.98]
SSP3	97.23 [87.85, 107.55]	43.65 [35.74, 54.15]	6.39 [2.61, 17.83]
SSP4	97.23 [87.85, 107.55]	43.76 [35.85, 54.25]	7.34 [3.13, 19.36]
SSP5	97.16 [87.79, 107.47]	44.23 [36.31, 54.71]	2.15 [0.63, 9.06]

Table 6.8: Median estimated and forecasted U5MR in Kenya in 1990, 2015 and 2030, with 95% credible intervals.

of reaching SDG target 3.2 for NMR is not that sensitive to the choice of GDP model, with 0.07% and 0.05% for model 14M and 15M, respectively. The probability of reaching target 3.2 using model 14 and 15 is much higher, almost 30%. The probability of reaching target 3.2 for U5MR using model 8M and 9M is higher than when using model 8 and 9. However, the change in probability is not as large as the change for NMR. Model 8M gives a probability of reaching the target of 99.97%, whereas model 9M gives a probability of 99.96%. Thus, if the model 8M and 9M are accurate, Kenya will most likely reach SDG target 3.2 for U5MR but not for NMR.

When comparing the NMR values for model 12 in Table 5.6 to the values for model 12M in Table 6.7, one can detect a small decrease. As also seen from Figure 6.7, the credible intervals are in general tighter for model 12M in Table 6.7 than for model 12 in Table 5.6. The median predictions for 2030 using model 14M and 15M have increased compared to the ones from model 14 and 15. An increasing trend in NMR can also be seen in Figure D.1 and Figure D.2. If model 14M and 15M are accurate, this is an indication that interventions should be implemented in Kenya.

In general the median values for U5MR in 2015 and 2030 presented in Table 6.8 are lower than the corresponding values in Table 5.7 and the credible intervals are tighter. Using the median values presented in Table 6.8 one gets that MDG target 4A becomes 33 per 1000 live birth in 2015 for U5MR using model 12M and 32 using model 8M and 9M (rounded to the nearest child). Thus the conclusion is once again that MDG target 4A was not reached for Kenya.

Judging by three-step prediction CRPS and MSE it appears that the best model for

NMR is model 15, while the best model for U5MR is model 8M. Using these models, the probability of reaching SDG target 3.2 is 29.63% for NMR and 99.97% for U5MR. However, the probability of reaching SDG target 3.2 is sensitive to the choice of model and somewhat sensitive to the choice of GDP projections. MDG target 4A was not reached for Kenya.

The analysis in this section concludes that NMR predictions for Kenya are not improved by including information about Tanzania and Uganda, whereas U5MR predictions are. The probability of reaching SDG target 3.2 for NMR is low but approximately equal for the three investigated multivariate models. The probability of reaching SDG target 3.2 for U5MR is approximately 10% for the multivariate model without covariates, and approximately 100% for the multivariate models with *GDP* as a covariate. By comparing this to the probabilities obtained in Chapter 5, it is clear that the probability of reaching SDG target 3.2 is sensitive to the choice of model.

6.2 Effect of borrowing strength from neighboring countries on estimation

Besides being interested in producing accurate forecasts, it is also desirable to produce accurate estimates of the past and present. DHS surveys are normally undertaken every fifth year, so imagine the following scenario: KDHS14 is not published yet, but information from UDHS 16 and TDHS16 is available up to and including the year 2013. Let the focus be on obtaining accurate estimates of NMR and U5MR in Kenya in the years 2009-2013. It is desirable to determine if joint modeling of the mortality rates in Kenya, Uganda, and Tanzania leads to more accurate estimates for Kenya than a model that only considers data from Kenya.

Model 12, 12M, 14, 14M, 15 and 15M are fitted for NMR, without using data from the KDHS14 survey, for the period 1980-2013. For U5MR, model 12, 12M, 8, 8M, 9 and 9M are fitted in the same way. The resulting predictions for the KDHS14 survey plotted together with KDHS data are presented in Appendix E. From the plots, one can see that in general, the multivariate models have tighter credible intervals than previously seen plots, and it looks like the predicted values for the multivariate models are closer to the expected values estimated using data from the KDHS14. This is more apparent for the models including covariates than for model 12 and 12M. To further investigate this observation, the average CRPS (calculated using Equation (2.19)) and MSE for KDHS14 estimates for 2009-2013 are computed, and presented in Table 6.9 and Table 6.10. When calculating CRPS and MSE, estimates for SSP2 are used for 2011-2013, for the models that include GDP.

From Table 6.9 one can see that model 12M is not an improvement compared to model 12, but model 14M and 15M have lower average CRPS and MSE than model 14 and model 15. Model 14M and model 15M performs very similarly. That is not that surprising as $\log(\text{GDP1})$ and $\log(\text{GDP2})$ are identical up to and including the year 2010 for all three countries, so only three years with different GDP values are considered. Despite this, model 14 and 15 perform quite differently, with model 15 still being the best out of these two.

Model	CRPS $\times 10^{-01}$	MSE $\times 10^{-01}$
12	1.80	14.61
14	1.47	11.70
15	1.08	8.08
12M	2.92	20.05
14M	0.28	0.90
15M	0.28	0.91

Table 6.9: Average CRPS and MSE for KDHS14 NMR estimates, 2009-2013. The values marked in green are the most favourable in each column.

Model	CRPS $\times 10^{-01}$	MSE $\times 10^{-01}$
12	2.54	17.67
8	1.97	12.78
9	1.71	10.82
12M	2.14	11.61
8M	0.15	0.30
9M	0.31	0.77

Table 6.10: Average CRPS and MSE for KDHS14 U5MR estimates, 2009-2013. The values marked in green are the most favourable in each column.

From Table 6.10, one can see that all multivariate models are an improvement on their corresponding univariate models. There is a small difference in the performance of model 8M and 9M, with model 8M still being the preferred one. It is interesting to note that model 9 have lower CRPS score and MSE score than model 8. That is unexpected since model 8 performs better than model 8 in Chapter 5.

From this analysis, one can learn that historical and current estimates of mortality rate in a country can be improved by borrowing strength from similar countries for which newer data is available. This is a significant result, as it shows that joint modeling of mortality rates produce the most reliable estimates for a country for which recent data is unavailable.

Discussion

The primary sources of information when estimating child mortality rates in low-and medium-income countries are censuses and surveys. The DHS surveys used in this thesis are of high quality, the DHS data is simple to access, and the coding is explained in the DHS recode manuals. However, exploring the different datasets reveals that they are not coded in the same way even though the recode manuals indicate that they are. This indicates issues with data accuracy and the need to check the coding of the designs. The DHS reports are structured in the same way for different surveys, but information about the sample design is given differently in the different reports. This means that there are issues of the quality of the data provided.

The unaccounted for non-sampling error in the data affects all findings of this thesis. The non-sampling error is caused by, for example, recollection bias, migration or epidemic of diseases such as HIV/AIDS. Data from UNAIDS (2018) shows that the number of deaths related to AIDS peaked in Kenya around 1998-2006. Women who died in that period were not able to report births and deaths and are more likely to have had children who died young (p. 5, Wakefield et al. (2018)). Thus, estimates of NMR and U5MR for an extended period of time are inaccurate. Section 2.2 of Wakefield et al. (2018) describes a method of adjusting for HIV, which could have been interesting to apply in this thesis. Despite the non-sampling errors and the challenges discussed in the previous paragraph, the DHS data is of high quality, and there is no other data source that is comparable with respect to quality, content, and accessibility.

In this thesis, the designs of the DHS surveys are accounted for by splitting the modeling process into two parts; first, design-based yearly mortality rates are estimated, then a temporal model is applied. If one were to use the observation units directly in a temporal model for mortality rates without design-based estimation, one would have to develop a model that is valid for every location. Then the model would have to account for urban/rural effects, cluster effects, stratum effects, and so on. That would make the modeling process much more complicated, require detailed urban/rural maps, and spatial sparsity of data could lead to problems with parameter estimation. In addition, unknown information about birth density is needed to produce estimates at county or national level. The two-step

approach taken in this thesis, on the other hand, directly targets national mortality rates.

The fact that the true national mortality rate is not known for any year makes it difficult to assess the accuracy and precision of the models. The CRPS score balances accuracy and precision. Therefore CRPS is considered most important in this thesis. If two models are equally inaccurate, the model that is best at capturing the uncertainty in the predictions will have the lowest CRPS score. However, predictions with too high uncertainty are of little use. Assessments of point predictions are arguably less important, as one cannot know if the observation or the prediction is at that point. Ideally, one should assess predictions using techniques such as cross-validation, but that is challenging in this situation.

To perform cross-validation, one could randomly select some clusters to leave out of the modeling process and then try to predict the child mortality rates in those clusters. However, the developed method estimates child mortality rates at a national scale, not for specific enumeration areas. Therefore the comparison would not be useful for evaluating the models' performance at the national level. Computing design-based mortality estimates using all observations, then performing cross-validation by randomly selecting some years to be a part of the training set is not possible as the observations are time-dependent. Only observations from the final years can be left out when estimating models, as is done when computing step-wise predictions.

As mentioned in Chapter 5, 17-step ahead prediction CRPS scores should have been considered not only in Chapter 4 but also in Chapter 5 and in Chapter 6. However, 17-step ahead predictions bring with it two problems. First one cannot know if the trend from 17 years ago will continue. From Figure 3.3, one can see that U5MR behaves differently between 1980-1996 than from 1996-2013. Therefore models that could have predicted 1996-2013 accurately would not necessarily predict 2013-2030 equally well. Second, when aiming to evaluate 17-step predictions, very few years of estimated mortality rates are left for model-estimation. If one were to train models on data for years ≤ 1996 , only one 17-step prediction CRPS score could have been computed. In Chapter 5 and Chapter 6, six three-step ahead CRPS scores are computed and the average taken. If one were to do the same for 17-step prediction scores, only data for years prior to 1990 would have been considered for the first model.

Throughout the thesis, the choices of best models have been sensitive to the choices of assessment criteria and different models have performed relatively similar. Three-step ahead prediction CRPS scores are considered most important in this thesis, as a compromise between the need for long-term predictions and the problems discussed in the previous paragraph. However, models with the lowest three-step ahead CRPS score have not always had the lowest one- and two-step CRPS scores or the lowest MSE scores. If a different assessment criterion had been deemed most important, different models would have been considered the best models.

The predictions produced in this thesis answers the question "what will happen if the current trend continues?" That is an important question as it identifies where one has to implement interventions in order to reach the UN goals. However, it is not possible to know what will change in the future and directly impact mortality rates. If for example, the weather changes and that affects mortality rates, a need for human interventions that were not identified by the models could arise. It is possible to include covariates that explain changes in trend, but the covariates also have to be forecasted.

The covariates used in this thesis are bound to have associated uncertainty, which was not accounted for in the models. Instead, the uncertainty in the covariates is, to some extent, accounted for by the five different scenarios. The estimated coefficients of the covariates have large credible intervals, indicating that the effect of the covariates is difficult to determine. The uncertainty is amplified for predictions if the covariates have high values so that they have a big impact or if confounded effects interact differently in future years.

This thesis aimed not only to produce accurate forecasts but to do so in a transparent, easily reproducible manner. Computing design-based estimates of yearly NMR using the `survey` package is straight forward, and `SUMMER` is helpful when computing yearly design-based estimates of U5MR. The function `inla()` was efficient and easy to use. All tested models were estimated in a matter of seconds, including the multivariate models with covariates, making the analysis conducted in this thesis feasible. Only one argument of `inla()` had to be modified to change between temporal processes and, in addition to adjusting the input, only minor changes had to be made to estimate multivariate models. Thus the modeling process is transparent, and it is possible for non-experts to produce forecasts.

The models developed in this thesis are relatively simple, and the default priors in `R-INLA` are used. It would have been interesting to examine the effect of the choice of the family of priors and choice of hyperparameters in the priors. However, a proper analysis of the effect of different priors and hyperparameters would require inspection of a vast amount of models. It would not be feasible to conduct given the time and computational resources available for this thesis. If an extensive investigation of priors were to be undertaken, it would have been interesting to test the priors developed in Fuglstad et al. (2019). Fuglstad et al. (2019) models the random effect through a total latent variance and attribution of total variance to the model components.

The models would be more realistic if $\text{NMR} \leq \text{U5MR}$ was ensured. One way to do that is to jointly estimate separate temporal models for each of the six age intervals and make predictions for each age interval prior to computing the U5MR, similar to what is done in Wakefield et al. (2018). Then the NMR would be an inherent part of the U5MR and $\text{NMR} \leq \text{U5MR}$ also for predictions. Furthermore, it would be interesting to expand the model so that it could produce estimates and forecasts at sub-national levels. Then model-based methods are needed to estimate NMR and U5MR in areas where there is too little data for design-based estimation.

This thesis shows that Latent Gaussian models can be used to forecast NMR and U5MR, and do so most successfully with `AR(2)` and `RW(1)` as the temporal effect, for NMR and U5MR respectively. Applying the developed models to data from Kenya reveals that the probability of Kenya reaching SDG target 3.2 is sensitive to the choice of model. The best models give a probability of reaching SDG target 3.2 of 29.63% for NMR and 99.97% for U5MR, given that the current development continues. Based on this, interventions attempting to reduce NMR should be implemented in Kenya. The best models for U5MR developed in this thesis all conclude that Kenya did not reach MDG target 4A, consistent with the conclusion one can draw from the official estimates presented in Table 1.1. However, those conclusions are based on a consideration of mean predictions only, even though the joint posterior marginal of the linear predictor for 1990 and linear predictor for

2015 should have been considered.

This thesis finds that estimates for one country can be improved by borrowing strength from neighboring countries for which newer data is available. Given the assumption that historical data is used, the result does not depend on forecasted covariates. Therefore, they are deemed more reliable than the prediction results. However, the effect of different past interventions in different countries is not studied. If different interventions have been implemented in the neighboring countries compared to the country of interest, the conclusion might not hold. Despite this, the findings are of importance. The DHS should continue undertaking surveys in neighboring countries in different years, as that allows for maximal exploitation of the data. The results provide arguments for the current practice of simultaneously estimating child mortality rates in several countries.

This thesis demonstrates that the two-step approach for modeling child mortality rates is efficient and valuable. Temporal models can be used to predict mortality rates for 2030 and to assess the probability of reaching SDG target 3.2. Identifying a trend in NMR is difficult. Predictions for U5MR are improved by borrowing strength from neighboring countries. Estimates of past values of NMR and U5MR are improved by including information from neighboring countries for which newer information is available. Based on the findings of the thesis, models with temporal effect RW(1) and AR(2) best forecasts child mortality rates and forecasts can be improved by including covariates and borrowing strength from nearby countries. Therefore, it is recommended that *SUMMER* includes AR(2) as a possible temporal effect and makes it possible to include covariates and borrow strength from neighboring countries.

Bibliography

- Alkema, L., New, J. R., 2014. Global estimation of child mortality using a bayesian b-spline bias-reduction model. *The Annals of Applied Statistics* 8 (4), 2122–2149.
- Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Brit, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., Lynch, M., Lindgren, F., Cohen, J. M., Murray, C. L. J., Smith, D. L., Hay, S. I., Cibulskis, R. E., Gething, P. W., 2015. The effect of malaria control on plasmodium falciparum in africa between 2000 and 2015. *Nature* 8 (526), 207–211.
- Brass, W., Coale, A. J., 1968. *Methods of Analysis and Estimation*.
- Brockwell, P. J., Davis, R. A., 2016. *Introduction to Time Series and Forecasting*.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76 (1).
- Carruthers, E., Lewis, K., McCue, T., Westley, P., 2008. *Generalized linear models: model selection, diagnostics, and overdispersion*. Memorial University of Newfoundland.
- Croft, T. N., Marshall, A. M. J., Allen, C. K., 2018. Guide to dhs statistics. https://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_DHS-7.pdf, [Accessed: 28.01.19].
- Dehaene, G., 2018. *Lecture notes, math-435, bayesian computation*, epfl.
- DHS, 2008. *Standard recode manual for dhs-4*. <https://www.dhsprogram.com/pubs/pdf/DHSG4/Recode4DHS.pdf>, [Accessed: 28.01.19].
- DHS, 2012. *Standard recode manual for dhs-5*. https://www.dhsprogram.com/pubs/pdf/DHSG4/Recode5DHS_23August2012.pdf, [Accessed: 28.01.19].
- DHS, 2013. *Standard recode manual for dhs-6*. https://www.dhsprogram.com/pubs/pdf/DHSG4/Recode6_DHS_22March2013_DHSG4.pdf, [Accessed: 28.01.19].

-
- Efron, B., 1988. Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of the American Statistical Association* 83 (402), 414–425.
- Foremann, K. J., Maraquez, N., Doglert, A., Fukutaki, K., Fullman, N., McGaughey, M., Pletcher, M. A., Smith, A. E., Tang, K., Yuan, C.-W., Brown, J. C., Friedman, J., He, J., Heuton, K. R., Holmberg, M., Patel, D. J., Reidy, P., Carter, A., Cercy, K., Chaplin, A., Douwes-Schultx, D., Frank, T., Goettsch, F., Liu, P. Y., 2018. Forecasting life expectancy, years of life lost, and all-cause and specific-cause mortality fr 250 causes of death: reference and alternative scenarios for 2016-2040 for 195 countries and territories. *Lancet* 392, 2052–2090.
- Fuglstad, G.-A., Hem, I. G., Knight, A., Rue, H., Riebler, A., 2019. Intuitive principle-based priors for attributing variance in additive model structures. arXiv:1902.00242v1 [stat.ME].
- Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., van Vuuren, D. P., van den Berg, M., Feng, L., Klein, D., Calvin, K., Doelman, J. C., Frank, S., Fricko, O., Harmsen, M., Hasegawa, T., Havlik, P., Hilaire, J., Hoesly, R., Horing, J., Popp, A., Stehfest, E., Takahashi, K., nov 2018. Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development Discussions*, 1–42.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B., 2014. *Bayesian Data Analysis*. CC Press, Taylor and Francis Group.
- Gneiting, T., Raftery, A. E., 2007. Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association* 102 (477), 359–378.
- Golding, N., Burstein, R., Longbottom, J., Browne, A. J., Fullman, N., Osgood-Zimmerman, A., Earl, L., Bhatt, S., Cameron, E., Casey, D. C., Dwyer-Lindgren, L., Farag, T. H., Flaxman, A. D., Fraser, M. S., Gething, P. W., Gibson, H. S., Graetz, N., Krause, L. K., Kulikoff, X. R., Lim, S. S., Mappin, B., Morozoff, C., Reiner, R. C. J., Sligar, A., Smith, D. L., Wang, H., , Weiss, D. J., Murray, C. J. L., Moyes, C. L., Hay, S. I., 2017. Mapping under-5 and neonatal mortality in africa, 2000-2015: a baseline analysis for the sustainable development goals. *The Lancet* 390 (10108), 2171–2182.
- Groves, R. M., Fowler Jr., F. J., Couper, M., Lepkowski, J. M., Singer, E., Tourangeau, R., 2004. *Survey Methodology*. John Wiley & Sons.
- Halonon, J. I., Blangiardo, M., Toledano, Mireille B. Fecht, D., Gulliver, J., Anderson, H. R., Beevers, S. D., Dajnak, D., Kelly, F. J., Tonne, C., 2016. Long-term exposure to traffic pollution and hospital admissions in london. *Environmental Pollution* 208, 48–57.
- Hoffman, M. D., Gelman, A., 2014. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research* 15 (1), 1593–1623.
- Hubin, A., Storvik, G., 2016. Estimating the marginal likelihood with integrated nested laplace approximation (inla). arXiv:1611.01450v1 [stat.CO].

-
- IGME, 2017. Most recent child mortality estimates, 2017. https://childmortality.org/files_v22/download/UN%20IGME%20Child%20Mortality%20Report%202018.pdf, [Accessed:30.04.19].
- Kenya National Bureau of Statistics, Ministry of Health and National AIDS Control Council, Kenya Medical Research Institute, National Council for Population and Development, The DHS Program, ICF International, 2015. Kenya demographic and health survey. <https://dhsprogram.com/pubs/pdf/fr308/fr308.pdf>, [Accessed: 31.01.19].
- Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with r-inla. *Journal of Statistical Software* 63 (19), 1–25.
- Lohr, S. L., 2010. *Sampling: Design and Analysis*. Brooks/Cole.
- Lumley, T., 2010. *Complex Surveys: A Guide to Analysis Using R*. John Wiley & Sons.
- Lumley, T., Scott, A., et al., 2017. Fitting regression models to survey data. *Statistical Science* 32 (2), 265–278.
- Martins, T. G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with inla: new features. *Computational Statistics Data Analysis* 67, 68–83.
- MEASURE, 2019. Measure evaluation. https://www.measureevaluation.org/prh/rh_indicators/womens-health/nb/neonatal-mortality-rate-nmr, [Accessed: 28.01.19].
- MEASURE DHS, ICF International, 2012. Sampling and household listing manual. https://dhsprogram.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf, [Accessed: 21.02.19].
- Mercer, L., Wakefield, J., Pantazis, A., Lutambi, A., Masanja, H., Clark, S., 2015. Space-time smoothing of complex survey data: Small area estimation for child mortality. *The Annals of Applied Statistics* 9 (2), 1889–1905.
- Nosedal, A., 2016. The autocorrelation function and ar(1), ar(2) models. <https://mcs.utm.utoronto.ca/~nosedal/sta457/ar1-and-ar2.pdf>, [Accessed: 13.02.19].
- O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., Solecki, W., 2017. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change* 42, 169–180.
- Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpender, F., Silva, L. A. D., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J. C., Kainuma, M., Klimont,

-
- Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., jan 2017. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42, 153–168.
- Rogelj, J., Popp, A., Calvin, K. V., Luderer, G., Emmerling, J., Gernaat, D., Fujimori, S., Strefler, J., Hasegawa, T., Marangoni, G., Krey, V., Kriegler, E., Riahi, K., van Vuuren, D. P., Doelman, J., Drouet, L., Edmonds, J., Fricko, O., Harmsen, M., Havlík, P., Humpender, F., Stehfest, E., Tavoni, M., mar 2018. Scenarios towards limiting global mean temperature increase below 1.5 °c. *Nature Climate Change* 8 (4), 325–332.
- Rue, H., Held, L., 2005. *Gaussian Markov Random Fields*.
- Rue, H., Martino, S., 2019. Bayesian analysis of structured additive models (r documentation).
- Rue, H., Martino, S., Chopin, N., 2009. Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion). *Royal Statistics Society, Series B* 71 (2), 319–392.
- Rue, H., Riebler, A., Sørbye, S., Illian, J. B., Simpson, D. P., 2016. Bayesian computing with inla: A review. *Annual Review of Statistics and Its Application* (4), 395–421.
- Saghagen, M., Vik, H., 2018. A beginners guide to complex survey methodology. (Project assignment at NTNU, TMA4500).
- Sørbye, S. H., Rue, H., 2017. Penalised complexity priors for stationary autoregressive processes. *Journal of time series analysis* 38, 923–935.
- UN, 2019a. Goal 3: Ensure healthy lives and promote well-being for all at all ages. <https://www.un.org/sustainabledevelopment/health/>, [Accessed: 28.01.19].
- UN, 2019b. Goal 4: Reduce child mortality. <http://www.un.org/millenniumgoals/childhealth.shtml>, [Accessed: 31.01.19].
- UNAIDS, 2018. Kenya hiv and aids estimates. <http://www.unaids.org/en/regionscountries/countries/kenya>, [Accessed: 30.04.19].
- Vehtari, A., Gelman, Andrew Gabry, J., 2017. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing* 27 (5), 1413–1432.
- Wakefield, J., Fuglstad, G.-A., Riebler, A., Godwin, J., Wilson, K., Clark, S. J., 2018. Estimating under-five mortality in space and time in a developing world context. *Statistical methods in medical research*, 1–21.
- Wang, H., Liddell, C. A., Coates, M. M., Mooney, M. D., Llevitz, C. E., Schumacher, A. E., Apfel, H., Iannarone, M., Phillips, B., Lofgren, K. T., et al, 2014. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: A systematic analysis for the global burden of disease study 2013. *The Lancet* 384, 957–979.

Watanabe, S., 2010. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11.

Zivot, E., Wang, J., 2006. *Modeling Financial Time Series with S-PLU®*. Springer, New York, NY.

Additional Simulation Results

In this section, additional plots referred to in Chapter 4 is displayed.

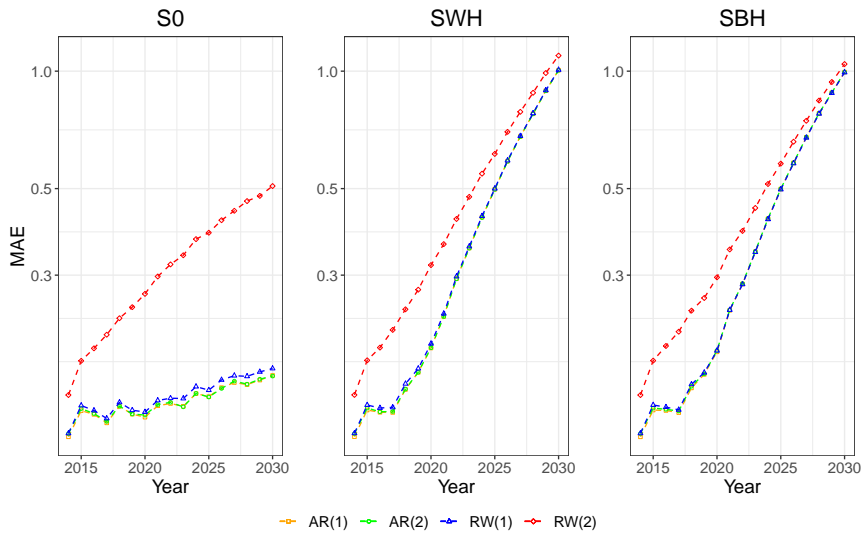


Figure A.1: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(NMR) simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

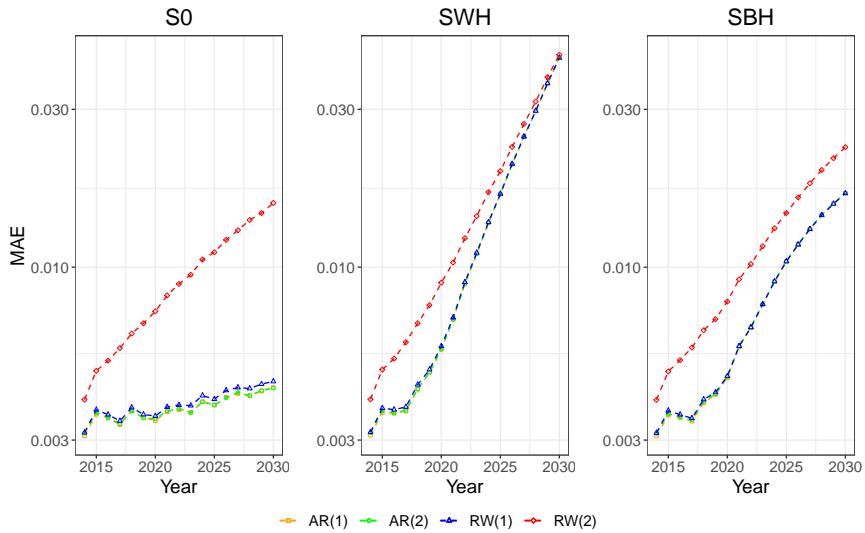


Figure A.2: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of NMR pr. 1000 live birth simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

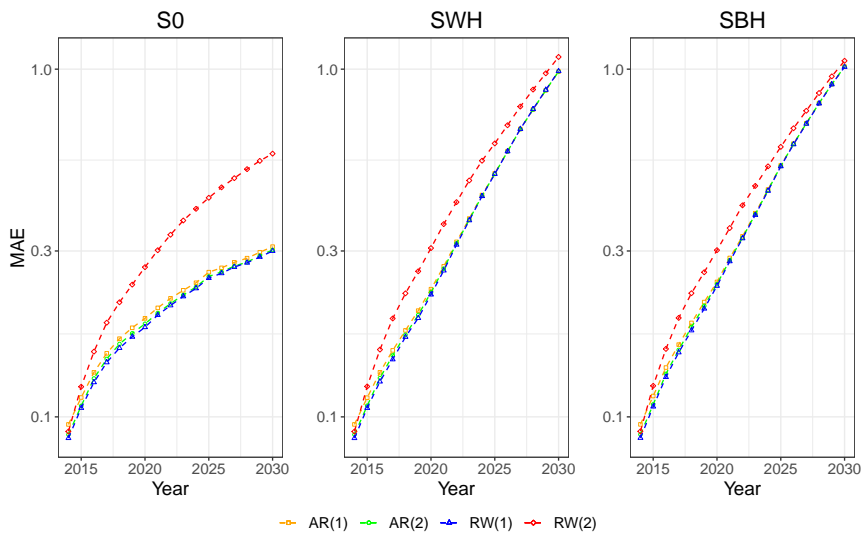


Figure A.3: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(U5MR) simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

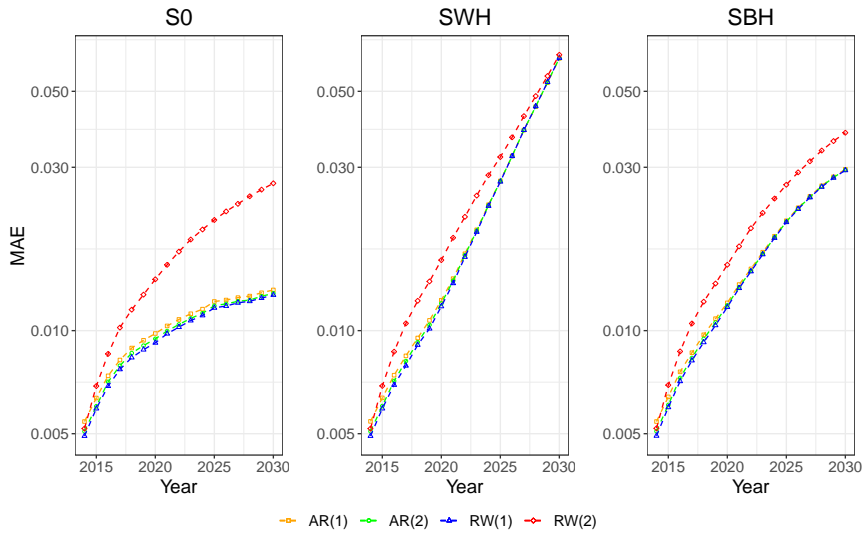


Figure A.4: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of U5MR pr. 1000 live birth simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

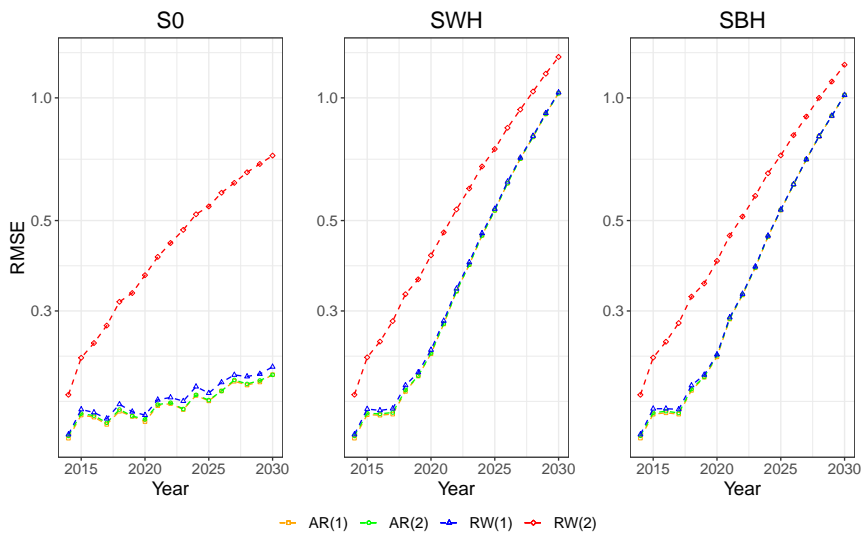


Figure A.5: RMSE of prediction error for models with with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(NMR) simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

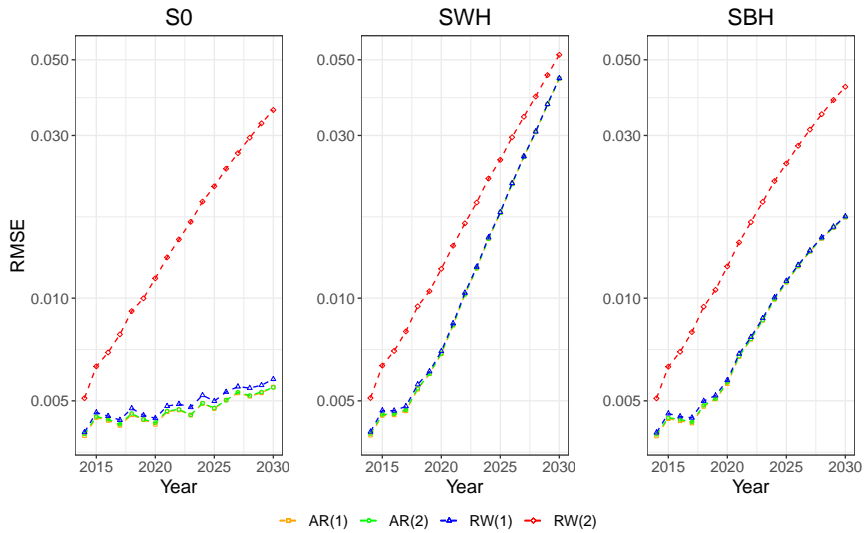


Figure A.6: RMSE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of NMR pr. 1000 live birth simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

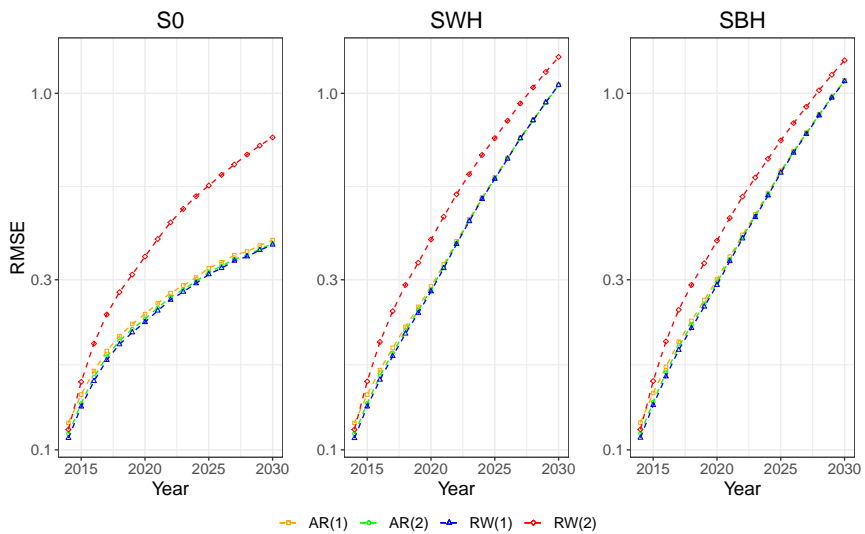


Figure A.7: RMSE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(U5MR) simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

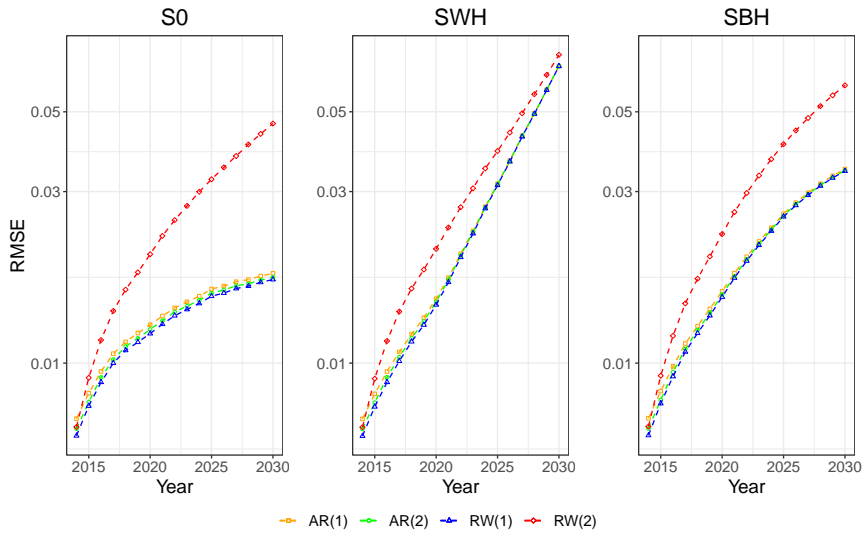


Figure A.8: RMSE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of U5MR pr. 1000 live birth simulated by a model with temporal effect AR(1), for scenarios S0, SWB and SBH.

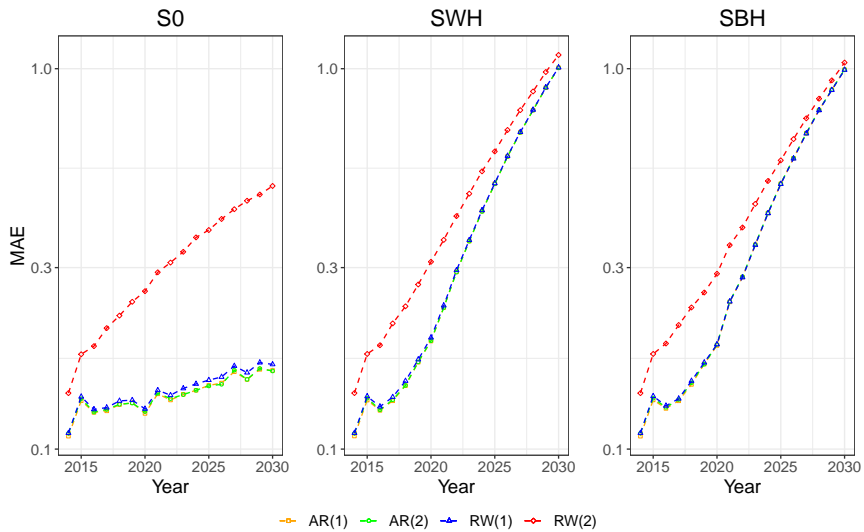


Figure A.9: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(NMR) simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.

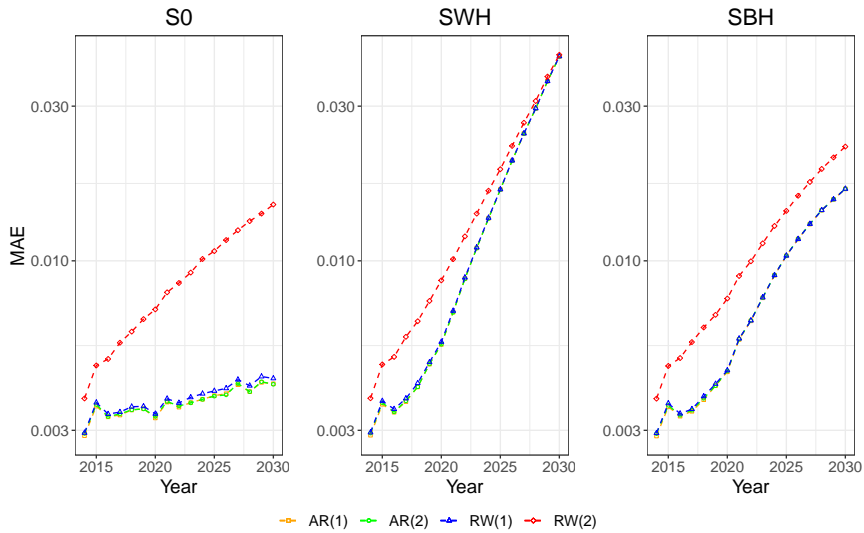


Figure A.10: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of NMR pr. 1000 live birth simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.

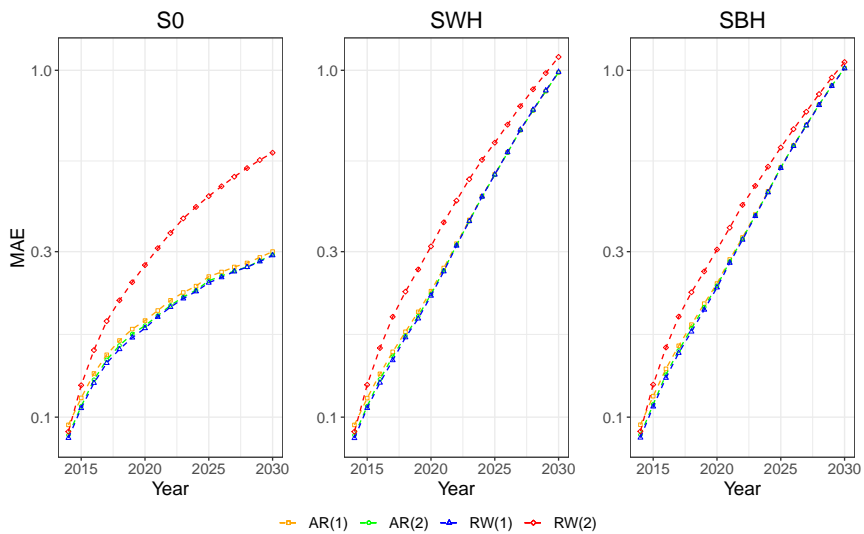


Figure A.11: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(U5MR) simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.

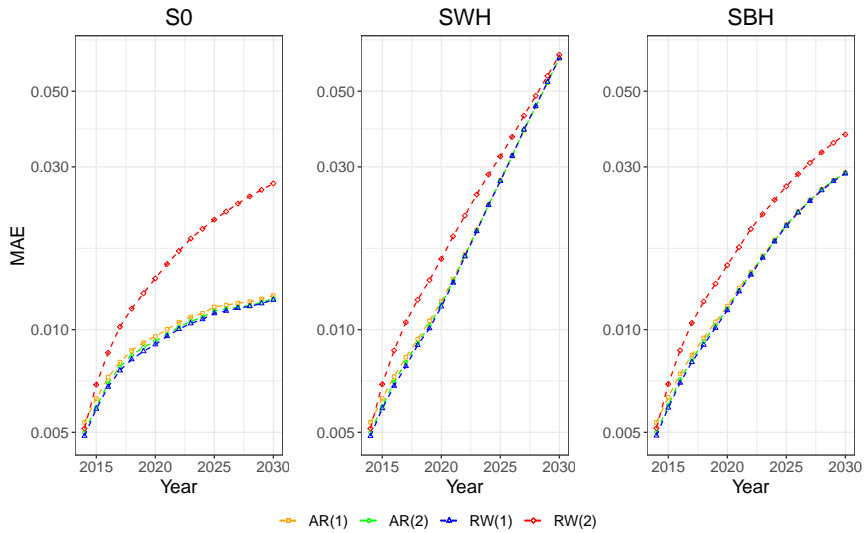


Figure A.12: MAE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of U5MR pr. 1000 live birth simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.

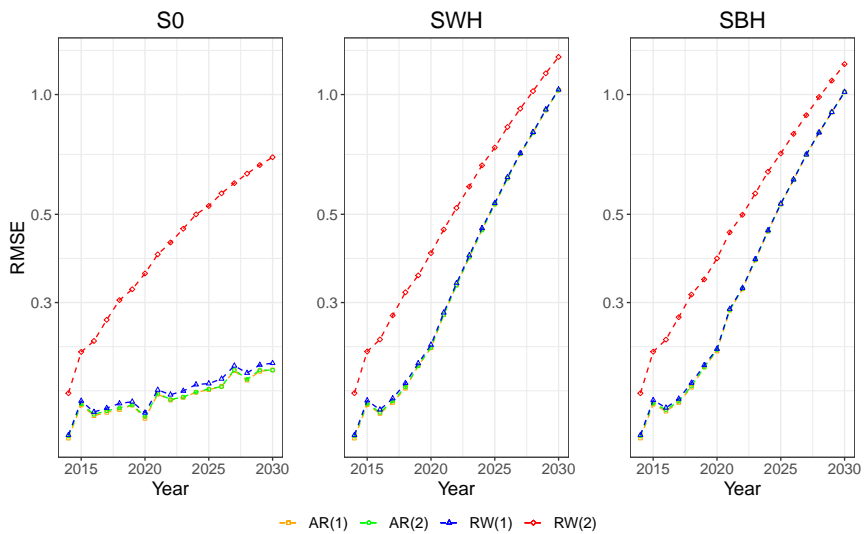


Figure A.13: RMSE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(NMR) simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.

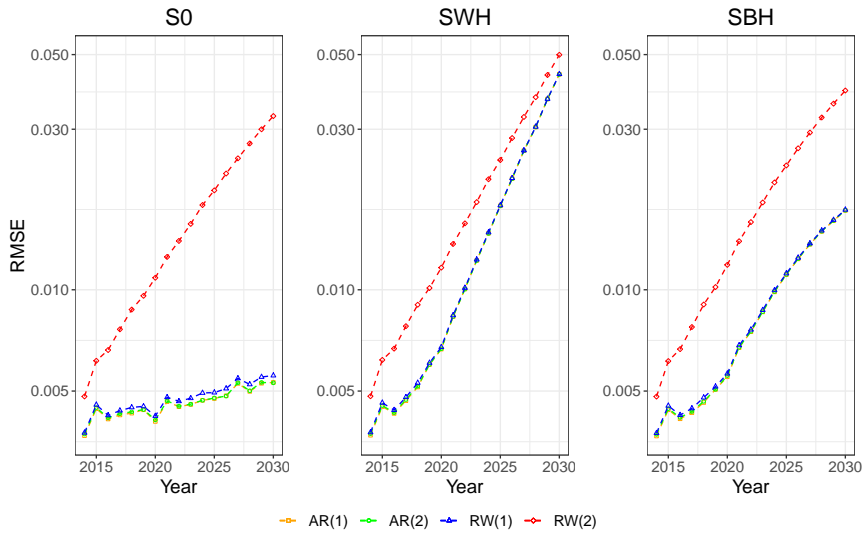


Figure A.14: RMSE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of NMR pr. 1000 live birth simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.

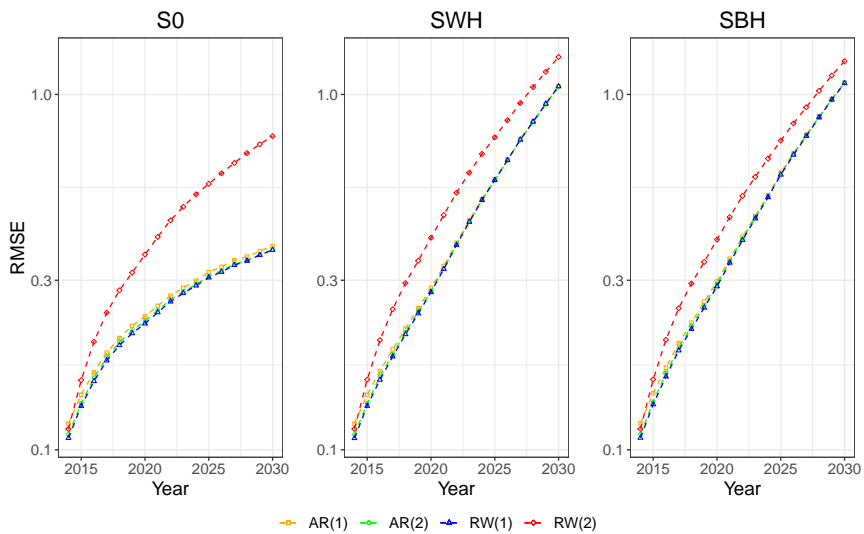


Figure A.15: RMSE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of logit(U5MR) simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.

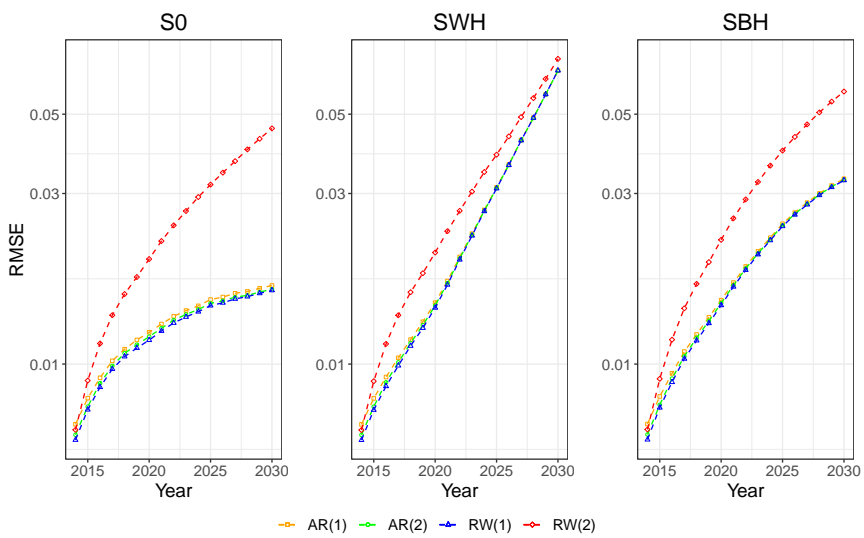


Figure A.16: RMSE of prediction error for models with temporal effect AR(1), AR(2), RW(1) and RW(2) of U5MR pr. 1000 live birth simulated by a model with temporal effect AR(2), for scenarios S0, SWB and SBH.)

Appendix B

Cross Plots of Responses

In this appendix, cross plots of potential covariates explored in Chapter 6 are presented.

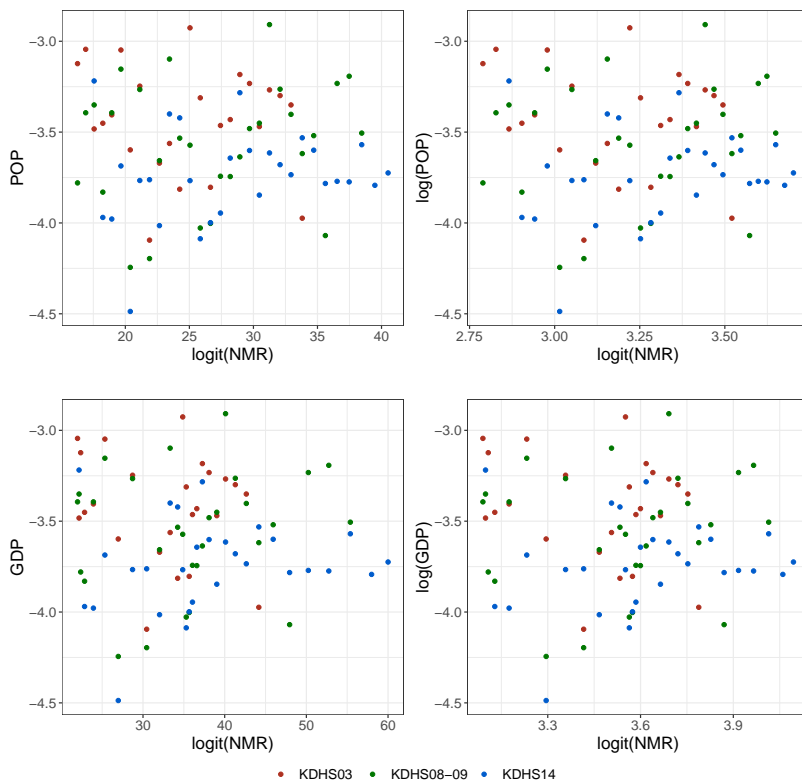


Figure B.1: logit(NMR) plotted against possible covariates, for Kenya years 1980-2010.

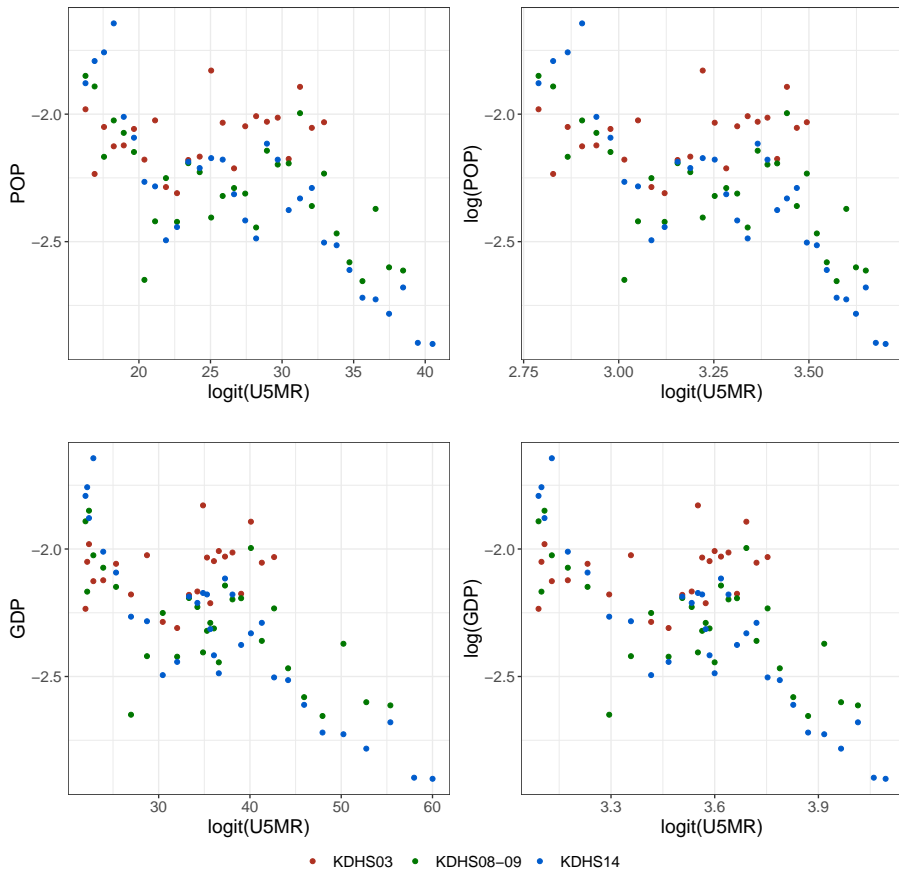


Figure B.2: logit(U5MR) plotted against possible covariates, for Kenya years 1980-2010.

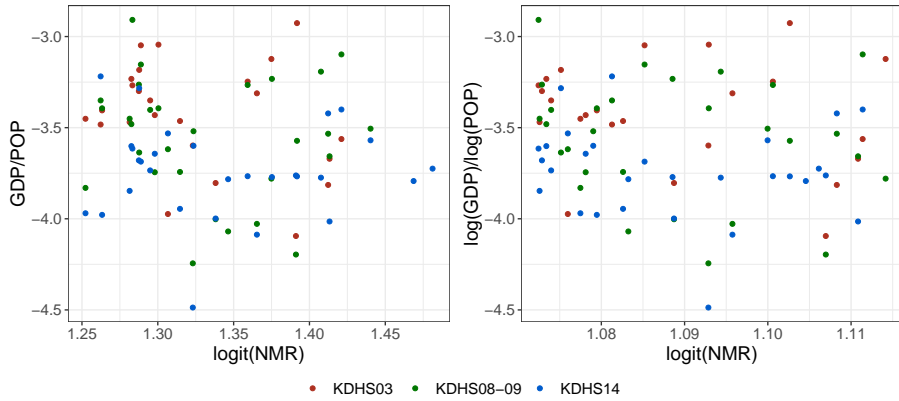


Figure B.3: logit(NMR) plotted against ratio and ratio of log of GDP and POP, for Kenya years 1980-2010.

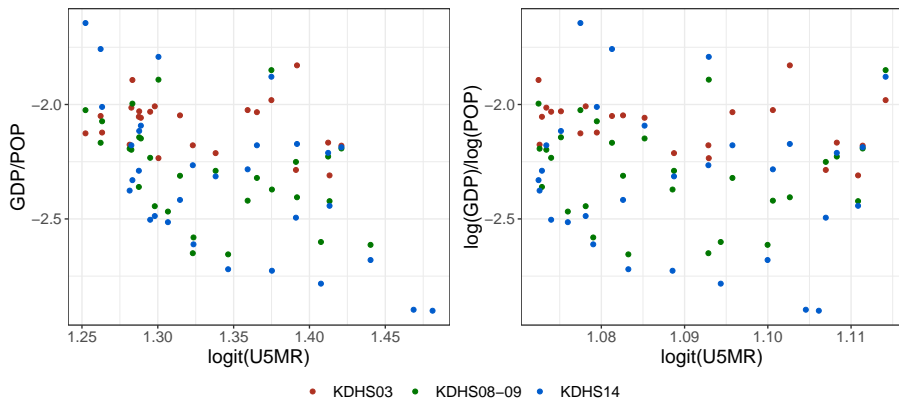


Figure B.4: logit(U5MR) plotted against ratio and ratio of log of GDP and POP, for Kenya years 1980-2010.

Appendix C

Estimated Coefficients, Chapter 5

In this appendix, the estimated coefficients estimated in Chapter 5 and their credible intervals are presented. The coefficients of NMR models are presented in Table C.1 and U5MR coefficients are presented in Table C.2.

Model name	Covariate	AR(2)	RW(1)
2	log(POP)	-0.07 [-0.45, 0.25]	-0.12 [-0.96, 0.38]
3	log(GDP1)	-0.09 [-0.42, 0.21]	-0.15 [-0.83, 0.32]
4	log(GDP2)	-0.10 [-0.41, 0.19]	-0.17 [-0.80, 0.25]
5	log(POP)	0.97 [-1.02, 2.76]	0.82 [-1.87, 2.37]
	log(GDP1)	-0.96 [-2.61, 0.84]	-0.87 [-2.25, 1.29]
6	log(POP)	1.03 [-0.63, 2.56]	1.00 [-1.10, 2.32]
	log(GDP2)	-0.98 [-2.34, 0.47]	-0.98 [-2.11, 0.61]
7	POP	0.00 [-0.02, 0.01]	0.00 [-0.03, 0.01]
8	GDP1	0.00 [-0.01, 0.01]	0.00 [-0.02, 0.01]
9	GDP2	0.00 [-0.01, 0.00]	0.00 [-0.02, 0.01]
10	POP	0.03 [-0.04, 0.09]	0.02 [-0.07, 0.08]
	GDP1	-0.02 [-0.06, 0.02]	-0.02 [-0.05, 0.04]
11	POP	0.03 [-0.02, 0.08]	0.03 [-0.04, 0.07]
	GDP2	-0.02 [-0.05, 0.01]	-0.02 [-0.05, 0.01]
12	<i>t</i>	0.00 [-0.01, 0.01]	0.00 [-0.03, 0.01]
13	<i>t</i>	0.01 [-0.11, 0.14]	0.01 [-0.12, 0.20]
	log(POP)	-0.50 [-5.08, 3.61]	-0.51 [-6.96, 3.90]
14	<i>t</i>	0.03 [-0.03, 0.08]	0.03 [-0.05, 0.07]
	log(GDP1)	-1.06 [-2.72, 0.71]	-0.92 [-2.30, 1.05]
15	<i>t</i>	0.04 [-0.01, 0.09]	0.04 [-0.02, 0.08]
	log(GDP2)	-1.22 [-2.72, 0.33]	-1.17 [-2.35, 0.30]
	<i>t</i>	0.03 [-0.09, 0.14]	0.02 [-0.10, 0.18]
16	log(POP)	0.17 [-3.94, 3.93]	0.11 [-6.41, 4.01]
	log(GDP1)	-1.04 [-2.83, 0.94]	-0.88 [-2.36, 1.46]
	<i>t</i>	0.04 [-0.07, 0.16]	0.03 [-0.06, 0.14]
17	log(POP)	-0.19 [-3.89, 3.29]	0.05 [-3.87, 3.10]
	log(GDP2)	-1.21 [-2.75, 0.41]	-1.18 [-2.34, 0.29]
18	<i>t</i>	0.06 [-0.11, 0.21]	0.08 [-0.08, 0.21]
	POP	-0.07 [-0.25, 0.13]	-0.10 [-0.26, 0.09]
19	<i>t</i>	0.02 [-0.02, 0.06]	0.02 [-0.05, 0.05]
	GDP1	-0.02 [-0.05, 0.02]	-0.02 [-0.05, 0.03]
20	<i>t</i>	0.02 [-0.02, 0.05]	0.02 [-0.03, 0.05]
	GDP2	-0.02 [-0.04, 0.01]	-0.02 [-0.04, 0.01]
21	<i>t</i>	0.01 [-0.19, 0.20]	0.04 [-0.12, 0.21]
	POP	0.01 [-0.28, 0.31]	-0.03 [-0.28, 0.20]
	GDP1	-0.02 [-0.07, 0.04]	-0.02 [-0.05, 0.03]
22	<i>t</i>	-0.09 [-0.36, 0.18]	-0.04 [-0.27, 0.19]
	POP	0.16 [-0.25, 0.57]	0.10 [-0.26, 0.43]
	GDP2	-0.04 [-0.09, 0.02]	-0.03 [-0.07, 0.02]

Table C.1: Coefficients and their 95% credible intervals for NMR, estimated using data from 1980-2010. No coefficients are significantly different from zero.

Model name	Covariate	AR(2)	RW(1)
2	log(POP)	-0.74 [-1.37, -0.11]	-0.73 [-1.57, 0.11]
3	log(GDP1)	-0.74 [-1.24, -0.26]	-0.78 [-1.42, -0.14]
4	log(GDP2)	-0.68 [-1.12, -0.23]	-0.69 [-1.29, -0.07]
5	log(POP)	1.01 [-1.05, 2.94]	0.86 [-1.11, 2.64]
	log(GDP1)	-1.62 [-3.37, 0.21]	-1.45 [-2.96, 0.20]
6	log(POP)	0.64 [-1.46, 2.49]	0.47 [-1.62, 2.33]
	log(GDP2)	-1.21 [-2.79, 0.55]	-1.04 [-2.51, 0.62]
7	POP	-0.03 [-0.05, -0.01]	-0.03 [-0.06, 0.00]
8	GDP1	-0.02 [-0.03, -0.01]	-0.02 [-0.04, -0.01]
9	GDP2	-0.02 [-0.03, -0.01]	-0.02 [-0.03, 0.00]
10	POP	0.04 [-0.02, 0.09]	0.03 [-0.03, 0.09]
	GDP1	-0.04 [-0.08, 0.00]	-0.04 [-0.07, 0.00]
11	POP	0.00 [-0.07, 0.06]	0.00 [-0.08, 0.06]
	GDP2	-0.02 [-0.05, 0.02]	-0.01 [-0.05, 0.03]
12	<i>t</i>	-0.02 [-0.04, -0.01]	-0.02 [-0.05, 0.00]
13	<i>t</i>	-0.09 [-0.22, 0.07]	-0.07 [-0.22, 0.10]
	log(POP)	2.15 [-3.17, 6.72]	1.54 [-3.96, 6.66]
14	<i>t</i>	0.01 [-0.05, 0.07]	0.01 [-0.04, 0.07]
	log(GDP1)	-1.04 [-2.82, 0.79]	-1.09 [-2.64, 0.51]
15	<i>t</i>	0.01 [-0.06, 0.07]	0.00 [-0.06, 0.07]
	log(GDP2)	-0.84 [-2.60, 1.02]	-0.80 [-2.39, 0.91]
16	<i>t</i>	-0.09 [-0.18, 0.04]	-0.08 [-0.18, 0.06]
	log(POP)	3.99 [-0.98, 7.25]	3.74 [-1.42, 7.22]
17	log(GDP1)	-1.65 [-3.02, 0.02]	-1.60 [-2.81, -0.01]
	<i>t</i>	-0.06 [-0.20, 0.09]	-0.05 [-0.20, 0.11]
	log(POP)	2.59 [-2.79, 6.64]	2.07 [-3.36, 6.67]
18	log(GDP2)	-1.00 [-2.57, 0.83]	-0.92 [-2.42, 0.81]
	<i>t</i>	0.10 [-0.09, 0.26]	0.08 [-0.12, 0.26]
19	POP	-0.15 [-0.34, 0.08]	-0.12 [-0.34, 0.12]
	<i>t</i>	0.03 [-0.02, 0.06]	0.02 [-0.02, 0.06]
20	GDP1	-0.04 [-0.07, -0.01]	-0.04 [-0.06, 0.00]
	<i>t</i>	0.00 [-0.04, 0.04]	0.00 [-0.05, 0.04]
21	GDP2	-0.02 [-0.05, 0.01]	-0.02 [-0.04, 0.02]
	<i>t</i>	0.04 [-0.14, 0.20]	0.03 [-0.16, 0.20]
22	POP	-0.02 [-0.25, 0.23]	-0.01 [-0.25, 0.24]
	GDP1	-0.04 [-0.08, 0.01]	-0.04 [-0.07, 0.01]
22	<i>t</i>	0.07 [-0.22, 0.37]	0.06 [-0.24, 0.36]
	POP	-0.11 [-0.55, 0.33]	-0.09 [-0.53, 0.34]
	GDP2	-0.01 [-0.06, 0.06]	0.00 [-0.06, 0.06]

Table C.2: Coefficients and their 95% credible intervals for U5MR, estimated using data from 1980-2010. The green coefficients are significantly different from zero.

Additional Plots Section 6.1

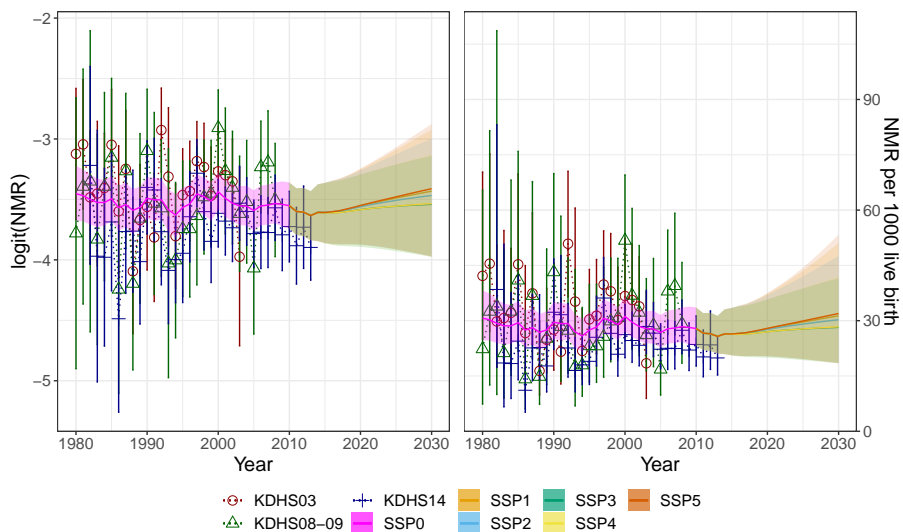


Figure D.1: logit(NMR) and NMR for Kenya fitted using model 14M with temporal effect AR(2).

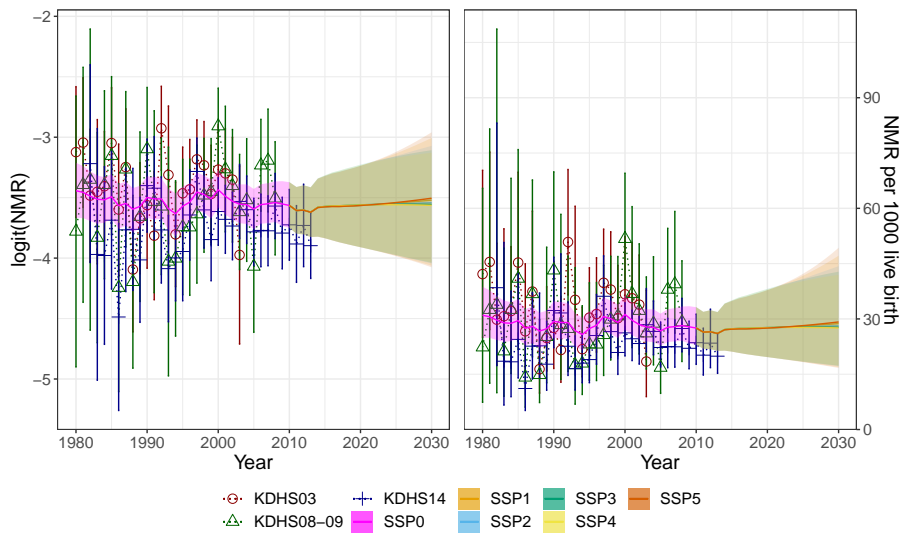


Figure D.2: $\text{logit}(\text{NMR})$ and NMR fitted using model15M with temporal effect AR(2).

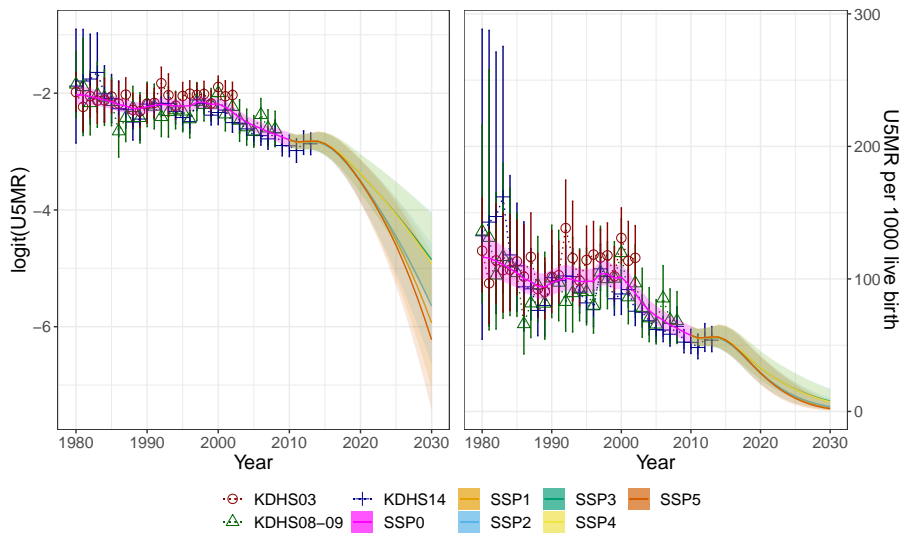


Figure D.3: $\text{logit}(\text{U5MR})$ and U5MR for Kenya fitted using model 8M with temporal effect RW(1).

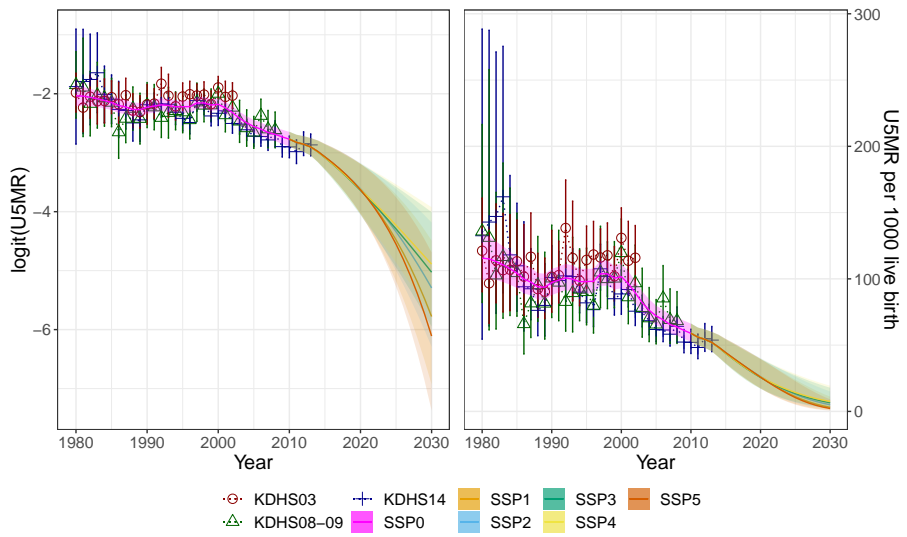


Figure D.4: $\text{logit}(U5MR)$ and $U5MR$ for Kenya fitted using model 9M with temporal effect $RW(1)$.

Appendix E

Additional Plots Section 6.2

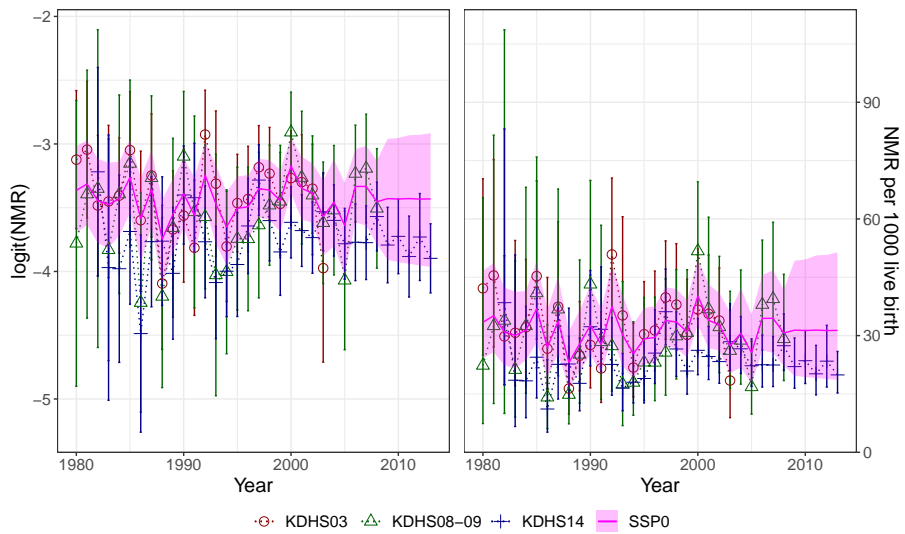


Figure E.1: logit(NMR) and NMR per 1000 live birth for Kenya. Model 12 with temporal effect AR(2), trained using data from KDHS03 and KDHS08-09 from 1980-2008.

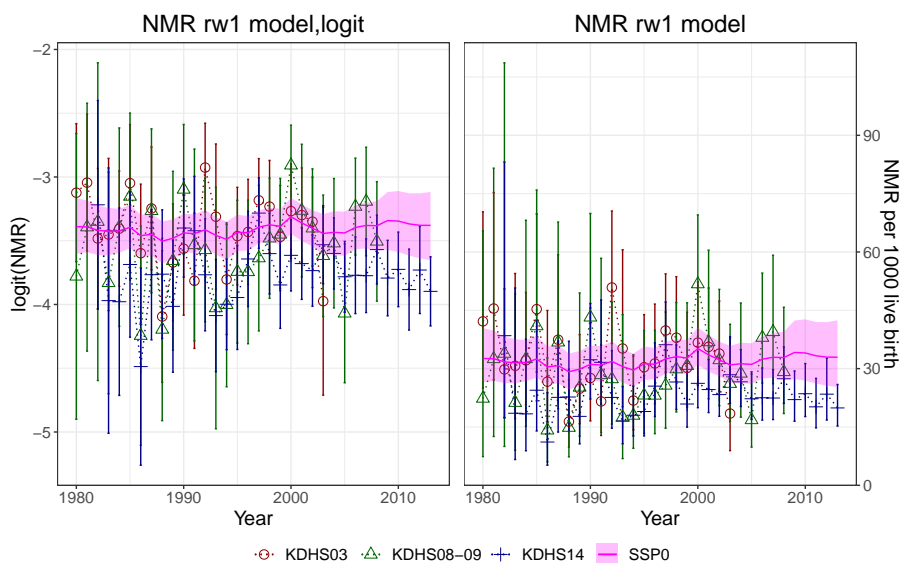


Figure E.2: logit(NMR) and NMR per 1000 live birth for Kenya. Model 12M with temporal effect AR(2), trained using data from KDHS03, KDHS08-09, UDHS06, UDHS11, UDHS16, TDHS05, TDHS10 and TDHS16, from 1980-2013.

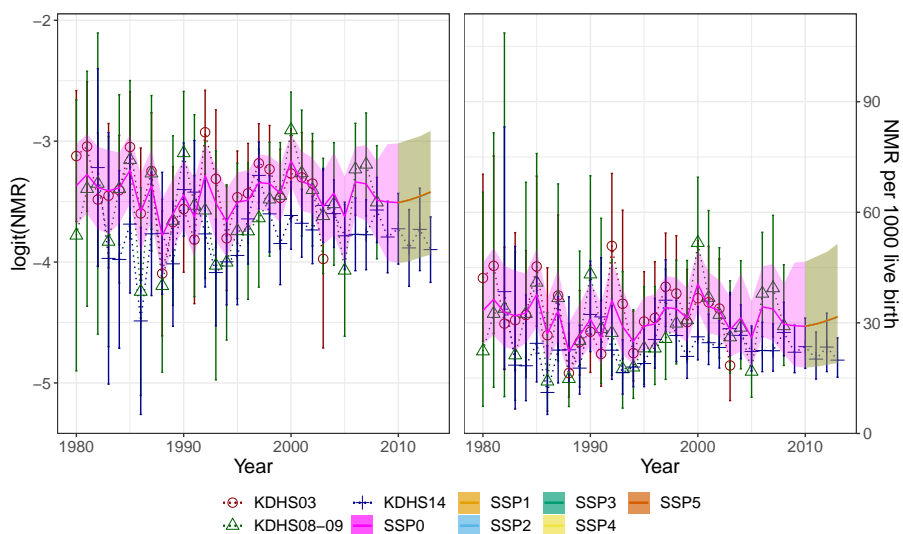


Figure E.3: logit(NMR) and NMR per 1000 live birth for Kenya. Model 14 with temporal effect AR(2), trained using data from KDHS03 and KDHS08-09 from 1980-2008 and log(GDP1) as covariate.

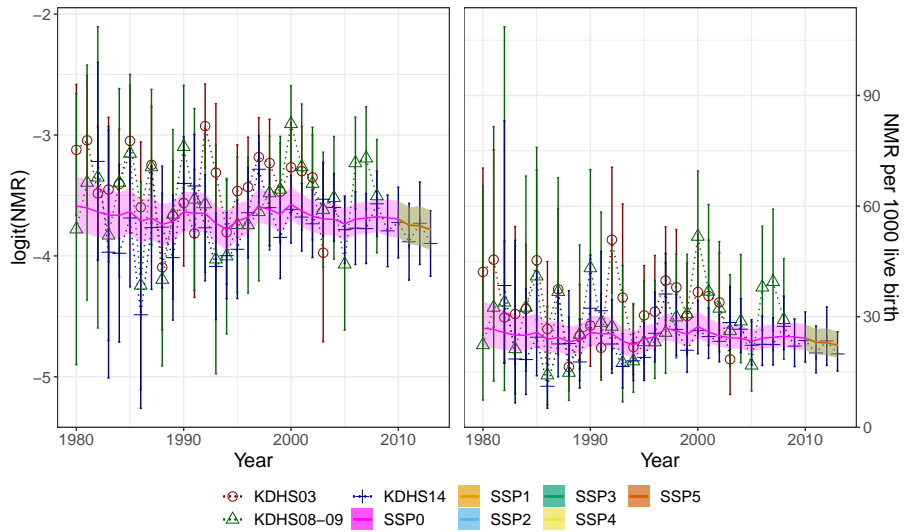


Figure E.4: logit(NMR) and NMR per 1000 live birth for Kenya. Model 14M with temporal effect AR(2), trained using data from KDHS03, KDHS08-09, UDHS06, UDHS11, UDHS16, TDHS05, TDHS10 and TDHS16, from 1980-2013 and log(GDP1) as covariate.

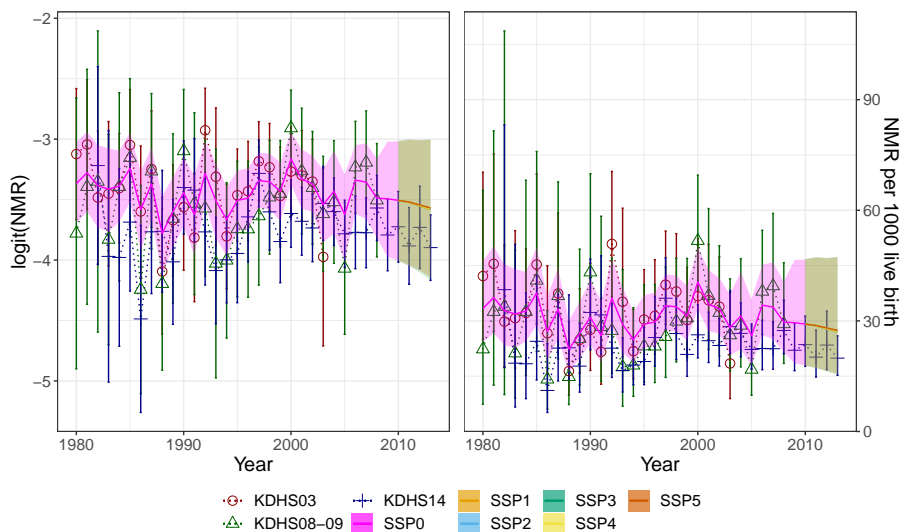


Figure E.5: logit(NMR) and NMR per 1000 live birth for Kenya. Model 15 with temporal effect AR(2), trained using data from KDHS03 and KDHS08-09 from 1980-2008 and log(GDP2) as covariate.

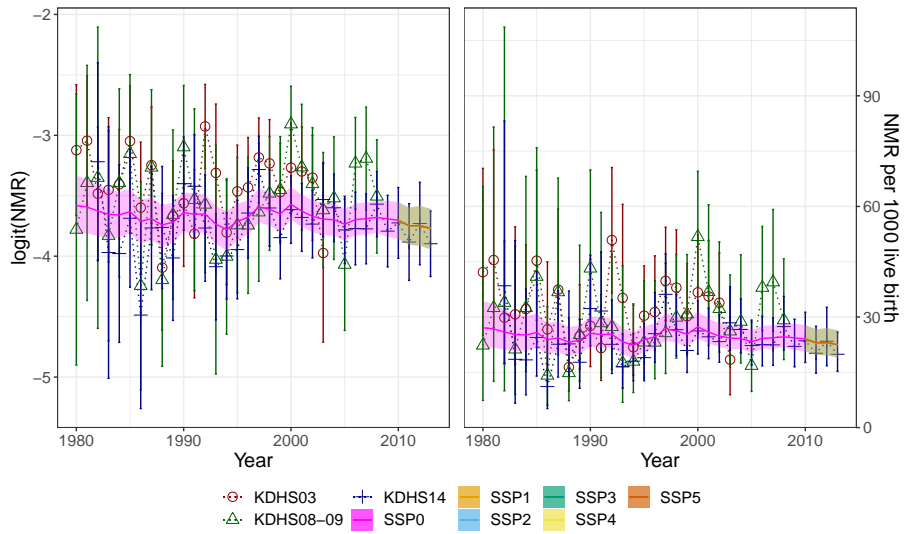


Figure E.6: $\text{logit}(\text{NMR})$ and NMR per 1000 live birth for Kenya. Model 15M with temporal effect AR(2), trained using data from KDHS03, KDHS08-09, UDHS06, UDHS11, UDHS16, TDHS05, TDHS10 and TDHS16, from 1980-2013 and $\text{log}(\text{GDP}2)$ as covariate.

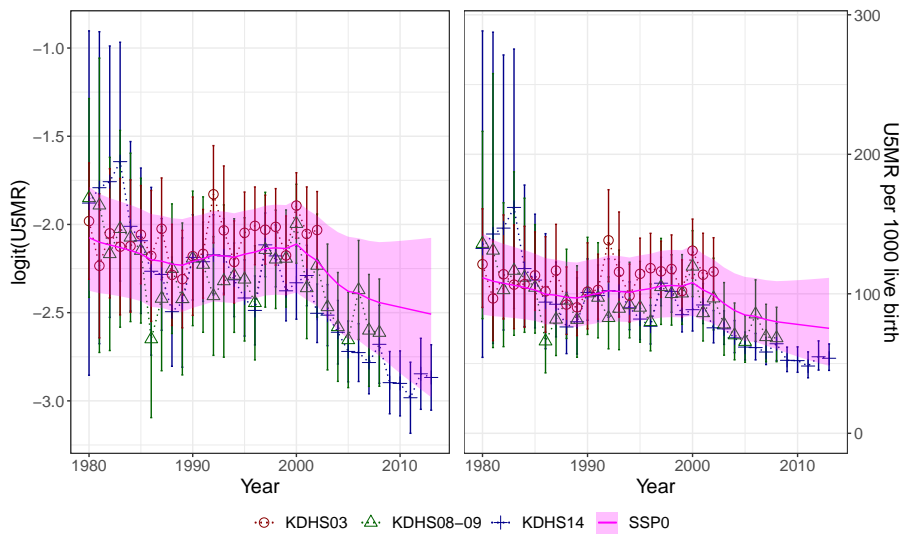


Figure E.7: $\text{logit}(\text{U5MR})$ and U5MR per 1000 live birth for Kenya. Model 12, trained using data from KDHS03 and KDHS08-09 from 1980-2008.

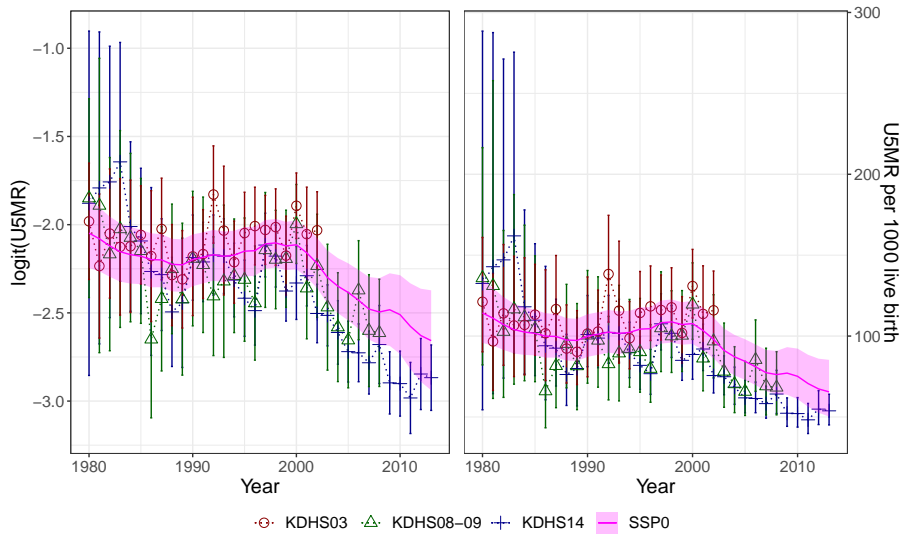


Figure E.8: $\text{logit}(U5MR)$ and $U5MR$ per 1000 live birth for Kenya. Model 12M with temporal effect $RW(1)$, trained using data from KDHS03, KDHS08-09, UDHS06, UDHS11, UDHS16, TDHS05, TDHS10 and TDHS16, from 1980-2013.

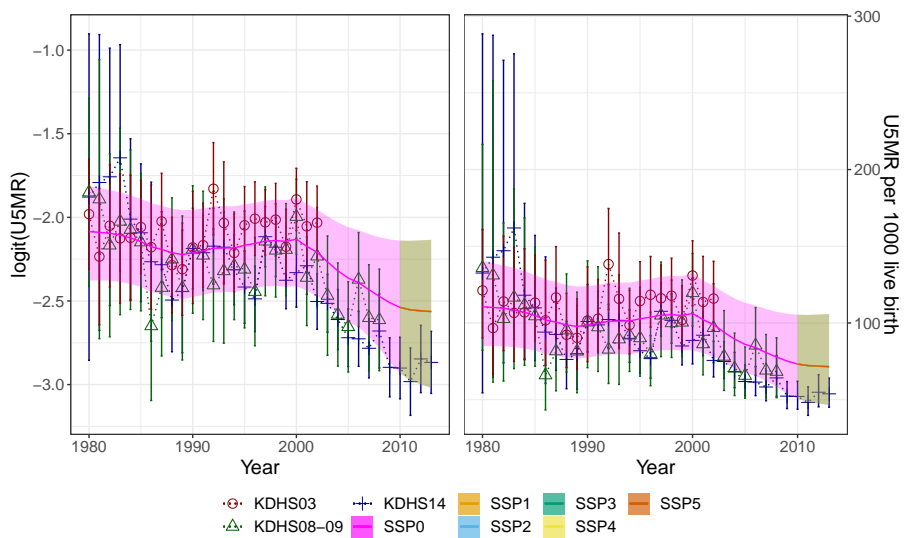


Figure E.9: $\text{logit}(U5MR)$ and $U5MR$ per 1000 live birth for Kenya. Model 8 with temporal effect $RW(1)$, trained using data from KDHS03 and KDHS08-09 from 1980-2008 and GDP1 as covariate.

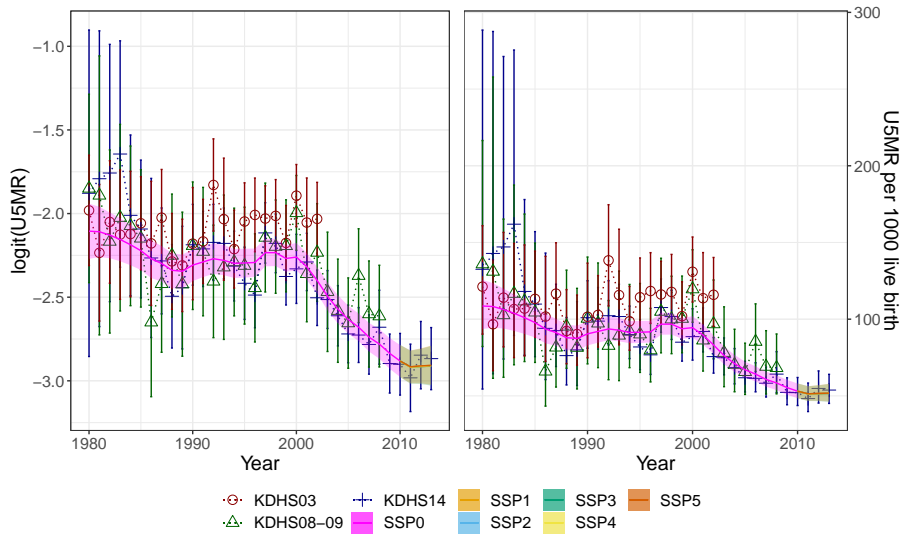


Figure E.10: $\text{logit}(U5MR)$ and $U5MR$ per 1000 live birth for Kenya. Model 8M with temporal effect $RW(1)$, trained using data from KDHS03, KDHS08-09, UDHS06, UDHS11, UDHS16, TDHS05, TDHS10 and TDHS16, from 1980-2013 and GDP1 as covariate.

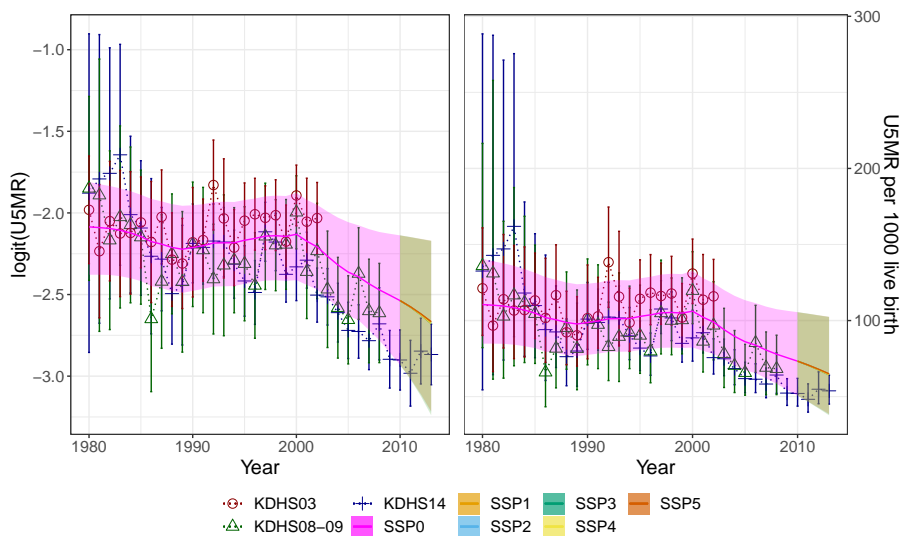


Figure E.11: $\text{logit}(U5MR)$ and $U5MR$ per 1000 live birth for Kenya. Model 9 with temporal effect $RW(1)$, trained using data from KDHS03 and KDHS08-09 from 1980-2008 and GDP1 as covariate.

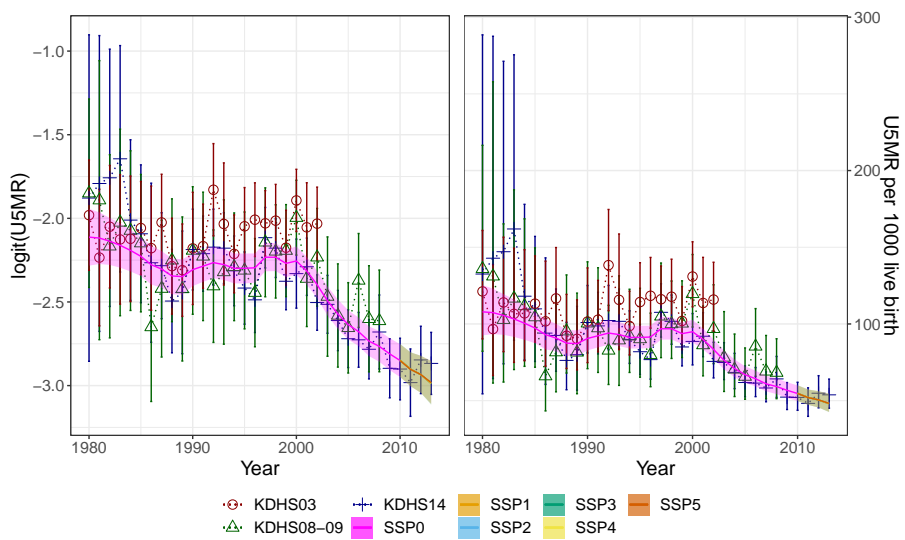


Figure E.12: logit(U5MR) and U5MR per 1000 live birth for Kenya. Model 9M with temporal effect RW(1), trained using data from KDHS03, KDHS08-09, UDHS06, UDHS11, UDHS16, TDHS05, TDHS10 and TDHS16, from 1980-2013 and GDP1 as covariate.

