Bayesian Models in Population Projections and Climate Change Forecast

Peiran Liu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Adrian E. Raftery, Chair

Adrian Dobra

Don Percival

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Bayesian Models in Population Projections and Climate Change Forecast

Peiran Liu

Chair of the Supervisory Committee:
Adrian E. Raftery
Department of Statistics and Sociology

The goal of this dissertation is to develop new methods for probabilistic projections in demography and climate science.

In the first project, I propose new models for improving estimates and projections of total fertility rates (TFR) for most countries. I develop a new Bayesian hierarchical model for projection that incorporates an additional layer modeling bias and measurement error in estimates of past TFR. In this way, the uncertainty of past TFR estimates is evaluated and incorporated into the projections. The proposed method addresses the issue of data correction across different versions of the UN's *World Population Prospects*. The resulting projections are of an appropriate form to be taken as alternatives to results from Alkema et al. [2011], the current default probabilistic projections of TFR produced every two years by the United Nations Population Division. Moreover, the UN Population Division has until now used used five-year average TFR, but plans to switch to annual values for greater precision. The five-year model is inadequate to represent the autocorrelation of the annual series. I extend the projection model to this situation by adding an autoregressive component on the random distortion to the fertility transition model. I implemented the new model computationally, and issued as a major update of the R package **bayesTFR**, currently used by the UN. The package has been updated to be generic for different settings, and computational difficulties has been resolved by vectorization in the Markov Chain Monte Carlo sampling

process.

In a separate project, I propose a method for probabilistic forecasting of global carbon emissions, on a country-specific basis. These forecasts are further linked with atmospheric-oceanic global circulation models to generate probabilistic forecasts of global mean temperature. By connecting representative concentration pathways to probabilistic forecasts of carbon emissions, I generated the first probabilistic forecasts of global mean temperature. In out-of-sample predictive experiments, I showed that the resulting method provides accurate forecasts and well-calibrated forecasting intervals. I also built a dataset summarizing the country-specific National Determined Contributions in carbon emission reductions signed on the Paris Agreement. By making reasonable assumptions on different scenarios, I compute the probability for major countries achieving their NDC promises, and the additional efforts need to achieve the Paris Agreement goals of limiting global warming by 2100 to 2°C or 1.5°C.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

UN: The United Nations.

UNPD: The United Nations Population Division.

TFR: Total Fertility Rate.

WFD: The World Fertility Data.

UNFCCC: The United Nations Framework Convention on Climate Change.

CMIP: The Coupled Model Intercomparison Project.

GCM: General circulation model.

AOGCM: Atmosphere-ocean general circulation model.

ASFR: Age-specific Fertility Rate.

IPCC: Intergovernmental Panel on Climate Change.

RCP: Representative concentration pathways.

EMIC: The earth system models of intermediate complexity.

ESM: The earth system models.

DHS: The Demographic and Health Surveys.

MICS: Multiple Indicator CLuster Surveys.

VR: Vital Registration.

WPP: World Population Prospects.

MCMC: Markov Chain Monte Carlo.

MAE: Mean Absolute Error.

UNICEF: The United Nations Children's Fund.

(I)NDC: (Intended) National Determined Contributions.

EU: The European Union.

GDP: Gross Domestic Product.

PPP: The Purchasing Power Parity.

JAGS: Just Another Gibbs Sampler.

(P)ACF: (Partial) auto-correlation function.

IPAT: IPAT model.

BAU: Business-as-Usual.

USA: The United States of America.

GT: Giga-tons.

ACKNOWLEDGMENTS

First and foremost, I want to thank Adrian Raftery for his mentorship. From the beginning of my PhD program, his professionalism led my research and helped me form rigorous working habits. Additionally, I cannot forget that during the worst time of my PhD career, his kindness and sincere help gave me courage and brought me through the bad time. I do not believe I would be here without his support, guidance and mentorship.

I also want to thank you to Adrian Dobra, Don Percival and David Battisti for serving on my supervisory committee, offering advice on my research, and giving me recommendations for my job hunting. I also thank Darryl Holman for serving as my graduate school representative, and making helpful suggestions.

I would like to thank the Department of Statistics at the University of Washington, both for the financial support and the great environment. Specifically I want to thank the team members of my research group, Hana Ševčíková, Jon Azose, Yicheng Li, Daphne Liu and Nathan Welch. I have got tremendous help during my PhD career.

Finally, I am grateful to my parents, not only for supporting my life in the US over these years, but also for raising me up. I believe my father could be proud of me in the heaven. I would also like to thank Sherry for standing with me through the journey.

DEDICATION

to my family, and for a better world.

Chapter 1

INTRODUCTION

1.1 Motivation

Demography is a foundation of the social sciences. It is hard to imagine a social science able to advance steadily without first processing basic information about the human population that it studies [Xie, 2000]. According to United Nations [2019b], the world population has reached 7.8 billion and is expected to surpass 10 billion in 2060. Apart from the population increase, highly diverse populations exist across each continent and country. Yet the major industrialized world has seen the average number of children per woman (as measured by the total fertility rate; TFR) drop for over the last century, from an average of 5 children in the 19^{th} century Europe to the current 1.6. In addition, substantial progress in public health systems across the world has increased life expectancy significantly, changing age distributions of industrialized countries. This has influenced decision makers and their policies towards schools, healthcare, and retirement systems.

This pattern is shared by a great number of developing regions (notably China), but it remains unclear whether all countries will go through a similar transition at a similar pace [Alho and Spencer, 2006]. For example, it seems that the fertility transition that happened in Asia may differ from Africa's transition. The UN [United Nations, 2019b] noted Thailand's TFR decreasing from six to two over thirty years ending in 1990, and these phenomena were shared by many Asian countries, leading to severe aging problems in Japan and other nations. However, many African countries have experienced a much slower fertility transition. Researchers have identified this phenomenon and tried to explain its existence [Kebede et al., 2019, Shapiro and Hinde, 2017, Bongaarts and Casterline, 2013, Shapiro and Gebreselassie, 2008]. Population heterogeneity gives demographics an important role in the day-to-day

decision-making of national and local governments.

Relatively simple mathematical methods have traditionally been used to estimate and project demographic trends. Population projections have been produced routinely since the 1940s using a deterministic method called the "cohort-component method" [Cannan, 1895, Whelpton, 1928, 1936, Preston et al., 2000]. These methods will continue to serve demography well, but there are reasons for expanding demographers' toolkits in a statistical direction. In 2015, the United Nations Population Division (UNPD) produced its first official probabilistic population projections for all countries by combining probabilistic projections of fertility and mortality rates, setting a new standard for applying statistics to demographic projection [United Nations, 2015b].

The second and third chapters of this dissertation seek to improve UNPD TFR estimates and forecasts. The UNPD publishes population projections (by age and sex), mortality and fertility rates, and net migration estimates for all countries using five-year age groups to cover five-year periods up to the year 2100. However, after reviewing TFR estimates from multiple World Fertility Data sources [United Nations, 2015b, 2019b], past TFRs are uncertain and the five-year estimates do not account for fluctuating demographic data. As the UNPD switches from forecasting for five-year age groups in five-year periods to annual reporting on one-year age groups, it will incorporate the unstated uncertainty of past estimates into future publications of both estimation and projections. Chapters 2 and 3 of this dissertation attempt to resolve this.

We plan to use statistical tools to both improve population forecasting and analyze demographic change outcomes. Population increase will intensify competition for arable land, clean water, and raw materials, which will accompany needs for food, housing and goods, increased in soil erosion and deforestation activities, and related climate change contributions [Alho and Spencer, 2006]. The 2015 Paris Agreement has set a target of 1.5°C for annual global temperature increases [UNFCCC, 2015]. Whether such a targets could be achieved is of great interest and motivates the work in Chapter 4 of this dissertation.

As uncertainty is discussed according to scenario, Chapter 4 will also provide a method

for making forecasts probabilistically, combining climate models with probabilistic forecasts of socioeconomic factors. The Coupled Model Intercomparison Project, Phase 5 [Hurrell et al., 2011] (CMIP 5), is a standard experimental protocol for studying the output of coupled atmosphere-ocean general circulation models (AOGCMs). This tool is used by climate scientists making deterministic climate change forecasts.

1.2 Background

This section is based in part on articles by Liu and Raftery [2020b], Raftery et al. [2017], with additional background provided in Sections 2.1, 3.1, and 4.1.

1.2.1 Population modeling and fertility rate projection

The most widely-used population projections for individual countries are produced by national statistical agencies [U.S. Census Bureau, 2017] as well as international organizations, including the UNPD or Eurostat. Most projections are based on the cohort-component method [Cannan, 1895, Whelpton, 1928, 1936, Preston et al., 2000], which requires assumptions about future fertility, mortality, and migration rates by age and sex. As the work in this dissertation focuses mainly on TFRs, explaining how TFR relate to age-specific fertility rates (ASFR).

Raftery et al. [2012] indicated after TFRs are estimated and projected, they can be converted into ASFRs using model fertility schedules. The United Nations [2012] obtained future age patterns of fertility were obtained by interpolating linearly between a starting proportionate age pattern of fertility and a target model pattern. The latter was chosen from 15 proportionate patterns, with mean childbearing age ranging from 24 to 28.5 years, with low fertility countries being assumed to reach a target model pattern (selected from five target patterns of either Europe or countries undergoing economic transition) by 2025–2030. Ševčíková et al. [2016] proposed a new method for projecting percent-ASFR that combined observed and assumed national trends towards global age fertility patterns.

This dissertation focuses on developing TFR forecasting. Existing methods for TFR

forecasting shall be classified based on how they estimate and project TFRs: either by curve fitting or extrapolation. Curve-fitting method research, such as that of Peristera and Kostaki [2007], makes assumptions about ASFR patterns, and can present different model interpretations by introducing observations, such as younger mothers or educational changes. Extrapolation methods , by contrast, focus specifically on fertility rates and forecast based on historical and population base information instead of factors that potentially affect TFRs. For example, Cheng and Lin [2010] built a model based on fertility dynamics (age, birth cohort and calendar year), while Myrskylä et al. [2013], produced a simplified model that used principal component analysis (i.e., the Lee–Carter model) for its forecasting.

Admittedly, curve-fitting methods go into more depth on specific factors driving the changes in fertility rates. For example, Evans [1986] went into detail about racial differences in relation to fertility rates, while Coale and McNeil [1972] focused on how age distributions of a female's first marriage affected the representative schedule and fertility rates. Such insights are difficult to obtain through extrapolation. However, extrapolation methods do significantly outperform curve-fitting methods when projections are made [Bohk-Ewald et al., 2018], and many curve-fitting methods cannot improve upon freeze-rate methods when forecasting assumes the most recent observed data to be a future constant.

There are two major reasons for the poor performance of curve-fitting methods in fertility rates projections. The first reason is that on the macro-level, the fertility rate change is a combined results of multiple social-economic factors, such as race and culture [Evans, 1986], female eduction [Kebede et al., 2019], and family planning resources [Poston Jr and Gu, 1987]. It is way too difficult to include and predict all of these factors and to combine their effects on predicting fertility rates, and considering part of them will not improve the model forecast automatically. On the other hand, TFR is resistent to policy makers, and some factors that might appear important for curve fitting exert little influence on TFR. As such, curve-fitting methods produce poor forecasts, which can be illustrated with China. In 1979, the One Child Policy was introduced accompanied by extensive discussion about forecasting. For example, Merli and Raftery [2000] attempted to correct Chinese TFR estimates for the

year 2000. However, the interesting fact is that while the TFR in China drops from 6.3 from 1965-1970, to 2.5 from 1980-1985, many other Asian countries with One Child Policy saw similar drops in TFR in the same period, such as South Korea from 5.6 in 1960-1965 to 2.2 in 1980-1985. It has been well argued that the impetus of fertility rate changes is deeply internal, and that policies could only exert little influence.

Therefore, despite the notable contribution from curve-fitting methods, we eschew all of them from our model, as it will not support forecasting with noisy data, such as education or economic development, to the smoothly-changing time series, TFR. Alkema et al. [2011] laid the foundation of probabilistic TFR forecast by building hierarchical models, which were officially adopted by the UN in2015 [United Nations, 2015b]. Chapter 2 will discuss this method, its limitations, and how it can be improved.

1.2.2 Global mean temperature forecast

CMIP5 was adopted by Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report, and integrates AOGCMs, the standard models used by climate scientists [Taylor et al., 2012]. Based on the deterministic nature of the models, climate modelers need time series for future concentrations of greenhouse gas emissions and air pollutants as well as land-use changes as inputs to "projects" future climate change [Van Vuuren et al., 2011]. In IPCC reports, climate change projections are based on four different pathways for emissions and land use up to 2100, with each representative concentration pathway (RCP) based on a different socioeconomic scenario for the world's future and developed by a different research group [Raftery et al., 2017]. Chapter 4 explains part of the background, and what follows is a brief introduction of CMIP 5 models, RCPs, and their limitations.

CMIP phases, such as the multimodal dataset of CMIP 3 [Taylor et al., 2012], that provided the basis for hundreds of peer-reviewed papers, and played a prominent role in the IPCC's Fourth Assessment Report on climate variability and change. The CMIP project provides a multimodel context for assessing mechanisms responsible for model differences, examining climate predictability and determining why similarly forced models produce a

range of responses. AOGCMs and Earth system models of intermediate complexity are used to perform many CMIP experiments. Standard AOGCMs are possibly coupled to biogeochemical components as Earth system models, and are used for some long-term experiments. With most modeling groups worldwide participating, CMIP 5 is considered useful because of its solid climate and atmospherical science foundation, and is of relevance to national and international climate science assessments.

The future concentrations and emissions of greenhouse gases, air pollutants, and land-use changes used in CMIP 5 forecasts, are summarized by RCPs. RCPs are based on scenarios developed independently by modeling groups and published to be, as a set, representative of all emissions and concentrations literature. Four RCP scenarios were selected and named according to their radiative forcing target level for 2100. They include one mitigation scenario leading to a very low forcing level (RCP2.6), two medium stabilization scenarios (RCP4.5/6), and one very high baseline emission scenario (RCP8.5) [Van Vuuren et al., 2011].

However, the major contradiction lies between RCPs description and their usage in CMIP 5, which treats RCPs as emissions forecast and uses them as input when forecasting future climate changes. Van Vuuren et al. [2011] noted that "RCPs should not be interpreted as forecasts or absolute bounds, or be seen as policy prescriptive." Thus, CMIP 5 climate change "forecast" are actually based on descriptive assumptions of the future that lack statistically-based forecasting and validation. Therefore, Chapter 4 will include statistical predictions of socioeconomic factors along with resulting climate change forecasts.

Chapter 4 of his thesis will also contain models built for socioeconomic factors based on population projections made in Chapters 2 and 3 as well as other research [Raftery et al., 2012, 2014a], and will link CMIP 5 models to those socio-economic factors. Raftery et al. [2017] focused on predicting socioeconomic factors, and Chapter 4 will describe the methods for making probabilistic climate changes forecast with the probabilistic socioeconomic factors forecast. Notably, a key contribution of my work is a method for coupling the Paris Agreement with a model assessing its impact. I have assessed how likely countries are to meet their commitments, what the climate change outcome would be assuming that the Paris

Agreement is realized, and how much more is needed beyond realizing the emission reduction commitments to limiting the global warming by 2°C or 1.5°C. More detailed background of the Paris Agreement can be found in Section 4.5.

1.3 Summary of Contributions to Statistics

This dissertation solves statistical problems in demography and climate science by developing new statistical models and improving estimation methods. These contributions can be summarized according to six areas.

1.3.1 New statistical model for estimating past TFR uncertainty

A Bayesian hierarchical model is developed to estimate the uncertainty of past TFRs for most countries, incorporating the estimated uncertainty in TFR projections. Currently, United Nations [2015b] makes its official TFR projections based on its previous TFR estimates for all countries, and the uncertainty of these estimates is not considered in the projection model, causing underestimation the uncertainty in TFR projections.

Specifically, by analyzing the distribution of TFR estimates from different sources in different countries, a generic model is built for all estimates in the World Fertility Data (WFD) [United Nations, 2015a], accounting for bias and measurement error variance for past estimates separately. The model in Alkema et al. [2011] is subsequently extended by adding the model for WFD as another layer on top of the existing Bayesian hierarchical model. The model is grounded in demographic sciences, and with its hierarchical structure, it can generate robust estimates with little data. A validation study has shown that the proposed model can solve the problem of underestimating uncertainty in forecasting by generating better mean absolute errors and calibrating the right level of uncertainty.

1.3.2 New statistical model for annual TFR estimation and projection

A model for annual TFR estimation and projections is also developed that differs from the model for five-year average TFR; this model uses annual data, is more high-dimensional,

and has a autocorrelation structure that is more significant than the five-year average data. To estimate annual TFR robustly, a new statistical model is built on top of the Bayesian hierarchical model described in Section 1.3.1 by extending the Alkema et al. [2011] model with extra autoregressive (AR)(1) term in the fertility transition phase. This provides robust bias and measurement error standard deviation estimates, especially for countries with high-quality Vital Registration data history.

As the number of parameters to estimate has increased from 5 per country in Alkema et al. [2011] to 80 per country, it is more likely to generate ill-distributed estimates. To address this, the Metropolis-Hastings algorithm is adjusted with larger step sizes and prior distributions of hyper-parameters are adjusted, especially those related to the uncertainties. With these carefully designed new models, robust estimates of past annual TFR estimates and reasonable forecasts for all countries can be obtained.

1.3.3 Computational improvements in TFR estimation procedure

As described in Section 1.3.2, the number of parameters to estimate has increased from 5 per country to 80, causing computational costs to increase by an even larger scale since with more parameters to estimate, it will take longer for the Markov Chain Monte Carlo (MCMC) process to converge. To produce converged estimates within the time limit, the computational performance of the model estimation was improved by implementing the Metropolis-Hastings algorithm in a vectorized manner. With our new method, we were able to generate the same number of posterior samples in twice the time with the same computation settings.

The method is implemented in the R package `bayesTFR`, which is widely used for statistical demographers. With this extension, we have further improved the impact of statistical models on demographic research. The package is well designed for demographers who run statistical estimation and projections for TFR and who may not be very familiar with Bayesian hierarchical modeling techniques.

1.3.4 New statistical model for CO_2 emission forecasts for most countries

To make a probabilistic forecast of CO_2 emission levels for most countries, we use a simple form of the Kaya identity [Kaya et al., 1997], which expresses future emission levels in a country as a product of three components: population, GDP per capita, and carbon intensity (CO_2 emissions per unit of GDP). We have built Bayesian hierarchical models for GDP per capita and carbon intensity and combined these models with the population projection model.

To determine GDP per capita, we have built a Bayesian hierarchical model for all countries based on the idea of a world technology frontier (represented by the US for 1960-2015, the period of our data), towards which countries may converge. This model is motivated by that of Lucas [2000], which allows countries with high current growth rates to continue growing fast in the short to medium term while avoiding unrealistically high long-term forecasts.

For carbon intensity, most countries have reached peak intensity by 2015; subsequently, their carbon intensity has been trending downwards. We detected the peaks with LOWESS smoothing and modelled the post-peak data as a linear trend as well as an autoregressive random process. This model was extended with a Bayesian hierarchical structure for all countries so that countries with a short history of intensity decline could be projected realistically.

We also evaluated the within-country correlation between model errors of population, GDP per capita, and carbon intensity and built a joint model for GDP and intensity to incorporate the correlations between model errors. The model is elaborated in more detail in Raftery et al. [2017].

1.3.5 New model for global mean temperature forecast

We have developed statistical models for the probabilistic forecast of global mean temperature. This is achieved by combining previous CO_2 emission forecasts and two new statistical

models. We first developed a Bayesian hierarchical model quantifying the uncertainty of the CMIP 5 model forecast of global mean temperature by building a hierarchical autoregressive model for the measurement errors of past global mean temperature estimates from each CMIP 5 models. We then developed statistical models for the relationship between cumulative CO_2 emissions and temperature projections in CMIP 5.

In order to make probabilistic projections of global mean temperature statistically, we have taken the projected CO_2 emissions in Section 1.3.4 as input, using the model for the relationship between cumulative emissions and temperature projections to obtain corresponding global mean temperature projections from all CMIP 5 models. We then combined these forecasts with their uncertainty to generate the final probabilistic projections of global mean temperature changes. This model is grounded in climate science, but it is fully statistical and avoids running global circulation models, which are expensive, deterministic, and not actual projections with statistical validity.

1.3.6 Solving climate assessment inverse problems

Solving the inverse problem is important in applying statistical models in social sciences, which could be used to understand the need for changes in input for certain objectives. These questions are not answered for probabilistic projection models, such as the model in forecasting global mean temperatures as described in Section 1.3.5.

We have made use of the bisection method in our global mean temperature forecast model, so that we could compute the required emission reduction for achieving certain global mean temperature objectives, such as having an even chance of limiting global warming by 2°C Celcius by the end of this century. In the process, we introduce another ratio parameter, controlling the ratio of assumed CO_2 emission trajectories and actual forecasts, and use the bi-section method on the ratio to achieve a sufficiently accurate percentage reduction for mitigating global warming.

1.3.7 Summary

In summary, we have developed new Bayesian hierarchical models for statistical projections of TFR, CO_2 emissions, and climate change. All these problems share two similar characteristics: high dimensionality and limited data points. In TFR estimation and projections, the number of raw data points is 20,000, while we have to estimate 16,000 TFR values over 201 time series. Moreover, 120 of these time series do not have high-quality historical estimates, and some of them have only a few historical data points. Similarly, in forecasting CO_2 emissions projections, we have built a model for carbon intensity using only post-peak data, but some countries have only a few years of relevant data. We have carefully built hierarchical structures on the basis of these data to solve this problem, and through the process, the Bayesian hierarchical model has demonstrated its utility in future use in these areas. The Bayesian hierarchical model can generate robust estimates for cases with limited data, and time series with poor estimates can use information from other time series to generate reasonable forecasts.

Moreover, we have developed a new framework to build statistical models and make probabilistic forecasts in climate science. This is achieved by analyzing the statistical relationship between socio-economic factors and corresponding climate outcomes, which in our case is the global mean temperature. We then converted the problem of running expensive global circulation models into making probabilistic forecasts of socio-economic factors, where statistical models, such as the models we proposed in Section 1.3.4, can be effective. This framework can be used by future statisticians to model climate changes statistically.

1.4 Outline of the Dissertation

The dissertation is organized as follows.

Chapter 2 describes the method for quantifying the uncertainty of past TFR estimates and improving TFR prediction by accounting for this uncertainty. We further extend the model in Chapter 3 to make annual estimations instead of a five-year average and implement

this methodology in the **bayesTFR** package. Chapter 4 focuses on the application of population projection in climate change forecasting, providing methods for probabilistic forecasting climate change, as well as the effect of the Paris Agreement. Finally, Chapter 5 summarizes this work, and offers ideas for future directions of research.

Chapter 2

ACCOUNTING FOR UNCERTAINTY ABOUT PAST VALUES IN PROBABILISTIC PROJECTIONS OF THE TOTAL FERTILITY RATE FOR MOST COUNTRIES

2.1 Introduction

Population projections or forecasts consist of forecasts of future population numbers and also of the components of population change, namely births, deaths and migration, broken down by age and sex, and possibly also by other categories such as race. They are used by governments at all levels (local, regional, state, national and international) for planning and policy decision-making, since knowing the future numbers of people is key to government policy-making. They are also used by the private sector for strategic decisions, and by researchers in the health and social sciences.

The most widely used population projections for many individual countries are produced by their national statistical agency, such as the U.S. Census Bureau in the United States [U.S. Census Bureau, 2017], as well as international organizations, including the UN Population Division or Eurostat. The United Nations publishes projections of population by age and sex, and mortality and fertility rates, as well as estimates of net migration for all countries by five-year age-groups in five-year periods to the year 2100, updated every two years in the UN's *World Population Prospects*, whose most recent edition was published in 2019 [United Nations, 2019b]. The UN's population projections are widely viewed as the gold-standard and regularly updated projections for all countries [Lutz and Samir, 2010].

Since the 1940s, population projections have in most cases been produced by a deterministic method called the cohort-component method [Cannan, 1895, Whelpton, 1928, 1936,

Preston et al., 2000]. This is based on the *demographic balancing equation*, namely

$$\text{Population}_{t+1} = \text{Population}_t + \text{Births}_t - \text{Deaths}_t + \text{Immigrants}_t - \text{Emigrants}_t,$$

where Population refers to the number at time t, and Births, Deaths, Immigration and Emigration refer to the numbers in the time interval from time t to time $t+1$. The cohort-component method uses an age-structured version of this, of which a simple form is

$$\begin{aligned} \text{Population}_{a+1,t+1} &= \text{Population}_{a,t} \times \text{Survival Rate}_{a,t} + \text{Net Migration}_{a,t}, \text{ for } a \geq 0 \\ \text{Population}_{0,t+1} &= \sum_a \text{Women}_{a,t} \times \text{Fertility Rate}_{a,t}, \end{aligned}$$

where the first subscript (a for the top equation and 0 for the second) refers to age, and net migration is equal to the number of immigrants minus the number of emigrants. For a full treatment of the method, see [Preston et al., 2000].

This method is simple to implement, but it requires assumptions about future fertility, mortality and migration rates by age and sex. These have typically been produced subjectively by experts, either in-house experts working at the agency producing the projections, or a panel of outside experts assembled by the agency. Uncertainty has commonly been communicated by scenarios; for example the UN traditionally published High, Medium and Low variants, in which the total fertility rates (TFR, definition can be found in Section 2.2.1) for all countries and all future periods were increased or decreased by half a child per woman. This deterministic approach has been criticized on the grounds that it has no probabilistic basis, and that it can give implausible results over multiple projection periods [Keyfitz, 1981, Stoto, 1983, Lee and Tuljapurkar, 1994]; for a review and summary of this literature see the National Research Council report on the topic [National Research Council and Committee on Population and others, 2000].

Many methods for probabilistic forecasting of future fertility rates have been proposed, including those of [Lee, 1993, Alders et al., 2007, Alho et al., 2006, 2008, Booth et al., 2009],

in each case either for individual countries or groups of countries, typically in the developed world. However, these methods cannot be easily applied to the UN's task of producing forecasts for all countries. No fewer than 20 major methods, with 162 variants altogether, were identified and evaluated by [Bohk-Ewald et al., 2018], although not all of these were probabilistic. They found that only three probabilistic methods outperformed the simplest method of assuming that future fertility rates would be the same in the future as in the past, namely those of [Myrskylä et al., 2013, Schmertmann et al., 2014, Ševčíková et al., 2016]. The latter method is based on the probabilistic methods for projecting the total fertility rate used by the UN.

In 2015, the UN adopted a different method for their official population projections for all countries [United Nations, 2015b]. This method was probabilistic and statistically-based, replacing the previous deterministic method, thus responding to the critiques. The UN used Bayesian hierarchical models to produce probabilistic projections of the total fertility rate [Alkema et al., 2011, Raftery et al., 2014a, Fosdick and Raftery, 2014, Ševčíková et al., 2011], and life expectancy [Raftery et al., 2013, 2014b]. The UN then simulated trajectories from the resulting predictive distributions, and translated each trajectory into age-specific fertility and mortality rates. These in turn were input into the cohort-component method to yield many possible future population trajectories of all countries [Ševčíková et al., 2016, Ševčíková and Raftery, 2016]. This method indicated that world population was likely to be higher than had previously been thought, reaching 11.2 billion (95% prediction interval 9.5 to 13.2 billion) in 2100, from 7.4 billion now [Gerland et al., 2014, United Nations, 2017]. The main reason for this is that fertility in high-fertility countries, many of them in Sub-Saharan Africa, has been declining more slowly than experts had expected, and the statistical approach took this into account more fully than the expert-based assumptions.

The method discussed above takes UN estimates of past and present TFR as the source of data, which is also the standard in other studies, such as [Murray et al., 2018]. Although the new UN method takes account of uncertainty more systematically than previous methods, there are still sources of uncertainty that it does not account for. The Bayesian hierarchical

model used by the UN is conditional on estimates of present and past population, as well as fertility and mortality rates. In countries with long-established high quality vital registration systems, and hence accurate counts of births and deaths, this is not a large source of uncertainty; this is the case for 80 of the world's 201 countries, which have at least 30 years of Vital Registration Records according to the World Fertility Database 2015 Version [United Nations, 2015a]. However, the remaining 120 or so countries do not have longstanding high quality vital registration systems, and their fertility and mortality rates are typically estimated from surveys that can be subject to poor coverage in time and space, biases and measurement error. For example, the Demographic and Health Surveys (DHS) are one of the most important and reliable sources of data on fertility rates in countries without good long-term vital registration data [Federal Republic of Nigeria National Population Commission and ICF Macro, 2009], but they have suffered from large underestimates of TFR in some countries in Sub-Saharan Africa [Schoumaker, 2010, 2011, 2014, Pullum et al., 2013].

Thus the uncertainty of the estimated present and past vital rates and population numbers for certain countries are not negligible, and in the existing framework, this uncertainty is not accounted for in the projections. This can cause uncertainty in the projections to be underestimated [Abel et al., 2016]. Demographers have developed methods for correcting estimates of TFR for specific forms of bias, such as recall errors, developing indirect estimation methods for this purpose [Brass, 1964, 2015]. Bias and uncertainty of past and present estimates were modeled by [Alkema et al., 2012], using multiple data quality indicators, such as the source of the data, the estimation method (e.g. direct or indirect), and recall time for retrospective birth history surveys. But these methods do not account for the uncertainty in total fertility rate projections, and consequently population projections, that is due to uncertainty about past and present values.

In this chapter we extend the UN probabilistic projection method to account for uncertainty about past and present total fertility rates, which may be the most important remaining unaccounted source of uncertainty. This is made possible by the recent publication of a new dataset by the UN Population Division that contains not just estimates of past

and present fertility rates for all countries, but also the data from the primary data sources on which the estimates are based, including censuses, vital registration systems, partial and sample vital systems, international surveys such as the DHS and the Multiple Indicator Cluster Surveys, or MICS [UNICEF, 2016], and national, regional and local surveys [United Nations, 2015a]. We do this by developing a new Bayesian hierarchical model that extends the UN model to account for bias and measurement error in the different information sources.

The chapter is organized as follows. The data and proposed methodology are described in Section 2.2. In Section 2.3 we report the method's performance using out-of-sample predictive validation. We then provide more detail in Section 2.4, which is a case study of how the method works for Nigeria, which is one of the most important countries for uncertainty about future world population, because it is the most populous country in Africa, has high fertility, and does not have a long-established high-quality vital registration system. We conclude with a discussion in Section 2.5.

2.2 Method

2.2.1 Notation

We restrict our attention to estimation of the TFR of each country. The TFR is a period measure. It is defined as the number of children a woman would bear if she survived to the end of the reproductive interval and at each age she experienced the age-specific fertility rates prevalent in the period to which it refers. It is defined in units of children per woman.

We use the symbol y to denote TFR estimates from different data sources and the symbol f to denote the true (unobserved) TFR. Although the UN Population Division's estimates of past TFR values do contain error, we assume that they are unbiased, in the sense that the errors do not tend to be systematically in one direction or the other; for discussion of this assumption see [Alkema et al., 2012]. These official UN estimates of past TFR values will be denoted by u. All of these parameters will be indexed by country c and time t. Data from different sources y are also indexed by their source, denoted by s. Here, by source

we mean the type of data involved, for the given country. For example, one source would be the direct estimates of TFR from the Census, for Nigeria. The bias and measurement error variance of these estimates are denoted by δ and ρ^2, respectively. The quantities of interest are the unknown past, present and future TFR, f. We estimate past TFR for the time period $[t_0, t_1]$, while prediction will be for the period $[t_1, t_2]$. In practice in this chapter, $t_0 = 1950$, $t_1 = 2015$ and $t_2 = 2100$.

The three-phase Bayesian hierarchical model of [Alkema et al., 2011] will be used to model the total fertility rates. For describing the Bayesian hierarchical model, the vector of five country-specific parameters controlling the evolution of total fertility rates of country c is denoted by θ_c, and the vector of global parameters is denoted by ψ.

2.2.2 Data

We use the World Fertility Data 2015 [United Nations, 2015a] from the UN Population Division for 201 countries in the world. This database is publicly available and includes estimates of TFR from surveys, censuses and sample or partial vital registration data for countries without high-quality vital registration systems. It includes data available as of November 2015 and covers the time period from 1950 to 2015. These data were used to produce the estimates of past TFR in the United Nations World Population Prospects (WPP) 2015 Revision. These estimates were in turn part of the basis for the UN's 2015 population projections for all countries.

We use TFR estimates from national and international surveys, indirect estimates and vital registration for all 201 countries to estimate the bias and variance of different data sources. We take the estimates in the WPP 2015 revision as a baseline, assuming that they are unbiased (but not that they are without error). This assumption, also used by [Alkema et al., 2012], is made because the analysts producing past estimates were often aware of sources of bias in datasets and tried to correct for them. While this assumption is not perfect, it seems reasonable to argue that WPP provides the least biased set of estimates available.

To estimate TFR from multiple sources, it is necessary to make some assumption about a baseline unbiased (although not error-free) estimate, and we have used the official UN estimate. An alternative would be to choose one of the individual information sources as unbiased; a possible candidate would be the direct estimates from the DHS surveys. However, these estimates (and estimates from any one source) are not without flaws, and some of them were of poor quality, as discussed by [Alkema et al., 2012] and others. Another possibility would be to use an average of the available estimates, but similar comments apply. The UN analysts used all the available data sources, while being aware of problems with them, and adjusted or downweighted them accordingly in developing their own estimates. The information used by the UN analysts is currently often available only implicitly, through the official UN estimates they produced, and so these estimates seem likely to be of higher quality than the individual estimates.

In the 2015 revision of the WPP, the UN estimated the five-year average TFR, $u_{c,t}$, for country c in time period $(t, t+5)$, for each five-year period from 1950 to 2015. The outcome in each five-year period $(t, t+5)$ is an estimate of the average TFR between July 1 of year t and June 30 of year $t+5$, and so is centered close to January 1 of year $t+3$. This chapter constructs trajectories and estimations in five-year intervals, and projects TFR up to year 2100 probabilistically according to these estimated trajectories of the past.

2.2.3 Model

Three Phase Bayesian Hierarchical Model Our methodology builds on that of ([Alkema et al., 2011, pg.818-824], [Raftery et al., 2014a, pg.64-65]) for fertility transition phase and post-fertility transition phase respectively, and was implemented by [Ševčíková et al., 2011] . This divides the evolution of TFR in a country into three phases: pre-transition, transition and post-transition, as illustrated in Figure 2.1.

During the fertility transition or decline phase (Phase II), the total fertility rate is modeled

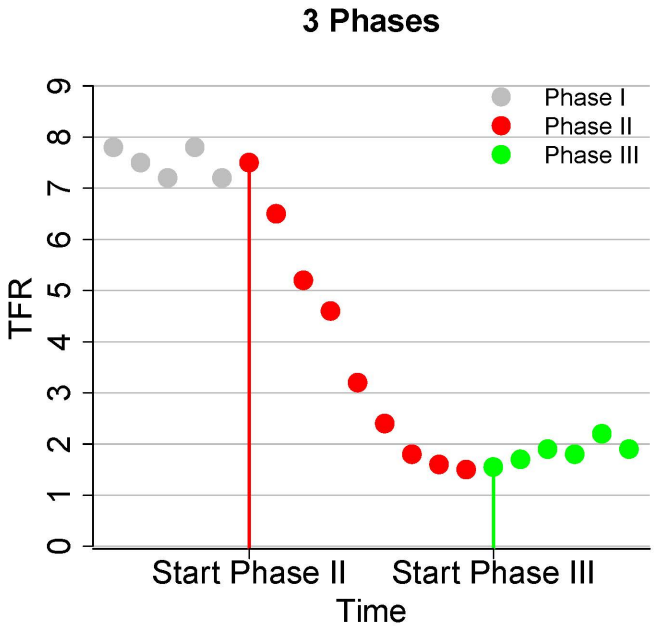


Figure 2.1: Illustration of the three phases in the typical evolution of fertility in a country: pre-transition high fertility (Phase I; grey), transition from high to low fertility — (Phase II; red), and post-transition fertility fluctuations and recovery (Phase III; green).

as a random walk with negative drift, namely

$$f_{c,t} = f_{c,t-5} - g(f_{c,t-5}|\theta_c) + \varepsilon_{c,t}, \tag{2.1}$$

where $g(\cdot|\theta_c)$ is the expected five-year decrement in the TFR over the next period, modeled by a double logistic function governed by the country-specific parameter vector $\theta_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$, and $\varepsilon_{c,t}$ is random noise (independent $N(0, \sigma^2_{c,t})$) around the expected

decrement. The double logistic curve we are using is given by

$$g(f_{c,t}|\theta_c) = \frac{-d_c}{1+\exp\left(-2\frac{\ln(9)}{\Delta_{c1}}(f_{c,t}-\sum_i \Delta_{ci}+0.5\Delta_{c1})\right)} + \frac{d_c}{1+\exp\left(-2\frac{\ln(9)}{\Delta_{c3}}(f_{c,t}-0.5\Delta_{c3}-\Delta_{c4})\right)}.$$

Parameters all have a clear interpretation and could be illustrated in Figure 2.2.

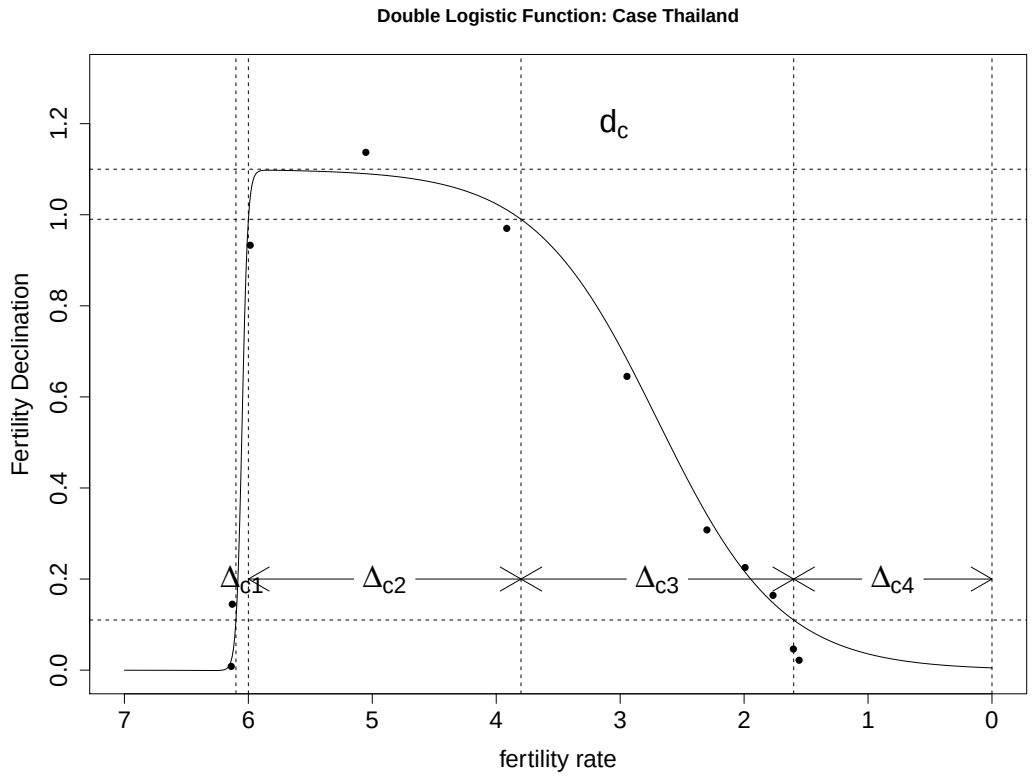


Figure 2.2: Illustration of the parameters and the corresponding double logistic curve. d_c controls the maximum of five year decrements, and Δ_{ci}s control the shape of the double logistic curve. Specifically, if we look at the fertility rates in a reversing order, Δ_{c1} matches the fertility change from the double logistic curve increasing from 10% of the maximum to 90% of the maximum. Δ_{c2} is difference of two fertility rates where all fertility rates in the middle is over 90% of d_c. Δ_{c3} is the fertility change from 90% to 10% of maximum and Δ_{c4} is for the remaining.

During the post-transition phase (Phase III), the total fertility rate is modeled by a Bayesian Hierarchical Autoregressive Model as:

$$f_{c,t} = \mu_c + \rho_c(f_{c,t-5} - \mu_c) + \varepsilon_{c,t}, \tag{2.2}$$

where μ_c is the mean of the TFR for country c in Phase III, and $\varepsilon_{c,t}$ is the random noise similar to that in phase II. As in our model, Phase III should be the ultimate Phase for all countries, μ_c can also be viewed as the long-term mean of the TFR for country c.

Since all or almost all countries have already started the fertility transition, modeling the TFR during the pre-fertility transition Phase I was not necessary for projection purposes in previous work. However, for constructing probabilistic estimation of past TFR from 1950 to 2015, we do need to model the Phase I data. They are modeled by a random walk model from year 1950 to the start of fertility transition as:

$$f_{c,t} = f_{c,t-5} + \varepsilon_{c,t}\,. \tag{2.3}$$

and the initial TFR $f_{c,1950}$ for countries in Phase I is modeled by a non-informative uniform prior $U(5, 8.5)$.

The country-specific parameters in all three phases, (θ_c, μ_c), follow a world distribution, which is governed by world parameters ψ, and these in turn have a prior distribution. These distributions are illustrated in the Appendix of chapter 2. The start and end of the fertility transition (Phase II) are defined based on the UN estimates $u_{c,t}$, by rules given in [Alkema et al., 2011].

Model of Imperfect Data The TFR estimates from different data sources $y_{c,t,s}$ are modeled based on the unobserved true value $f_{c,t}$. Building on [Alkema et al., 2012], we distinguish between the bias and measurement error variance in our model. The estimated TFR values

are modeled by a conditional normal distribution as:

$$y_{c,t,s}|f_{c,t} \sim \mathcal{N}_+(f_{c,t} + \delta_{c,s}, \rho^2_{c,s})\,, \tag{2.4}$$

$$\mathbb{E}[\delta_{c,s}] = \boldsymbol{x}^T_{c,s}\boldsymbol{\beta}\,, \tag{2.5}$$

$$\mathbb{E}[\rho_{c,s}] = \boldsymbol{X}^T_{c,s}\boldsymbol{\gamma}\,. \tag{2.6}$$

The bias and measurement error standard deviations, $\boldsymbol{\delta}_{c,s}$ and $\boldsymbol{\rho}_{c,s}$, are estimated using data quality indicators, denoted by $\boldsymbol{x}_{c,s}$ (a vector for each country source pair). In practice, we use the same covariates as data quality indicators (source and estimating methods of the data) for estimating bias and measurement error standard deviation, and thus we are using the same notation $\boldsymbol{x}_{c,s}$ here. The estimation process is described in the following sections.

Complete Model Layout We combine the three-phase Bayesian hierarchical model and imperfect data model into a four-level Bayesian hierarchical model with an additional level for the data sources. Estimation and prediction are then equivalent to obtaining the posterior distribution of the unknown TFR values $f_{c,t}$ in the estimation period $[t_0, t_1]$ and the prediction period $[t_1, t_2]$, based on the observed TFR estimates from different data sources.

The observed estimates of TFR can be measured for any time between t_0 and t_1. However, we seek estimates of the average over five-year periods. We approximate the true TFR at any time by assuming that the TFR evolves linearly between the centers of any two successive five-year intervals. This is a reasonable assumption because most demographic quantities, including the TFR, typically evolve fairly smoothly over time. Specifically, for any $t \in [t_\ell, t_\ell + 5]$, where t_ℓ and $t_\ell + 5$ are the centers of two successive five-year periods, we assume that

$$f_{c,t} = \frac{1}{5}[(t_{\ell+5} - t)f_{c,t_\ell} + (t - t_\ell)f_{c,t_{\ell+5}}]. \tag{2.7}$$

We also used the same rule as in Alkema et al. [2011] for determining when Phase II and

III kick in. Specifically, we identify the candidate period for the start period of the fertility transition as the most recent period with a local maximum within 0.5 child of the global maximum. If that local maximum is above 5, the corresponding period is defined as the start period of the fertility trainsition, and if not, we consider that the start of Phase II for this country is unobserved. The countries that have entered Phase III are defined as the countries in which two subsequent five-year increases below a TFR of 2 children have been observed.

Thus, the overall model is specified as follows:

$$
\begin{aligned}
\text{Level 1: } & y_{c,t,s}|f_{c,t} \sim \mathcal{N}(f_{c,t} + \delta_{c,s}, \rho^2_{c,s})\,, \\
& \mathbb{E}[\delta_{c,s}] = \boldsymbol{x}^T_{c,s}\boldsymbol{\beta}\,, \\
& \mathbb{E}[\rho_{c,s}] = \boldsymbol{x}^T_{c,s}\boldsymbol{\gamma}\,, \\
& f_{c,t} = \frac{1}{5}[(t_{\ell+5} - t)f_{c,t_\ell} + (t - t_\ell)f_{c,t_{\ell+5}}] \text{ for } t \in [t_\ell, t_{\ell+5}]\,; \\
\text{Level 2: } & \text{Phase I: } f_{c,t} = f_{c,t-5} + \varepsilon_{c,t}\,, \\
& \text{Phase II: } f_{c,t} = f_{c,t-5} - g(f_{c,t-5}|\theta_c) + \varepsilon_{c,t}\,, \\
& \text{Phase III: } f_{c,t} = \mu_c + \rho_c(f_{c,t-5} - \mu_c) + \varepsilon_{c,t}\,, \\
& \varepsilon_{c,t} \sim \mathcal{N}(0, \sigma^2_{c,t})\,; \\
\text{Level 3: } & \theta_c \sim h(\cdot|\psi)\,, \\
& \theta_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)\,, \\
& \mu_c \sim \mathcal{N}(\bar{\mu}, \sigma^2_\mu)\,, \\
& \rho_c \sim \mathcal{N}(\bar{\rho}, \sigma^2_\rho)\,; \\
\text{Level 4: } & \psi, \bar{\mu}, \sigma_\mu, \bar{\rho}, \sigma_\rho \sim \pi(\cdot)\,.
\end{aligned}
$$

Thus we model the observed TFR estimates in Level 1, conditional on the true total fertility rates. These are in turn modeled by the extant three-phase BHM in Level 2, conditional on the country-specific parameters. The country-specific parameters are then modeled con-

ditionally on the global parameters in Level 3, which have a prior distribution specified by hyperparameters (Level 4).

Here, g denotes the double logistic function, and h and π denote the conditional and unconditional distributions of the parameters of interest, respectively. The parameter vector θ_c, with five elements, controls the shape of the double logistic curve, and the hyper parameter ψ, with ten elements, specifies the mean and variance of the conditional normal distribution of some transformation of θ_c. The detailed functional form of the prior distribution $\pi(\cdot)$ and the conditional distribution $h(\cdot|\psi)$ can be found in the Supplementary Material, and are the same as specified by [Alkema et al., 2011].

Inference is based on the joint posterior distribution of $(f_{c,t}, \theta_c)$. The model is summarized graphically in Figure 2.3.

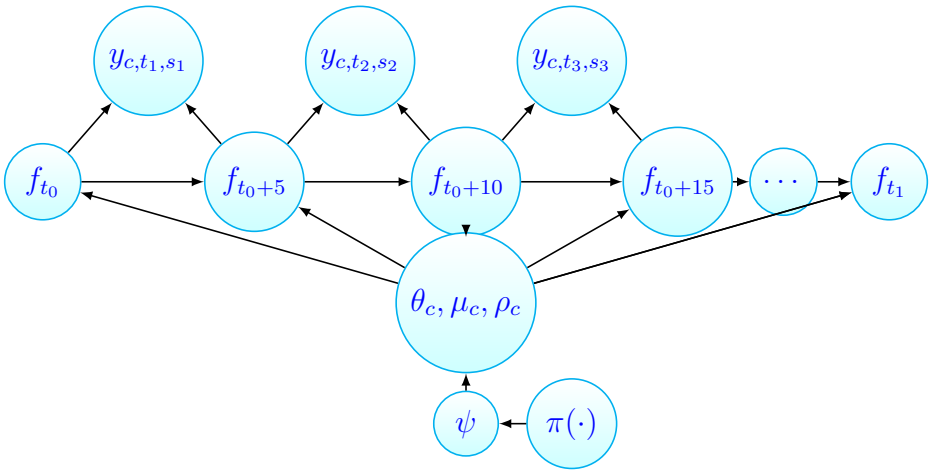


Figure 2.3: Model Specification: $y_{c,t,s}$ are the observed TFR, $f_{c,t}$ are the unknown TFR values, θ_c are the country-specific parameters and ψ are the global parameters. Detailed explanation of these parameters can be found in this section previously and the complete specification can be found also in the Section 1 of supplementary materials. Here, t_1, t_2, t_3 are representative of some arbitrary time points between $[t_0, t_0+5], [t_0+5, t_0+10], [t_0+10, t_0+15]$ respectively.

2.2.4 Estimation

Estimation of Bias and Measurement Error Variance The bias $\delta_{c,s}$, and measurement error variance, $\rho^2_{c,s}$, of the observed TFR estimates are estimated in a first stage, as input to the Bayesian hierarchical model, building on the method of [Alkema et al., 2012].

We first estimate the bias of TFR. As we discussed in Section 2.2.3, the UN estimates will be treated as unbiased but not error-free, providing a baseline reference. Then, for each observation $y_{c,t,s}$, we have

$$\mathbb{E}[y_{c,t,s} - u_{c,t}] = f_{c,t} + \delta_{c,s} - f_{c,t} = \delta_{c,s}.$$

Thus we can use the difference between the observed TFR values and the UN estimates, $(y_{c,t,s} - u_{c,t})$, as samples for our estimation of the bias and measurement error variance of each source. The parameters $\boldsymbol{\beta}$ are estimated by linear regression on data quality indicators $\boldsymbol{x}_{c,s}$, as in equation (2.5). The estimated biases $\hat{\delta}_{c,s}$ are then equal to the fitted values $\boldsymbol{x}_{c,s}\hat{\beta}$.

In this study, we used two data quality indicators: data source, and estimating method. Other potential data quality indicators were available, but we found that using them would not have improved the performance of the method.

We estimate the source-specific measurement error variance of the TFR estimates by regression on the data quality covariates $\boldsymbol{x}_{c,s}$ of the plug-in estimate $\rho_{c,s} = \sqrt{\frac{\pi}{2}}\mathbb{E}|z_{c,t,s} - \hat{\delta}_{c,s}|$, where $z_{c,t,s}$ is defined as the estimation error of the observed TFR estimates $y_{c,t,s}$, which is the difference between the TFR estimates $y_{c,t,s}$ and the true TFR $f_{c,t}$, namely $(y_{c,t,s} - f_{c,t})$. Since the true TFR is unknown, the unobserved true values $f_{c,t}$ are replaced by the UN estimates of TFR $(u_{c,t})$ (taken to be unbiased), which means that in practice, $z_{c,t,s}$ is replaced by $\tilde{z}_{c,t,s} = y_{c,t,s} - u_{c,t}$.

Estimation of the Complete Model: Given the estimated bias $\hat{\delta}_{c,s}$ and measurement error variance $\hat{\rho}^2_{c,s}$, we estimate the Bayesian hierarchical model for TFR using a purpose-built Markov Chain Monte Carlo (MCMC) algorithm coded in R. The roughly 3,600 parameters

and unknown TFR values are updated one at a time, using Gibbs steps, Metropolis-Hastings steps or slice sampling [Neal, 2003] for each parameter as appropriate. We monitored convergence by inspecting trace plots and using standard convergence diagnostics [Gelman and Rubin, 1992, Raftery and Lewis, 1996].

We thinned enough for the thinned sample to be roughly independent. In practice, for the final results we ran 3 chains, each of length 12,000 iterations with a burn-in of 2,000, and we thinned the resulting chains by 10, to obtain a final, approximately independent sample of size 3,000 from the posterior distribution. More information about the convergence diagnostics used is provided in the Supplementary Information.

2.2.5 Prediction of Future TFR

Unlike the projection process developed by [Alkema et al., 2011] and used by the UN, we have probabilistic rather than point TFR estimates of past rates over the time period $[t_0, t_1]$. Thus, instead of just sampling from the posterior trajectories of estimated country-specific parameters, we also generate posterior trajectories of past TFR values.

We proceed by repeating the following process many times. We first select a joint sample of model parameters and past and present TFR for all countries from the posterior distribution. Then, given the sampled model parameters and past and present TFR values, we simulate a trajectory of future TFR values, from 2015 to 2100 using the model specified by (2.1) and (2.2). This yields a sample from the joint posterior predictive distribution of future TFR in all countries and time periods considered, taking account of uncertainty about past values.

Our method also differs slightly from the extant method in how the end of the fertility transition. at which the model shifts from that for Phase II to that for Phase III, is determined. The current UN method uses deterministic rules based on the UN estimates [Alkema et al., 2011], and does not account for uncertainty about when the fertility transition ended. In our method, we retain the deterministic rules, but apply them separately to each sampled trajectory of past TFR values. Thus our method takes account of uncertainty about when

the fertility transition ended in a particular country, and hence which phase the country is in at the end of the estimation period.

2.3 Validation

We assess the predictive performance of our model using out-of-sample predictive validation, used for probabilistic forecasts, for example, by [Raftery et al., 2005]. We include all countries and regions in our validation exercise.

2.3.1 Study Design

The data we have cover the period from 1950 to 2015. We split this into the estimation period, $[t_0 = 1950, t_1 = 2005]$, and the prediction period, $[t_1 = 2005, t_2 = 2015]$. The inputs to our method consist of all TFR estimates from different sources referring to the estimation period.

For the UN estimates used as a reference, we take the values published in the WPP 2008 revision [United Nations, 2008]. The UN estimates of the past have been refined since then as more data have become available, but we deliberately do not take advantage of this in our estimation. This makes our validation exercise more analogous to the real prediction task at hand, for which we are using UN estimates in the WPP 2015 revision of past TFR values up to 2015 to predict values past 2015. It can be expected that these estimates of TFR values up to 2015 will become more accurate in the future as data accumulate, but we are not able to take advantage of this at the present time.

We are making probabilistic projections, and so we evaluate not only the point predictions, but also the predictive intervals. Our aim is to account for an important source of uncertainty ignored by the present state-of-the-art method, so the accuracy of the prediction intervals may be even more important than that of the point predictions. If our method is working well, we would expect the current state-of-the-art intervals to have less than nominal coverage (due to miss one source of uncertainty), and our method to give coverage closer to nominal. To evaluate our method, we compare our probabilistic projections with those

produced by the UN in WPP 2015.

Our out-of-sample validation proceeds as follows.

1. Choose the subset of the original data set $\mathcal{D}$ with TFR observed before year 2005 as the training data $\mathcal{D}_{\text{train}}$. We remove those observations before 2005 for those estimates in studies that provide series of estimates ending after 2005. For example, if a study lasts for 20 years and ends in 2008, yielding TFR observations for 1988 to 2008, we remove all observations from this study even though some of the estimates are for years before 2005. This is because these data would typically not have been available in 2005. This would account for 4000 samples (out of 22000 samples in total).

2. Estimate bias and measurement error variance for all the data points with the UN estimates, $u_{c,t}$ from the WPP 2008 revision as the reference. The reason that we use the 2008 revision as the reference instead of the 2015 revision is that UN is estimating historical TFR based only on current data available at the time of the revision. Thus if we were to use WPP 2015 as the reference, we would be using what is effectively future information in the out of sample validation.

3. Draw a sample from the joint posterior distribution of model parameters and past TFR values for 1950 to 2005, using MCMC.

4. For each sampled trajectory including the unobserved past TFR values and the model parameters, determine the TFR phase of country c for each time period for this trajectory, and make probabilistic projections for the projection period $[2005, 2015]$.

2.3.2 Out of Sample Validation Results

We produce results for all 201 countries using our method. For comparison, we also produce results using the method of [Alkema et al., 2011], which underlies the current UN methodology and does not take account of uncertainty about past TFR values.

We summarize the results in Table 2.1. This is based on the predictive intervals for each of the 201 countries and for both of the periods $[2005, 2010]$ and $[2010, 2015]$, so that each entry in Table 2.1 is an average over $201 \times 2 = 402$ values. For each TFR value to be predicted, we take the predictive median as the point estimate, and we compute the quantile-based 80% and 95% prediction intervals. The table shows the mean absolute error (MAE) of the point estimates (the smaller the better), and the coverage of the prediction intervals (the closer to the nominal value the better). If we split by periods and proportion of left-out UN estimates that fall above or below their 80% and 95% projection intervals, the performance of the new method is summarized in Table 2.2. (Note that the out of sample performance is different from that in [Alkema et al., 2011] because of the different forecasting horizons and the updated versions of the WPP data set used here for validation.)

Table 2.1: Mean Absolute Error and Coverage of Out of Sample TFR Point and Interval Predictions for Original Method [Alkema et al., 2011], and Proposed Method.

	Original Method	Proposed Method
Mean Absolute Error	0.250	0.242
Coverage of 80% interval	64.0%	77.8%
Coverage of 95% interval	80.1%	92.1%

Table 2.2: Proportion of Left-Out UN Estimates that Fall Above or Below their 95% and 80% Projection Intervals Split by Period.

	95% PI			80% PI		
	Below	Coverage	Above	Below	Coverage	Above
2005-2010	1.0%	91.7%	7.3%	1.0%	78.1%	20.9%
2010-2015	0.0%	92.7%	7.3%	4.7%	77.5%	17.8%

The proposed method improves the point predictions over the current method, as measured by the MAE, by about 3%, but is not statistically significant. It improves the coverage of the prediction intervals much more substantially. Under the current method, the coverage of the prediction intervals is somewhat below the nominal level, suggesting that some of the

uncertainty is being missed. Under the proposed method, the coverage of the prediction intervals is much closer to the nominal level, suggesting that the new method is capturing most of the missed uncertainty by taking account of uncertainty in past TFR values.

The overall coverage rate of the method of [Alkema et al., 2011] is worse than that reported in the original paper mainly because of the change of the historical data between WPP versions. For example, in WPP 2008, the estimate of TFR for Nigeria in 2000-2005 was 5.67, but that value changed to 6.05 in WPP version 2015; the latter is presumably more accurate because it is based on additional information available in 2015 but not in 2008. The model of [Alkema et al., 2011] performed extremely well for forecasting the UN estimates of the same version (WPP 2008), but not as well for forecasting the updated estimate of TFR in 2000-2005 that was made on the basis of data available in 2015.

For illustration, results of the out-of-sample validation exercise are shown in Figure 2.4 for Argentina, Botswana, Nigeria and the United States. Of these, only the United States has had a high-quality vital registration system for the entire period, while Argentina has a vital registration that was of lower quality in the early years, and the other two countries have no comprehensive vital registration systems, relying instead on censuses and periodic surveys.

The posterior intervals of past TFR values are very narrow for the United States, reflecting the high quality vital registration data available for the entire period, while for Argentina they are somewhat wider. For both Botswana and Nigeria the intervals are far wider, reflecting the much lower quality of the available data. For the earlier years, from the 1950s to the 1970s, the intervals for Botswana and Nigeria are especially wide, reflecting the sparsity of the data for these decades. The predictive distributions cover the observations in all cases, although in some cases they lie towards the edge of the intervals, as expected if the intervals are well calibrated.

Additional out of sample validation results for comparing our method with that of [Alkema et al., 2011] are available in the Supplementary Information. We find that the coverage rate remained close to the nominal levels for other out of sample forecasting peri-

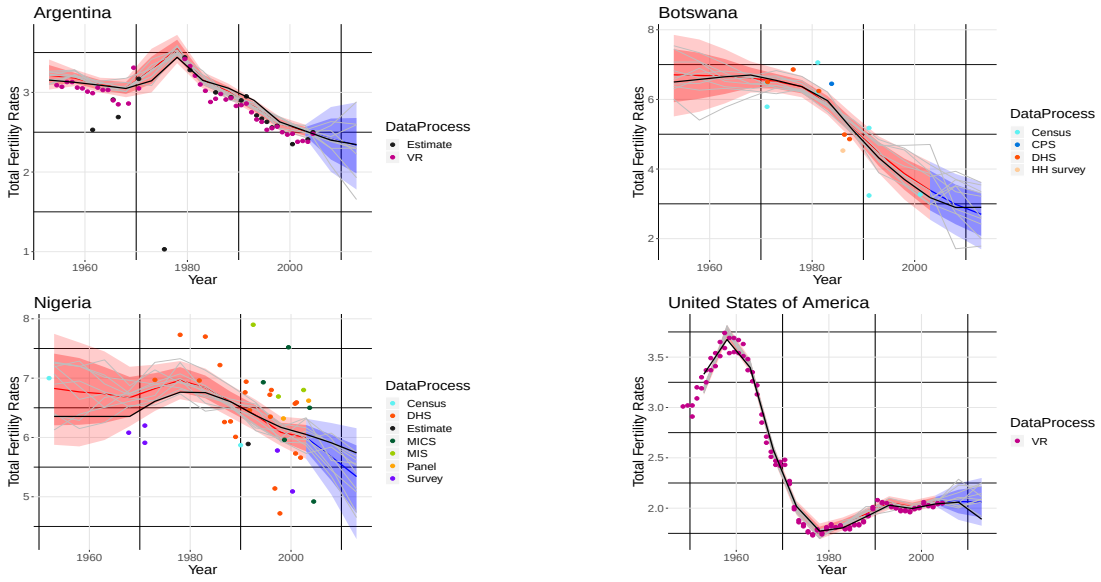


Figure 2.4: Out of Sample Validation Results for Argentina, Botswana, Nigeria and the United States. Estimates of past TFR values for $[1950, 2005]$ are shown by dots, with different sources corresponding to different colors, as described in the side captions. The UN estimates are shown in black. The posterior distributions of past values for $[1950, 2005]$ are shown in orange, with the posterior median as the solid line, the posterior 80% interval as the dark shaded region, and the posterior 95% intervals as the light shaded region. The corresponding posterior predictive distributions for $[2005, 2015]$ based on data up to 2005, are shown in blue.

ods.

2.3.3 Robustness Across WPP *Versions*

To illustrate the extent to which the new method is consistent across all WPP versions, we consider the same validation study as in Section 2.3.1, but evaluating our forecasts with reference to different WPP versions instead of only WPP 2015 in Section 2.3.2.

In order to compare with the WPP 2008 revision and the 2010 revision, we could not do validation on the 2010-2015 period. Thus, we train the model with the WPP 2008 Revision, and for the new method, we use only data before 2005. Then we make forecasts of TFR for 2005-2010 with the model trained, and measure the coverage rate of the forecast of TFR for 2005-2010 using the data from the WPP 2008, 2010, 2012, 2015 and 2017 revisions. The

results are summarized in Table 2.3.

Table 2.3: Coverage rate of 95% and 80% of prediction intervals based on model trained with data before 2005 and WPP 2008 Revision.

		Original Method		Proposed Method	
WPP Version	Testing Period	95% PI	80% PI	95% PI	80% PI
WPP 2008	2005-2010	98.5%	92.9%	98.5%	88.0%
WPP 2010	2005-2010	89.8%	73.0%	97.4%	83.3%
WPP 2012	2005-2010	79.9%	62.9%	95.4%	79.1%
WPP 2015	2005-2010	78.6%	61.3%	92.1%	77.7%
WPP 2017	2005-2010	78.4%	62.4%	92.4%	78.1%

The later the version of WPP, the more information the TFR estimates are based on, and so the more accurate they should be. Therefore, we are more interested in the coverage rate based on later revisions (especially the 2015 and 2017 revisions) than the earlier revisions. We found that our intervals overcovered the estimates from the earlier revisions, but had coverage close to nominal for the more accurate 2012 and later revisions, while the original method had coverage substantially below nominal for the later revisions.

2.4 Case Study: TFR Estimation and Projection For Nigeria

We illustrate the method by producing probabilistic forecasts of the TFR of Nigeria from 2015 to 2100, using data available up to 2015. As we have discussed, the method first estimates the bias and measurement error variance of the different data sources. It then estimates the uncertainty about past TFR values, and takes this uncertainty into account when making probabilistic projections.

2.4.1 Estimation of Bias and Measurement Error Variance of Different Data Sources

From 1950 to 2015, according to the UN's WPP 2015 revision, the TFR in Nigeria reached its peak around 1980 at about 6.7 children per woman. It then declined slowly, reaching about 5.7 in 2015. However, the data on which these estimates are based are surprisingly noisy, as

can be seen in Figure 2.5.

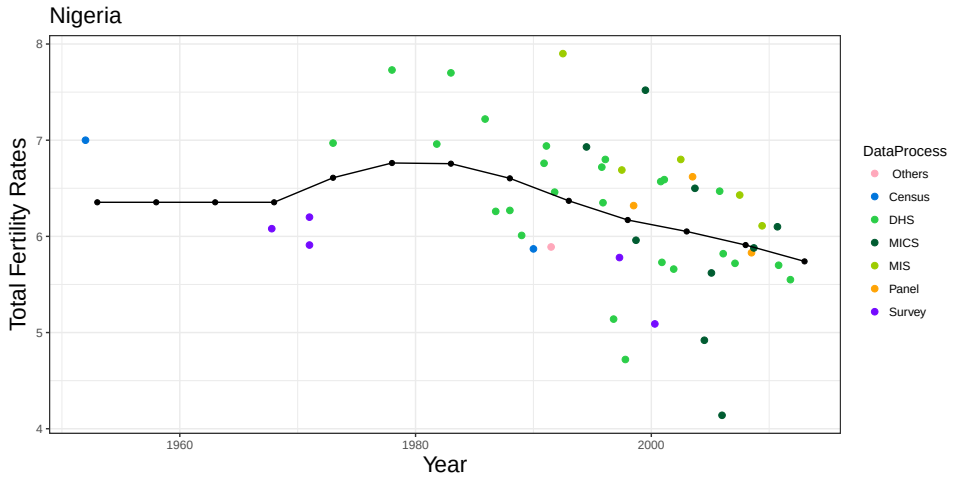


Figure 2.5: Nigeria TFR Estimates, 1950–2015.

These data come from several sources, including national censuses, which are comprehensive but are sparse in time and have issues of coverage. The other sources are mostly surveys, including the internationally organized Demographic and Health Surveys (DHS), the Multiple Indicator Cluster Surveys (MICS) run by UNICEF, and the Malaria Indicators Survey, or MIS, also run by DHS. There are also several occasional national cross-sectional and panel surveys. Some of the surveys, notably DHS and MICS, collect birth histories, which allow one survey to generate estimates for several past years, in some cases using indirect methods.

The black line represents the UN estimate of Nigeria's TFR in 1950 to 2015 from WPP 2015. We could also see the changes in the UN estimates of Nigeria's past TFR, from WPP 2008 to WPP 2017. We show those changes in Table 2.4.

We first estimate the bias and measurement error variance of the different data sources using the approach outlined in Section 2.2.3. From each observed TFR estimate $y_{c,t,s}$ we subtract the corresponding UN TFR estimate to obtain an estimate of the bias for that source, country and time, namely $z_{c,t,s} = y_{c,t,s} - u_{c,t}$. As data quality covariates, $\boldsymbol{x}_{c,s}$, we use

Table 2.4: WPP estimates of Nigeria's TFR from version 2008 to 2017. The most recent TFR estimates are changing significantly. For example, in WPP 2008, the TFR for 2005-2010 is estimated as 5.32, but in later versions, the TFR is estimated as 5.91, which is about 0.6 more children per woman. The values for 2000-2005 and 2005-2010, on which we focus, are bolded for emphasis.

year	WPP 2008	WPP 2010	WPP 2012	WPP 2015	WPP 2017
1950-1955	6.55	6.35	6.35	6.35	6.35
1955-1960	6.55	6.35	6.35	6.35	6.35
1960-1965	6.55	6.35	6.35	6.35	6.35
1965-1970	6.55	6.35	6.35	6.35	6.35
1970-1975	6.72	6.61	6.61	6.61	6.61
1975-1980	6.89	6.76	6.76	6.76	6.76
1980-1985	6.93	6.76	6.76	6.76	6.76
1985-1990	6.76	6.56	6.60	6.60	6.60
1990-1995	6.44	6.23	6.37	6.37	6.37
1995-2000	6.05	5.99	6.17	6.17	6.17
2000-2005	**5.67**	**5.79**	**6.05**	**6.05**	**6.05**
2005-2010	**5.32**	**5.61**	**6.00**	**5.91**	**5.91**
2010-2015	NA	NA	6.00	5.74	5.74

the source of the data and whether the estimate is direct or indirect. We then estimate the bias $\delta_{c,s}$ for country c and data source s as the fitted value from a regression of the $z_{c,t,s}$ on the data quality covariates $\boldsymbol{x}_{c,s}$, as in [Alkema et al., 2012].

The UN TFR estimates are for five-year periods, and we treat them as referring to the middle of the period. Thus, for example, we treat estimates for 2010–2015 as referring to the beginning of 2013. An observed TFR estimate can refer to any year between 1950 and 2015, and we use the linear interpolation of the two UN estimates closest to the time to which it refers as the corresponding UN estimate, $u_{c,t}$.

Similarly, after we get the fitted value of the bias estimates $\hat{\delta}_{c,s}$, we obtain the measurement error standard deviation estimates by regressing $|z_{c,t,s} - \hat{\delta}_{c,s}|$ on the same data quality covariates. The fitted biases and measurement error standard deviations are summarized in Table 2.5.

We can see from Table 2.5 that the direct estimates from the DHS are the highest quality

Table 2.5: Estimates of Bias and Measurement Error Variance for All Combinations of Source and Estimate Types, for Nigeria. Survey-NR are different Nigeria nationwide surveys and Survey represents other survey estimates. Under estimate type, D represents direct estimates, C cohort estimates and I indirect estimates. Here $\mu(\delta)$ and $\sigma(\delta)$ are the sample bias and measurement error standard deviations; when a hat is added they represent the estimates from the models. Estimated root mean squared errors are summarized in the column RMSE ($= \sqrt{\hat{\delta}^2 + \hat{\rho}^2}$). The number of observations for each combination is shown in the column n.

	Source	Estimate Type	$\mu(\delta)$	$\sigma(\delta)$	$\hat{\delta}$	$\hat{\sigma}(\delta) = \hat{\rho}$	RMSE	n
1	DHS	D	0.04	0.48	0.11	0.38	**0.40**	28
2	DHS	C	-0.26	0.51	-0.48	0.46	*0.66*	10
3	Census	D	0.00	0.91	-0.43	0.50	**0.66**	2
4	Census	C	-1.46	0.43	-1.02	0.58	1.17	2
5	MICS	D	-1.10	1.03	-0.33	0.81	**0.87**	2
6	MICS	C	-0.79	0.18	-0.92	0.89	1.28	2
7	MICS	I	0.29	1.64	0.20	1.35	*1.36*	15
8	MIS	D	0.70	0.48	0.22	0.56	**0.60**	5
9	MIS	I	0.68	1.37	0.75	1.09	*1.32*	30
10	Survey	D	-0.50	0.58	-0.47	0.42	**0.63**	4
11	Survey	C	-1.18	0.95	-1.06	0.49	1.17	8
12	Survey	I	0.14	0.98	0.06	0.95	*0.95*	15
13	Survey-NR	D	-0.40	0.18	-0.60	0.21	**0.64**	3
14	Survey-NR	C	-1.48	0.18	-1.18	0.29	1.22	2

estimates as measured by estimated root mean squared error (equal to $\sqrt{\hat{\delta}^2 + \hat{\rho}^2}$). Direct estimates generally have smaller variances than indirect estimates. Figure 2.6 plots the fitted biases and measurement error standard deviations against the observed ones; the model fit seems reasonably good.

2.4.2 Estimation of Past and Projection of Future TFR

The fertility transition, or Phase II, started in 1980 in Nigeria, according to the definition of [Alkema et al., 2011]. We initialize the MCMC algorithm with a warm start, simulating the starting values for the global parameters ψ and the country-level parameters θ_c from

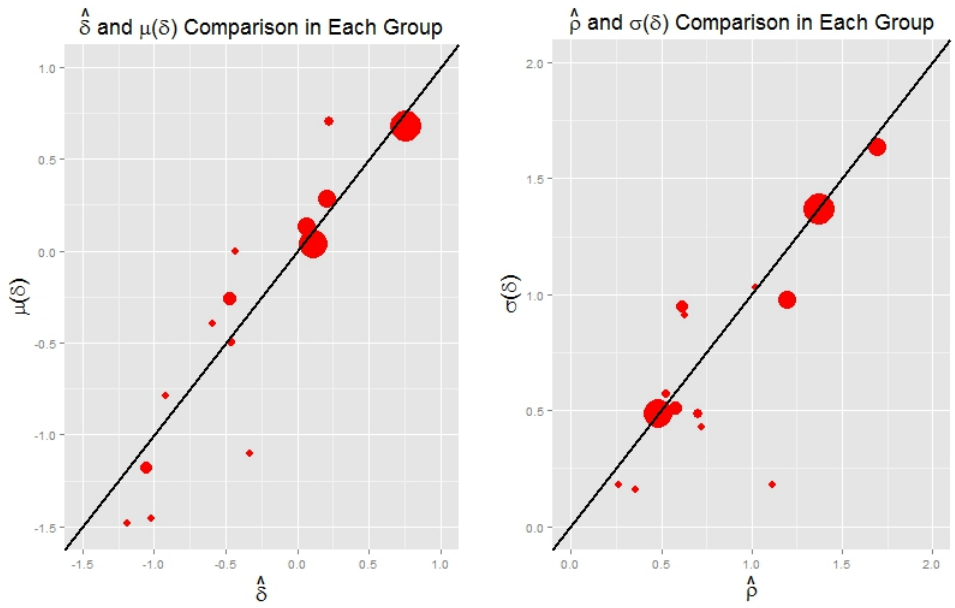


Figure 2.6: Bias and Variance Estimates for Nigeria: Fitted against Observed. The size of the dots represents the number of observations. Most large dots are along the diagonal line.

their posterior distribution from the model that does not take account of uncertainty about past TFR values [Alkema et al., 2011, Raftery et al., 2014a]. The true past fertility rates are initialized as the UN estimates.

The results are shown in Figure 2.7. These are based on data up to 2015, and can be compared with Figure 2.4(c), which is based on data up to 2005. The posterior distribution for the 2000-2005 period is tighter, because more data relevant to this period were available in 2015 than in 2005. The posterior distibution widens slightly for the past period, 2010–2015, again reflecting the relative paucity of data relevant to this period by 2015. We expect that this posterior distribution will tighten as more data relevant to 2010–2015 become available in the future.

We make projections in two steps. In the first step, we sample one trajectory from the MCMC results obtained in Section 2.4.2. Then given the sampled trajectory, the phase of the most recent year is determined by this trajectory, and then future TFR is sampled

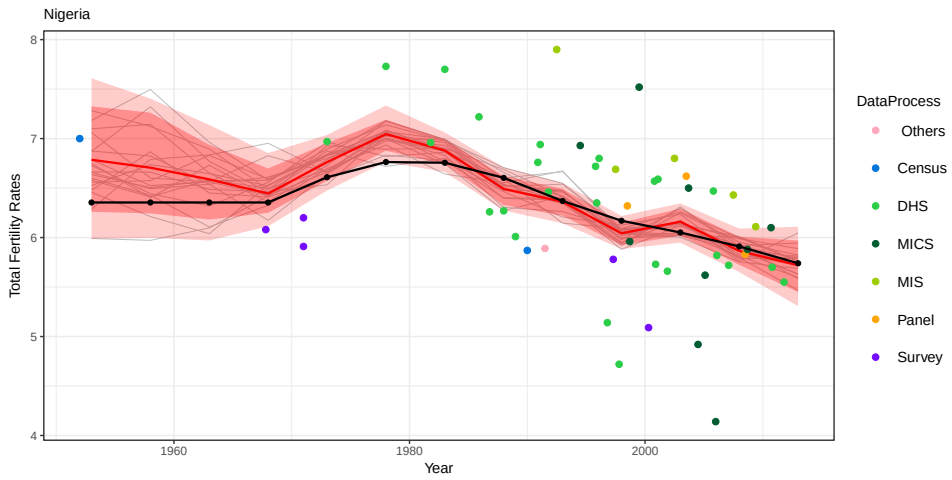


Figure 2.7: Past and Present TFR Estimates for Nigeria. Colored dots are observed TFR, red shaded areas are 95% estimation intervals and the black line is the UN TFR estimates (from WPP 2015).

according to the country-specific parameters of this trajectory. The resulting projection is shown in Figure 2.8. The black solid and dotted curves show the UN's 2015 probabilistic projection (not taking account of uncertainty about the past), while the blue line and shaded region show the projection from our method. Both project that Nigeria's TFR will likely decline, with a great deal of uncertainty about how fast this will happen. Our proposed method yields a similar predictive median to the current UN method, but somewhat wider prediction intervals. As we saw in the out-of-sample validation study, these wider intervals do incorporate an important additional source of uncertainty, and, on average, take the intervals from undercovering the truth to some extent to the nominal coverage.

2.4.3 Model Validation: Simulation Study

We now run a simulation study with input data on past TFRs chosen to resemble the Nigerian data, to see how accurately the proposed method captures past TFR values. Assume we have trained the MCMC process beforehand and have got the estimates of past TFR estimates for Nigeria as we showed in previous sections. Then for each simulation, we

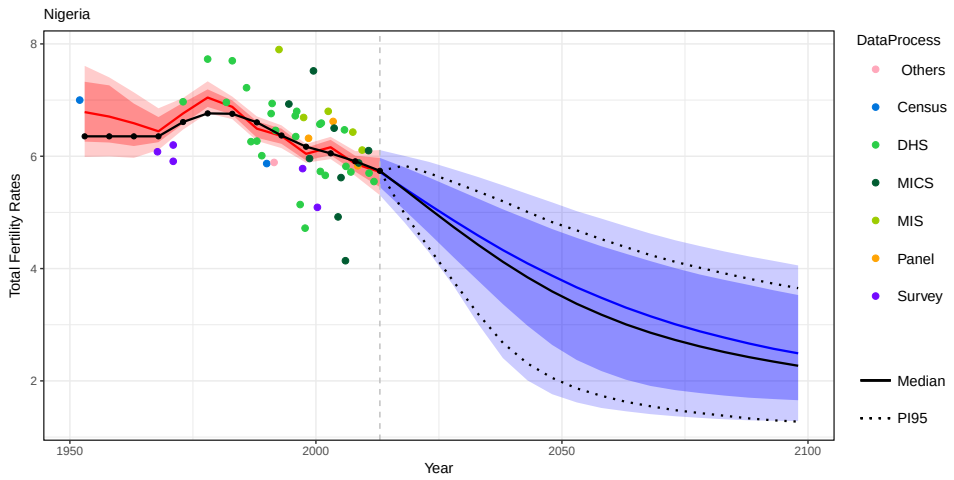


Figure 2.8: TFR projections for Nigeria. The red shaded areas are the estimated TFR with 95% estimation intervals, and the blue shaded areas are the projected TFR with 95% prediction interval, where the present is taken to be 2015, marked by a dashed vertical line. The black line and the black dotted lines represent the UN WPP 2015 median and 95% predictions.

sampled one TFR trajectory from the posterior distribution of these samples, and we assume the sampled trajectory to be the true (unobserved) TFR. Then we randomly generated TFR estimates from the normal distribution in Level 1 of the model, by assuming the bias of data points is the previous estimated bias ($\hat{\delta}_{c,s}$), and measurement error variances are the previous estimated variances ($\hat{\rho}_{c,s}$). We then treated sampled data points as the input data for the estimation process. We still treated the UN estimates as unbiased, as before.

We repeated the simulation process 1000 times, with 10,000 trajectories of fertility rates drawn in each simulation. The estimation results of one simulation are shown in Figure 2.9.

If we take the posterior median as the point estimate, the mean absolute error (MAE) for all 13 time periods is 0.157. We could also break it down by the 13 time periods, and the results are shown in Table 2.6.

The average coverage rate of 13 values the 80% interval was 85.9%, and the overall coverage rate of the 95% interval was 95.1%. The overall coverage rate was close to the

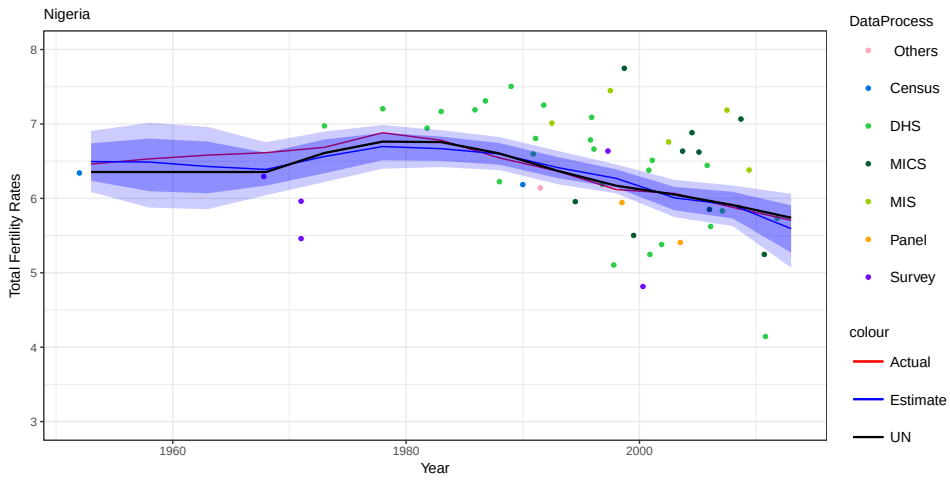


Figure 2.9: Past and Present TFR Estimates for Nigeria. Colored dots are simulated observations. The red curve is the assumed true TFR, which is assumed unobserved and the black line is the UN estimate. Shaded areas are 95% estimation intervals based on the simulated observations.

nominal rate for the 95% interval, and for 80% interval, the coverage rates have a larger uncertainty. In general the model gave accurate point and intervals estimates of past values in the simulation study.

2.5 Discussion

We have developed a new method for projecting the total fertility rate probabilistically for all countries that extends the UN method to take account of uncertainty about past TFR values. In a validation experiment, we found that the existing UN method leads to prediction intervals whose coverage is somewhat lower than nominal, while for our new method the coverage is close to nominal. For the countries with the highest quality data on past rates, mostly in Europe and North America, our method gives results that are similar to the current method. However, for countries with lower quality data where TFR estimates have been estimated based on surveys for at least part of the past 60 years, our method gives intervals that are noticeably wider than the current ones.

Table 2.6: Simulation Coverage and Mean Absolute Errors for 13 Estimation Periods.

	80% Interval Coverage	95% Interval Coverage	Mean Absolute Error
$f_{Nigeria,1953}$	0.865	0.917	0.259
$f_{Nigeria,1958}$	0.888	0.976	0.263
$f_{Nigeria,1963}$	0.895	0.982	0.222
$f_{Nigeria,1968}$	0.794	0.914	0.177
$f_{Nigeria,1973}$	0.847	0.967	0.139
$f_{Nigeria,1978}$	0.718	0.898	0.152
$f_{Nigeria,1983}$	0.797	0.933	0.118
$f_{Nigeria,1988}$	0.926	0.979	0.103
$f_{Nigeria,1993}$	0.936	0.980	0.097
$f_{Nigeria,1998}$	0.952	0.981	0.116
$f_{Nigeria,2003}$	0.848	0.940	0.090
$f_{Nigeria,2008}$	0.893	0.963	0.103
$f_{Nigeria,2013}$	0.805	0.935	0.198

These results call for improvements in the mechanisms and practice of data collection in those countries, which today have to rely on surveys in the absence of reliable vital registration. This is also one of the underlying aims of the Sustainable Development Goals agenda [United Nations, 2019c].

The long-term implications of these results are far reaching. The countries with the most uncertainty about past TFR values are also largely those with the highest current fertility levels and the greatest uncertainty about future levels, many of which are in Sub-Saharan Africa. Not surprisingly, therefore, our method indicates that these are also the countries for which the understatement of uncertainty was greatest. Thus our TFR results lead to a considerable increase in uncertainty for long-term population projections in these countries, especially as the effects of differences in TFR compound over generations. The population of Sub-Saharan Africa is currently around 1 billion, and current projections are that it will increase to between 3.4 and 4.8 billion in 2100 with 80% probability [Gerland et al., 2014, United Nations, 2017]. This interval will be wider still once uncertainty about past TFR has been factored in, with even larger implications for future population levels in Africa, and

hence for the world as a whole.

The method proposed in this chapter is in two stages. In the first stage we estimate the bias and measurement error variance of the different data sources by country using a classical analysis of variance method. In the second stage we estimate a Bayesian hierarchical model taking the point estimates from the first stage as input. In principle it would be possible to unify these two stages by including the estimation of the bias and variance of the different sources in the Bayesian hierarchical model. However, this would complicate the model considerably, making it harder to specify, code, debug and interpret, and it seems unlikely that it would change the results appreciably. It is also possible that the assumption that bias and measurement error variance is time invariant is not always true for all countries and regions and different sources of data, but considering these would also complicate the model. We feel that as our out-of-sample validation performs well in calibrating the uncertainty, we do not want to further complicate the model.

To use these projections of total fertility in population projections, one must convert them to age-specific fertility rates. The UN currently does this using the methodology described by [Ševčíková et al., 2016]. Each simulated future TFR value is converted to a corresponding age-specific fertility pattern, which is used with age-specific mortality and volume of net migration in the cohort-component projection method to project the corresponding future population by age and sex. A subtle point is that this takes account of uncertainty about future *total* fertility, but not about future *age-specific* fertility given total fertility, i.e. about the number but not the timing of future births. Because the age pattern of births is relatively concentrated regardless of their number, this is a much smaller source of uncertainty than uncertainty about the number of births [Ediev, 2013]. Nevertheless, it should be addressed in future research.

Our method does not explicitly use available information on survey design and design-based errors for DHS, MICS and other data sources, or possible measurement error in the vital registration data. Instead we estimate overall error as part of the method, including these specific contributors to error, thus bypassing these questions. We found that our

overall estimated standard errors were considerably larger than the design-based estimates of standard errors for DHS (for example four to five times larger for Nigeria), providing some support for the idea that the measurement error variance assessed by our method includes the sources of error assessed by the design-based approach. This also suggests that using the design-based estimates of standard errors alone would estimate overall uncertainty.

We have produced results for all the world's countries with populations over 100,000 as of 2015, except for one: China, the most populous country. We did not include China in our analysis because the estimates for its TFR suffer from a unique form of bias, which would require a different kind of analysis. This is due to the One Child Policy, introduced in 1979. The total fertility rates of China could still fall into the three-phase model scheme with the One Child Policy, but this has bring a specific data issue,. As a result of this policy, many Chinese families did not report births to the authorities, with the hope of being able to circumvent the policy and have additional children. This underreporting was particularly severe in the late 1990s, and [Goodkind, 2004] has argued that this was because the 1991 Decree pushed the responsibility of implementing family planning rules, especially the one-child policy, to local governments, giving them a greater incentive to underreport the number of births. This would make data of Chinese TFR suffering from a specific type of bias, which would affect the TFR estimation in 1990-2000.

There have been many efforts to correct for this underreporting. For example, [Yi, 1996, Retherford et al., 2005, Cai, 2008] and [Merli and Raftery, 2000] attempted to correct estimates of TFR in 2000. The clearest evidence of this underreporting comes from primary school enrollments several years later, which were typically substantially larger than the reported number of births during this period. [Zhai and Chen, 2007] used these enrollment data to correct the TFR estimates for the late 1990s. The UN has also been using enrollment data to correct the available estimates. Our method would not be sufficient to give good estimates of China's TFR in the period of severe underreporting. Instead, for China it would be desirable to extend our method to include enrollment data, taking account of uncertainty in the enrollment data in the model. A simpler approach would be to include enrollment-

corrected survey and census estimates as inputs to our method, but we felt that a more comprehensive approach was desirable given the great demographic importance of China and the unique data issues it presents, and so we omitted China from the present analysis.

This is a topic for a separate, tailored effort, targeted at a demographic audience. Scientists at the UN Population Division have been developing such a method and are currently preparing a paper to describe it [Wheldon, 2019].

Our method was developed to incorporate uncertainty about past fertility into projections of future fertility by the method of [Alkema et al., 2011], which is used by the UN. It focuses on uncertainty in the TFR rather than age-specific fertility rates, because for many countries without longstanding high-quality vital registration systems (the majority), estimates of past age-specific fertility rates are often not very good, while estimates of TFR are more solidly based. The same basic conceptual approach could be combined with other probabilistic fertility forecasting methods. In particular, [Bohk-Ewald et al., 2018] identified two other probabilistic fertility forecasting methods that also outperform the simple persistence, or "freeze rates", approach, namely those of [Myrskylä et al., 2013] and [Schmertmann et al., 2014]. These methods require good estimates of past age-specific fertility rates, not just total fertility, and are tailored for countries that do have well-established high-quality vital registration systems. Thus to be combined with these projection methods, our approach to assessing uncertainty about past total fertility would need to be extended to age-specfic fertility rates.

We developed an R package to implement this method and make it publicly available, building on and extending the existing `bayesTFR` R package that is used by the UN for their projections [Ševčíková et al., 2011]. This would be discussed in details in the Chapter 3

Chapter 3

A MAJOR UPDATE OF THE R PACKAGE bayesTFR: PROBABILISTIC ESTIMATION AND PROJECTIONS OF THE ANNUAL TOTAL FERTILITY RATE ACCOUNTING FOR PAST UNCERTAINTY

3.1 Introduction

In 2015 for the first time, the United Nations (UN) adopted the Bayesian method described by Alkema et al. [2011] for their official population projections for all countries, the World Population Prospects (WPP) 2015 [United Nations, 2015b]. This method is probabilistic and based on a principled statistical footing, replacing the previous deterministic method. One of the major components is the projection of the TFR which is implemented in the **bayesTFR** R package [Ševčíková et al., 2011]. This package is widely used in research on fertility rates and population projections (Abel et al. [2016], Gerland et al. [2017], Ševčíková and Raftery [2016], Ševčíková et al. [2018]).

While the projection of TFR is probabilistic, the method does not take uncertainty about the past into account. The Chapter 2 [Liu and Raftery, 2020b] addressed this issue by developing a Bayesian model that takes past TFR observations from World Fertility Data [United Nations, 2019a] as raw data, and combines the uncertainty from the data with the uncertainty from the model. Out-of-sample validation showed improved performance of the overall projection model, while providing users with information about the uncertainty of estimates of past fertility rates. A major overhaul of **bayesTFR** was required to incorporate the Liu and Raftery [2020b] methodology into the package.

The original framework implemented in **bayesTFR** was designed to work with five-year estimates and produced projections on a five-year time interval basis. This has the disadvan-

tage of missing TFR fluctuations and pattern changes within the five-year periods. There is a growing interest by the UN to publish population estimates and projections on an annual basis, and in response we have extended **bayesTFR** to work with annual data. The update revealed that an additional autoregressive component is needed to account for the larger autocorrelation and thus, to model the uncertainty in the fertility transition well.

The new version of the package now produces uncertainty information about the past which is propagated into the projections and is able to estimate and project on an annual basis. This chapter describes the methodological changes, and also provides instructions on how to generate probabilistic estimations and projections under different settings. These include with and without accounting for past TFR estimation, with annual or five-year data, and with and without the autoregressive component in the fertility transition phase of the model. Other updates in the package are also introduced and elaborated. The new version is currently available on the GitHub repository `PPgp/bayesTFR`, branch `dev`.

The chapter is organized as follows. Section 3.2 summarizes the theoretical models developed in Alkema et al. [2011], Raftery et al. [2014a], Liu and Raftery [2020b] and the autoregressive model in the fertility transition phase. Section 3.3 describes how to use the package, using a step-by-step approach with different model settings. Section 3.4 presents experiments on the performance of the models and the selection of the various settings. The article concludes with a discussion in Section 3.4.3.

3.2 Annual TFR model with uncertainty about the past

In this section, we summarize the model described in Chapter 2, and extend them for annual TFR models. We first summarize the original TFR model developed for five-year time periods [Alkema et al., 2011]. We then review the new methodology for probabilistic estimation and projection of TFR for all countries of the world accounting for uncertainty about the past, as proposed by Liu and Raftery [2020b]. Finally, we describe the changes in the methodology to work for annual estimation and projections.

Alkema et al. [2011] defined a three-phase model for the evolution of total fertility over time in a country:

1. Phase I: pre-transition phase with fluctuation at high fertility level.

2. Phase II: transition from high to low fertility where decrements are modeled by a random walk with drift given by a double logistic function.

3. Phase III: post-transition phase where fertility fluctuates around replacement level, modeled by an autoregressive AR(1) process.

We will use the same notation as in Ševčíková et al. [2011]. Specifically, $f_{c,t}$ denotes the TFR in country c and time period t, τ_c denotes the start period of Phase II for country c, λ_c is the start period of Phase III for country c, while $g(\boldsymbol{\theta}_c, f_{c,t})$ and $\boldsymbol{\theta}_c$ denote the parametric decline function and the corresponding country-specific parameters, respectively.

3.2.1 Existing model with five-year estimates

The pre-transition phase (Phase I) is not modeled as all countries in the world have already entered Phase II and thus, for the purpose of projecting into the future it is not needed. Note that this is the model based on Alkema et al. [2011]. In the chapter 2, we have already build the model for Phase I and will be discussed later.

The fertility transition phase (Phase II) is modeled by a random walk with drift. This is specified by

$$f_{c,t+1} = f_{c,t} - d_{c,t} \quad \text{for} \quad \lambda_c \leq t < \tau_c \,. \tag{3.1}$$

The decrement $d_{c,t}$ in (3.1) is modeled as the sum of a function of the level of the TFR and the noise, as follows:

$$d_{c,t} = d(\boldsymbol{\theta}_c, \lambda_c, \tau_c, f_{c,t}) = g(\boldsymbol{\theta}_c, f_{c,t}) + \varepsilon_{c,t} \tag{3.2}$$

where $g(\boldsymbol{\theta}_c, f_{c,t})$ are the double logistic decrements, which are determined by the country-specific parameter vector $\boldsymbol{\theta}_c$ and given by

$$\frac{-d_c}{1+\exp\left(-\frac{2\ln(p_1)}{\Delta_{c1}}(f_{c,t}-\sum_i \Delta_{ci}+0.5\Delta_{c1})\right)}+\frac{d_c}{1+\exp\left(-\frac{2\ln(p_2)}{\Delta_{c3}}(f_{c,t}-\Delta_{c4}-0.5\Delta_{c3})\right)}. \tag{3.3}$$

The random distortions $\varepsilon_{c,t}$ in each period are given by normal distributions as:

$$\varepsilon_{c,t} \sim \begin{cases} N(m_t, s^2), & \text{for } t = \tau_c \\ N(0, \sigma(f_{c,t})^2) & \text{otherwise}\,. \end{cases} \tag{3.4}$$

The quantity $\sigma(f_{c,t})$ is the standard deviation of the distortions during the later periods with

$$\sigma(f_{c,t}) = c_{1975}(t)\left(\sigma_0 + (f_{c,t}-S)(-aI_{[S,\infty)}(f_{c,t}) + bI_{[0,S]}(f_{c,t}))\right)\,. \tag{3.5}$$

For further details about the model and its priors, see Ševčíková et al. [2011]. For the purpose of this chapter, we only point to the definition of two parameters, namely the country-specific maximum decrement d_c, and the hyperparameter for the maximum standard deviation of the distortions σ_0. The d_c parameter is defined as

$$\begin{aligned} d_c^* &= \log\left(\frac{d_c - 0.25}{2.5 - d_c}\right)\,, \\ d_c^* &\sim N(\chi, \psi^2)\,. \end{aligned} \tag{3.6}$$

The prior distribution of σ_0 is $\sigma_0 \sim U[0.01, 0.6]$, which is wide enough to cover the range of standard deviation of all countires' TFR uncertainty.

The TFR in the post-transition phase (Phase III) is modeled by a first order autoregressive

time series model [Raftery et al., 2014a] as

$$f_{c,t+1} \sim N(\mu_c + \rho_c(f_{c,t} - \mu_c), \sigma^2) \text{ for } t \geq \lambda_c \tag{3.7}$$

where μ_c is the country-specific long-term mean fertility rate, and ρ_c is the autoregressive parameter with $\rho_c \in (0, 1)$. In **bayesTFR** these parameters can be estimated via the Markov chain Monte Carlo (MCMC) method. Alternatively, country-dependent values can be predefined or estimated via maximum likelihood.

The start period of Phase II, τ_c, is defined as

$$\tau_c = \begin{cases} \max\{t : (M_c - L_{c,t}) < 0.5\}, & \text{if } L_{c,t} > 5.5; \\ \text{first estimation year}, & \text{otherwise}, \end{cases} \tag{3.8}$$

where M_c is the maximum observed TFR outcome in country c, and $L_{c,t}$ denote local maxima.

The start period of Phase III for country c, λ_c, is defined as the period where two consecutive increases of TFR below 2 have been observed. More formally,

$$\lambda_c = \min\{t : f_{c,t} > f_{c,t-1}, f_{c,t+1} > f_{c,t} \text{ and } f_{c,p} < 2 \text{ for } p = t-1, t, t+1\}\,. \tag{3.9}$$

3.2.2 Probabilistic TFR estimation with uncertainty

This section is summary of models in Chapter 2, but with updated World Fertility Data [United Nations, 2019a] for past raw TFR estimates from surveys, reports and vital registrations for most regions in the world. We denote these data points by $y_{c,t,s}$, i.e. the raw TFR estimate for country c, time t and source s. The source s may refer to a census, a survey, vital registration statistics or other sources. For each observation, there are features $\boldsymbol{x}_{c,s}$ that describe the sources, estimating methods, recall lags and other aspects of data collection and

estimation, often measures of the quality of the data. The observed $y_{c,t,s}$ are modeled as:

$$\begin{aligned} y_{c,t,s}|f_{c,t} &\sim N(f_{c,t} + \delta_{c,s}, \rho^2_{c,s}), \\ E[\delta_{c,s}] &= \boldsymbol{x}^T_{c,s}\boldsymbol{\beta}, \\ E[\rho_{c,s}] &= \boldsymbol{x}^T_{c,s}\boldsymbol{\gamma}. \end{aligned} \tag{3.10}$$

The $\delta_{c,s}$ and $\rho_{c,s}$ are country-specific parameters which are estimated by least squares. In Liu and Raftery [2020b], the features used are the sources of the data and the corresponding estimation methods, but the model allows for any user-specified features. This part is combined with the existing Bayesian hierarchical model implemented in **bayesTFR**. Here, the past TFRs are considered as unknown, and are part of the parameters to estimate. The complete model is described in Appendix.

If we are using TFR for five-year intervals, as for example in the `tfr` dataset in the **wpp2019** package, the true TFR at any time stamp is considered as the linear interpolation of two consecutive five-year TFRs, computed by

$$f_{c,t} = \frac{1}{5}[(t_{\ell+5} - t)f_{c,t_\ell} + (t - t_\ell)f_{c,t_{\ell+5}}] \quad \text{for any } t \in [t_\ell, t_{\ell+5}]\,.$$

If we are estimating from annual TFR, for each observation of the raw data, we take the floor of t. For example, if an observation in the raw data is recorded at 1975.5, we consider this observation as an estimate of the calendar year 1975 since demographers are estimating demographic statistics based on calendar years.

Since we are now also modeling the past, not just the future as in the extant method, we need to model the pre-transition phase (Phase I). The TFR in this phase will be modeled by a random walk, specified by

$$f_{c,t+1} = f_{c,t} + \varepsilon_{c,t} \quad \text{for } t < \tau_c,$$

where the distributions of the random distortions in each period are given by

$$\varepsilon_{c,t} \sim N(0, s^2)\,.$$

Here, we simplify the model by setting the variance to be the same as the variance in the first period of the TFR decrements. This is a reasonable assumption because the starting period of Phase II is linked to Phase I, and the expected decline of TFR at the starting period of Phase II is small. Thus, the distortions of TFR share similar behavior.

The estimation of all country-specific parameters and hyperparameters conditional on the TFRs, other than the TFRs themselves, in the Phase II model remains the same as described by Ševčíková et al. [2011]. To estimate past TFR, the model for Phase III is estimated together with the Phase II model via the MCMC algorithm [Gelfand and Smith, 1990]. This algorithm is a combination of Gibbs sampling, Metropolis-Hastings (for Δ_{ci} in (3.3) and TFR), and slice sampling steps [Neal, 2003].

The estimation yields a set of TFR trajectories about the past. To project into the future, we apply the existing projection method as described in Ševčíková et al. [2011] starting with the last estimation period of each trajectory. This is in contrast with the previous version, where the projection for each country starts from a single data point, namely the last observed TFR.

3.2.3 Annual version of bayesTFR

The original model described in Section 3.2.1 was designed to work with five-year data. Several modifications needed to be made in order to estimate and project annual data well.

Most importantly, we found that unlike in the five-year version, the residuals of the Phase II model are highly autocorrelated when using annual data. We found a correlation coefficient of 0.7 for some major countries. Thus, we modified the Phase II model defined in Equations (3.1)-(3.2) by adding an additional order one autoregressive component. The

model for Phase II TFR is then specified as

$$d_{c,t+1} - g(\boldsymbol{\theta}_c, f_{c,t+1}) = \phi(d_{c,t} - g(\boldsymbol{\theta}_c, f_{c,t})) + \varepsilon_{c,t}. \tag{3.11}$$

The prior distribution of ϕ is set to be Uniform$(0, 1)$ and is not country-specific. For the random distortions $\varepsilon_{c,t}$, the distribution is considered to be the same as in Equations (3.4)-(3.5).

The same prior distributions as in the five-year version is used for most parameters. One exception is σ_0 where the lower bound was decreased by a factor of the square root of five, i.e. $\sigma_0 \sim U[0.0045, 0.6]$. The upper bound was kept the same to allow for the possibility of additional correlation.

The definition of the maximum decrement defined in Equation (3.6) was changed to be one-fifth of that for the five-year model:

$$d_c^* = \log\left(\frac{d_c - 0.05}{0.5 - d_c}\right).$$

No changes have been made to the model of the post-transition phase of TFR, Phase III. It is modeled by a first order autoregressive time series model as defined in Equation (3.7).

The rule for determining the start period of Phase II, τ_c, as defined in Equation (3.8), was unchanged. However, since the local maxima are calculated using annual TFR data, the results can differ from those obtained from a five-year dataset.

To determine the start periods of Phase III, λ_c, as defined in Equation (3.9), we first obtain five-year averages of TFR by subsampling every five values and apply the same rule as in the five-year version, namely that Phase III starts when two consecutive increases of TFR below 2 are observed.

3.2.4 Changes in TFR Projections

There are three main differences in the TFR projections between the new implementation and the one described by Ševčíková et al. [2011].

The first difference (which we alluded to at the end of Section 3.2.2), relates to the fact that by accounting for the past TFR uncertainty (Equation 3.10), instead of observed point estimates, the model results in a set of TFR trajectories about the past which changes the starting values of the forecast. To project $f_{c,T+1}$ where T is the last period of the estimation, the i-th sample from the MCMC output is given by $f^{(i)}_{c,T+1} = f^{(i)}_{c,T} - d^{(i)}_{c,T} + \varepsilon^{(i)}_{c,T}$. Thus, the uncertainty in the first forecast period will be wider than if we use a model without accounting for past uncertainty, in which case $f^{(i)}_{c,T} = f_{c,T}$ is the same for all trajectories.

The second difference relates to the annual model described in Section 3.2.3. When the additional autocorrelation of Phase II is taken into account (Equation 3.11), the past noise is carried over to the next time period. Specifically, to project $f_{c,t+1}$ for a country c that is in Phase II at time t, the i-th sample is given by $f^{(i)}_{c,t+1} = f^{(i)}_{c,t} - d^{(i)}_{c,t} + \varepsilon^{(i)}_{c,t}$, where $d^{(i)}_{c,t} = g(f^{(i)}_{c,t}, \boldsymbol{\theta}^{(i)}_c)$, and $\varepsilon^{(i)}_{c,t}$ is drawn from $N(\phi^{(i)}\varepsilon^{(i)}_{c,t-1}, \sigma^{(i)}(f^{(i)}_{c,t}))$. For the first forecast, i.e. at the time period $T+1$, the distortion of the last estimation period T is calculated before starting the projections.

Finally, the last difference regards the updated Phase III model as described in Raftery et al. [2014a], where country-specific long-term means μ_c and autocorrelation coefficients ρ_c were incorporated into the model (Equation 3.7) and estimated by MCMC. However, this change has been available in **bayesTFR** since version 3.0-0 published in 2013. Using this approach, to project $f_{c,t+1}$ for a country c that is in Phase III at time t, the i-th MCMC sample is drawn from a Normal distribution $N\left(\mu^{(i)}_c + \rho^{(i)}_c(f^{(i)}_{c,t} - \mu^{(i)}_c), \sigma^{(i),2}\right)$.

3.3 Using updated bayesTFR

Currently, the R package **bayesTFR** implements the model described in Section 3.2.1. It is used by UN analysts and others to train TFR projections models based on past five-

year estimates. New UN requirements motivated updating the package so that analysts can conduct estimation of past TFR with accounting for uncertainty, and make corresponding forecast for both, five year and annual time periods.

To make probabilistic TFR projections accounting for past TFR uncertainty, the following four steps are needed:

1. Data assembly (optional):

   (a) Prepare a dataset on raw TFR. By default, the World Fertility Data 2019 [United Nations, 2019a] is used.

   (b) Assemble a dataset on reference (initial) TFR. By default, the *the World Population Prospects* in the **wpp2019** package is used.

2. Model estimation:

   (a) Train linear models to estimate systematic bias and standard deviation for each observation from the raw TFR dataset.

   (b) Given the reference TFR, calculate the start period of Phase II and the start period of Phase III for each country (τ_c and λ_c).

   (c) Run the MCMC process to obtain posterior sample of the Phase II and Phase III model parameters, and the posterior sample of the past TFR for all countries.

3. Generate future TFR trajectories as discussed in Section 3.2.4.

4. Analyze outputs using a set of functions that summarize, plot, diagnose and export results of the three steps above.

As described in Ševčíková et al. [2011], steps 2 and 3 are expected to be relatively time-consuming. Adding the estimation uncertainty feature as well as working with annual estimates will add even more run time. Even though we optimized the code wherever possible, it might take several hours to complete these steps in a production-like setting.

The following sections describe the steps above in more detail, especially parts that are different from Ševčíková et al. [2011]. We will elaborate how to use the new features, as well as how to use it in the original way. We will demonstrate the package on a realistic example with a large number of MCMC iterations. To save time, users who merely want to test functionality should decrease the number of iterations by, say, a factor of 10. However note that since the Metropolis-Hastings step for TFR updates will give the acceptance rate at around 30%, a small number of iterations might result in significant differences in the estimation plots with respect to smoothness than what we will present.

3.3.1 Data assembly and estimation settings

The datasets assembled in this step will be passed to the main estimation function, `run.tfr.mcmc`, which now has additional arguments for this purpose. It can be specified what raw TFR data to use, whether to estimate and predict annually (logical argument `annual`), and whether to use the AR(1) model in Phase II as defined in Equation 3.11 (logical argument `ar.phase2`).

Argument `uncertainty = TRUE` specifies that estimation with taking into account uncertainty about the past is triggered. In such a case, a raw TFR dataset can be provided. By default, the World Fertility Data 2019 [United Nations, 2019a] is used. This dataset contains 12,079 records for 201 countries, each of which includes the corresponding estimation method and data source. These are then used by the model as data quality covariates in Equation 3.10.

The default raw TFR dataset can be viewed via

```
R> data("rawTFR")
R> head(rawTFR)

country_code year tfr method source
1        4 1965      7.97           Indirect      Census
2        4 1966      8.21           Indirect      Census
```

```
3        4 1967      8.32          Indirect      Census
4        4 1968      8.23          Indirect      Census
5        4 1969      8.07          Indirect      Census
6        4 1970      7.98          Indirect      Census
```

The default covariates are `c("source", "method")`. Users can provide their own file. For example, we could run the following code to create our own dataset.

```
R> setwd("path/to/working/directory")
R> data("rawTFR")
R> write.csv(rawTFR, "raw_tfr_user.csv")
R> tfr.raw.file <- "raw_tfr_user.csv"
```

Here, we create a data set by copying the data set in the package for illustration purpose, but the users can provide their own data sets with different covariates, or add other columns. Only `country_code, year, tfr` are not subject to change, and users could provide continuous covariates either. For example, users could provide data as in Table 3.1: Note that

country_code	year	tfr	method	source	lag
4	1965	7.97	Indirect	Census	0
...	...	...	...	...	...

Table 3.1: Format for the user-specific raw data. The names of country_code, year and tfr columns must be fixed and must be given, and the remaining covariates could have different names. In this example, `method`, `source` are categorical variables and `lag` is considered as continuous variable. Order of columns will not affect the simulation.

users are allowed to provide `csv` files with comma separately data set or `txt` files with `tab` separated data sets. Others formats are not supported yet. Throughout the paper, we would use the file `"raw_tfr_user.csv"` as the raw dataset when we are **not** using the default dataset.

The second dataset to assemble is a dataset on a reference, or initial, TFR. Its file name is passed into the argument `my.tfr.file`. If `uncertainty = FALSE`, this dataset is considered

in the estimation as the true observed TFR. Otherwise, it is used as the starting points of TFR in the MCMC process. By default, if `my.tfr.file` is not given, the `tfr` dataset from the **wpp2019** package is used for this purpose, which is a five-year dataset. Thus, if `annual = TRUE`, a linear interpolation of the default dataset is derived.

3.3.2 Fitting the TFR model

Most arguments of `run.tfr.mcmc` remain the same as described in Ševčíková et al. [2011]. Importantly, `start.year` and `present.year` set the first and the last year of the time series included in the computation, respectively. Arguments `nr.chains`, `iters` and `output.dir` determine the number of chains, the number of iterations and the directory to store MCMC estimations, respectively.

In the existing version of **bayesTFR**, the function `run.tfr.mcmc` is designed to obtain a posterior sample of Phase II model parameters. The estimation of Phase III parameters (as defined in Equation 3.7) is implemented in the function `run.tfr3.mcmc`. When building a full probabilistic model as described in Liu and Raftery [2020b], MCMC steps for updating TFR will affect both phases. Thus, if `uncertainty = TRUE`, the new `run.tfr.mcmc` function combines the estimation of both phases together and there is no need to invoke the `run.tfr3.mcmc` function explicitly. We'll call this method a "one-step estimation". However, this is not the case if uncertainty about the past is not taken into account. In such a case, the workflow of estimating Phase II and Phase III separately remains and is further called a "two-step estimation".

The various combinations of the possible settings are summarized in Table 3.2. We have marked each cell with a letter which will be referred to in the remainder of this section.

As described in Section 3.2.3, when using the annual model (cells A and B), adding the autoregressive component in Phase II as defined in Equation 3.11 should be considered. The option is controlled via the logical argument `ar.phase2` which should be passed to `run.tfr.mcmc` function. If set to `TRUE` the MCMC process performs an extra slice sampling step for estimating ϕ, an extra country-independent parameter in the model. The argument

annual	uncertainty	
	TRUE	FALSE
TRUE	**A** one-step estimation; Phase II - AR(1) allowed	**B** two-step estimation; Phase II - AR(1) allowed
FALSE	**C** one-step estimation	**D** two-step estimation

Table 3.2: Possible combinations in fitting TFR projection model

is ignored if `annual` is `FALSE`.

Starting a new simulation with two step estimation

The two-step estimation should be performed if uncertainty about the past is not taken into account (cells B and D in Table 3.2). The main differences between cells B and D are the setting of prior distributions as described in Section 3.2.3, and whether the autoregressive component can be included in the model. Here we give an example of a simulation with annual data (cell B) without the autoregressive component. However, we will not use this example further in the text, as the main focus of the article is on cell A which will be discussed in the next section.

Our example simulation consists of three MCMC chains, each of which is 5000 iterations long where thinning is disabled. (As noted earlier, the user is advised to decrease the number of iterations by a factor of ten for faster processing.) We will save the simulation results to a directory called "annual".

```
R> annual <- TRUE
R> nr.chains <- 3
R> total.iter <- 5000
R> thin <- 1
R> simu.dir <- file.path(getwd(), "annual")
```

The first step is to launch an estimation of Phase II:

```
R> m2 <- run.tfr.mcmc(output.dir = simu.dir, nr.chains = nr.chains,
+   iter = total.iter, thin = thin, annual = annual)
```

The second step is to start an estimation of Phase III:

```
R> m3 <- run.tfr3.mcmc(sim.dir = simu.dir, nr.chains = nr.chains,
+   iter = total.iter, thin = thin)
```

Here, we are using the same number of chains and iterations for Phase II and Phase III. However, this is not a requirement, but rather depends on the MCMC convergence. Even the 3×5000 iterations might be not enough to reach a convergence, but will give realistic outputs. Setting `total.iter = 62000` or `total.iter = "auto"` will most likely result in full convergence.

Starting a new simulation with one-step estimation

We now show an example of a simulation with uncertainty which is performed with one step only (cells A and C in Table 3.2). In particular, here we set `annual` to `TRUE` (cell A), but the same function would be used if `annual` is FALSE (cell C). We will also include the Phase II - AR(1) model (`ar.phase2`) which would not have any effect in cell C. The results will be saved in the directory "annual_unc". We will use this simulation throughout the article.

```
R> annual <- TRUE
R> ar.phase2 <- TRUE
R> nr.chains <- 3
R> total.iter <- 5000
R> thin <- 1
R> simu.dir.unc <- file.path(getwd(), "annual_unc")
```

As in the previous case, this setting may not be enough to yield fully converged MCMC simulations, but will still give realistic outputs. The processing time is within a range of a

couple of hours (with one MacBook Pro 2017 version, with 2.8GHz Quad-Core Intel Core i7 with 16 G RAM, 5000 iterations for 3 chains will take 3 hours). For faster processing, set `total.iter = 50` for a toy simulation. In addition, the `parallel` argument can be set to `TRUE`, in which case the three chains will be processed in parallel. In Section 3.4.2, we will give recommendations for settings that yield fully converged MCMC simulations. When appropriate, we will use results from such converged simulations to present outputs.

As mentioned in Section 3.3.1, additional arguments of `run.tfr.mcmc` allow one to pass user-specific raw TFR data (`my.tfr.raw.file`), names of categorical covariates (`covariates`), names of continuous covariates (`cont_covariates`), or to specify countries with unbiased vital registration data (`iso.unbiased`). If the `iso.unbiased` option is used, the `covariates` argument must include the variable "source" as in such a case, the function reduces the bias and standard deviation of records where the "source" column specifies "VR". In our example we will specify that the vital registration data of Canada and the USA (codes 124 and 840) are unbiased.

To estimate both Phase II and Phase III, one could do

```
R> m <- run.tfr.mcmc(output.dir = simu.dir.unc, nr.chains = nr.chains,
+   iter = total.iter, annual = annual, thin = thin, uncertainty = TRUE,
+.  ar.phase2 = ar.phase2, iso.unbiased = c(124, 840))
```

In comparison to the two-step model, the training process here has an extra Metropolis-Hastings step per iteration for generating posterior TFR samples.

Continuing an existing simulation

If an existing simulation needs to be extended by more iterations, one would proceed as in the current version of **bayesTFR**:

- Use `continue.tfr.mcmc` if the MCMCs were originally created via `run.tfr.mcmc`, regardless of whether one is in the one-step or the two-step estimation mode.

- Use `continue.tfr3.mcmc` if the MCMCs were originally created via `run.tfr3.mcmc`.

Now suppose we want to extend the simulation in the previous section by 100 iterations. Then we could do

```
R> m <- continue.tfr.mcmc(output.dir = simu.dir.unc, iter = 100)
```

(Set the `iter` argument to 10 if working with a toy simulation.) This will continue running both TFR phases in a one-step estimation while inheriting all settings from the original simulation, including `uncertainty`, `annual`, `ar.phase2` or settings about the raw data. At the end of the processing, each chain will be $5,100$ iterations long.

3.3.3 Generating projections

The main function for generating projections is called `tfr.predict`. The new version of the package adds the argument `uncertainty`. If it is `TRUE` and the model was estimated taking uncertainty about the past into consideration, then that past uncertainty will be carried over to the projections. In this case, each future trajectory starts from a trajectory representing the past.

Suppose we want to generate projections represented by 1,000 posterior trajectories until the year 2100 based on the simulation stored in the directory given by `simu.dir.unc`, with burn-in of the first 2100 iterations for each chain. This can be done using the following command:

```
R> pred <- tfr.predict(sim.dir = simu.dir.unc, end.year = 2100, burnin = 2100,
+     nr.traj = 1000, uncertainty = TRUE)
```

The function takes the existing $5,100$ iterations in each chain, removes the first 2,100 values and generates 1,000 TFR trajectories based on 1,000 equally spaced parameter values and past TFR, out of the remaining $3,000 \times 3 = 9,000$ iterations. For a toy simulation, use `burnin = 20` and `nr.traj = 50`.

If `uncertainty` is set to `FALSE`, all future trajectories start from the last observed data point. If the estimation process accounted for uncertainty, but the projection does not, the starting value of the projections is the initial TFR value for the last observed time period. This is however not recommended but may serve the purpose of apples-to-apples comparisons.

3.3.4 Analyzing Results

If results are to be explored at a later time point, one can load the estimation object from disk using the command

```
m <- get.tfr.mcmc(simu.dir.unc)
```

For one-step estimation, this object contains information about both phases. For a two-step simulation, or if the Phase III object is to be extracted explicitly, use

```
m3 <- get.tfr3.mcmc(simu.dir.unc)
```

Similarly, to load the prediction object from disk, do

```
pred <- get.tfr.prediction(simu.dir.unc)
```

Summary functions

For the summary statistics of the estimation object in this section, we will use the following thinning and burnin settings:

```
thin <- 3
burnin <- 2100
```

Use `thin <- 1` and `burnin <- 20` if you're working with the toy simulation.

To view a summary of country-independent parameters, one can use

```
R> summary(m, thin = thin, burnin = burnin)
```

Since the object `m` was obtained using a one-step estimation, the output includes estimation summaries for both phases. In comparison to the current version of the package, Phase II contains one additional parameter, namely "rho_phase2" which represents ϕ in model (3.11). As with any other parameter, the name, or multiple parameter names, can be passed to the function to view summary statistics for those selected parameters.

```
R> summary(m, par.names = c("rho_phase2", "sigma0"), thin = thin, burnin = burnin)

MCMCs of phase II
=================
Number of countries: 201
Hyperparameters estimated using 201 countries.

WPP: 2019
Input data: TFR for period 1950 - 2020 .

Iterations = 2103:5100
Thinning interval = 3
Number of chains = 3
Sample size per chain = 1000

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

Mean      SD  Naive SE Time-series SE
rho_phase2 0.6988 0.04117 0.0007516       0.005133
sigma0     0.0562 0.01111 0.0002028       0.002038

2. Quantiles for each variable:
```

```
2.5%     25%     50%     75%  97.5%
rho_phase2 0.59876 0.67408 0.71200 0.72904 0.7470
sigma0     0.04876 0.05162 0.05418 0.05694 0.0853
```

The full list of parameter names for Phase II can be obtained via

```
R> tfr.parameter.names(meta = m$meta)
```

```
[1] "alpha"  "alphat"  "delta"   "Triangle4"  "delta4"  "psi"  "chi"
[8] "a_sd"  "b_sd"  "const_sd"  "S_sd"  "sigma0"  "mean_eps_tau"  "sd_eps_tau"
[15] "rho_phase2"
```

Passing the `meta` argument is needed to identify that the simulation contains a Phase II - AR(1) estimation, and thus it contains the "rho_phase2" parameter. Phase III parameter names are not dependent on the simulation, thus no `meta` argument is needed:

```
R> tfr3.parameter.names()
```

```
[1] "mu"        "rho"       "sigma.mu"  "sigma.rho" "sigma.eps"
```

Specifying a country in the `summary` function will show results of country-specific parameters of that country. Functions `tfr.parameter.names.cs()` and `tfr3.parameter.names.cs()` list the allowed parameter names for Phase II and Phase III, respectively. For a simulation that took into account uncertainty about the past, there is an additional country-specific parameter, called "tfr," capturing that uncertainty. It is not listed explicitly via the above functions, but it can be explored like any other parameter. For the `summary` function it means passing it to the `par.names.cs` argument. For example, to view summary statistics of TFR estimation for Nigeria, we can do

```
R> summary(m, country = "Nigeria", par.names.cs = "tfr",
+     thin = thin, burnin = burnin)
```

The tabular sections of the output contain one row per past observed period each (by default 71, i.e. from 1950 to 2020). To select a subset we can specify which time periods we are interested in as `tfr_`x. For example, to view results for time periods 1, 30 and 70 (corresponding to 1950, 1979 and 2019) we do

```
R> summary(m, country = "Nigeria",
+     par.names.cs = c("tfr_1", "tfr_30", "tfr_70"),
+     thin = thin, burnin = burnin)
```

```
...
1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

Mean     SD Naive SE Time-series SE
tfr_1_c566  6.281 0.2390 0.004363       0.045831
tfr_30_c566 6.709 0.0762 0.001391       0.007952
tfr_70_c566 5.622 0.4785 0.008735       0.039332

2. Quantiles for each variable:

2.5%   25%   50%   75% 97.5%
tfr_1_c566  5.765 6.135 6.290 6.438 6.731
tfr_30_c566 6.577 6.655 6.704 6.760 6.879
tfr_70_c566 4.827 5.113 5.765 5.916 6.320
```

Exploring TFR estimation

In addition to summary statistics, one can explore the estimated trajectories as well as any quantiles of the past TFR estimates. For example,

```
R> nigeria_obj <- get.tfr.estimation(country.code = 566, sim.dir = simu.dir.unc,
+    burnin = burnin, thin = thin, probs = c(0.025, 0.1, 0.5, 0.9, 0.975))
```

returns a list where trajectories are contained in the element `tfr_table`. The number of rows corresponds to the number of trajectories (here $3000 = 3[\text{chains}] \cdot (5100 - 2100)/3[\text{thin}]$), while columns correspond to time periods (here 71).

```
R> dim(nigeria_obj$tfr_table)
```

```
[1] 3000   71
```

The quantiles are contained in the element `tfr_quantile`:

```
R> head(nigeria_obj$tfr_quantile)
```

```
2.5%      10%      50%      90%    97.5% year
1: 5.764658 5.968237 6.289810 6.585244 6.730593 1950
2: 5.803398 5.979124 6.287563 6.563627 6.684089 1951
3: 5.810477 6.013333 6.308801 6.538703 6.702456 1952
4: 5.837490 6.004007 6.314881 6.536717 6.703307 1953
5: 5.862094 6.017064 6.317159 6.540123 6.712419 1954
6: 5.846718 6.027068 6.325334 6.545919 6.701965 1955
```

This element is missing if the `probs` argument is not given.

To plot the estimation with user-defined intervals, do for example:

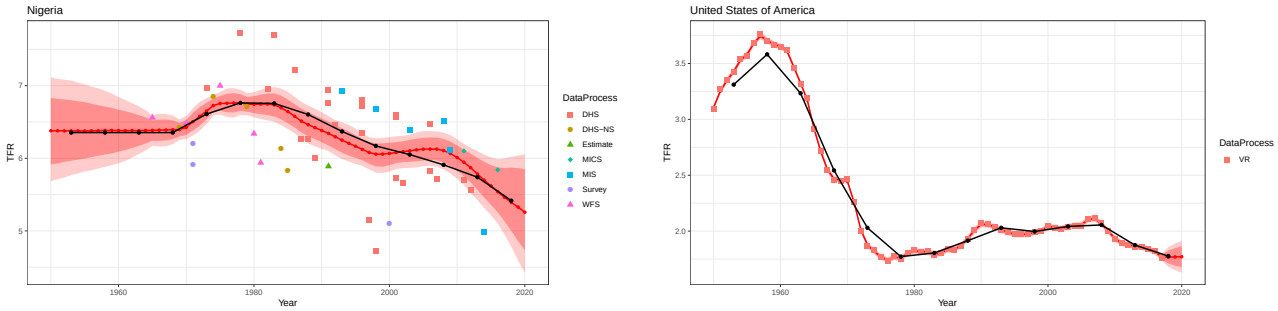


Figure 3.1: Annual TFR estimation for Nigeria (left panel) and the United States (right panel), resulting from a converged simulation. The red line shows the posterior median, while the red shaded area shows the pointwise 80% intervals, and the pink shaded areas shows the corresponding 95% intervals. The UN's 2019 WPP five-year estimates are shown by the black line.

```
R> plot <- tfr.estimation.plot(country.code = 566, sim.dir = simu.dir.unc,
+     burnin = burnin, thin = thin, pis = c(80, 95), plot.raw = TRUE)
R> print(plot)
```

The function uses the **ggplot2** package [Wickham, 2016] to visualize estimation uncertainty. Figure 3.1 shows results of the function call for Nigeria (as above) and the USA (`country.code = 840`) using a converged simulation.

Several arguments in this function need to be clarified:

1. `sim.dir`: Users can specify the location of the simulation set, or use the `mcmc.list` argument to pass the `m` object directly. For example `tfr.estimation.plot(mcmc.list = m, ...)`.

2. `country.code`: Either a name or numerical code of the country. Alternatively, users could provide the ISO-3 character code via the `ISO.code` argument, such as `ISO.code = "USA"` for the United States.

3. `pis`: Specifies which probability intervals will be plotted. It is a vector of maximum two elements.

4. `plot.raw`: Logical argument which determines whether the raw data used for estimating past uncertainty are plotted. If `TRUE` and the estimation process was not based on the default data, users need to provide the argument `grouping`, which should be one of the categorical covariates in his/her data set.

5. `save.image`: (not used in the call above) If `TRUE`, the plot will be saved as a pdf file in the directory specified in the argument `plot.dir`, named "tfr_country*code*.pdf".

Exploring bias and standard deviation of observations

Information about the bias and standard deviation of observations will give users an indication of the quality of the observations and whether these quantities were poorly estimated.

Now suppose we are interested in the bias and standard deviation estimates of the observations for Nigeria. Then we could use

```
bias_sd <- tfr.bias.sd(sim.dir = simu.dir.unc, country.code = 566)
```

The function will return a list with elements `model_bias`, `model_sd` and `table`. The `model_bias` and `model_sd` objects are of class `lm` and contain the linear models used to estimate the bias and standard deviation, respectively, while the `table` object includes the observed data points, data quality covariates, and the actual estimates for the specified country, here for Nigeria.

```
R> summary(bias_sd$model_bias)
R> head(bias_sd$table)
```

The results are shown in Tables 3.3 and 3.4.

To generate the estimates in the `table` object, the `predict` S3 method is applied to the `model_*` objects. Then the following adjustments are made if needed:

1. For some countries, the number of data points is very small for several groups. This could lead to a large bias, but a very small variance. As a result, the estimation will be

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.43	0.10	-4.37	0.00
covariate_1DHS-NS	-0.31	0.17	-1.76	0.09
covariate_1Estimate	-0.14	0.37	-0.39	0.70
covariate_1MICS	0.72	0.27	2.68	0.01
covariate_1MIS	0.29	0.16	1.79	0.08
covariate_1Survey	-0.21	0.23	-0.94	0.35
covariate_1WFS	-0.18	0.17	-1.06	0.30
covariate_2Indirect	0.80	0.11	7.05	0.00

Table 3.3: Linear model for bias obtained from `summary(bias_sd$model_bias)` for Nigeria.

method	source	bias	std
Indirect	WFS	0.18	0.13
Indirect	DHS-NS	0.06	0.09
Direct	Survey	-0.64	0.64
Indirect	DHS	0.37	0.28
Direct	WFS	-0.61	0.61
Direct	DHS-NS	-0.74	0.74

Table 3.4: Bias and standard deviation of each observations obtained from `bias_sd$table` for Nigeria.

unreasonably concentrated on the bias-adjusted data points. Thus, an adjustment to the standard deviation is made so that the estimated standard deviation is not smaller than the half of estimated bias.

2. For countries included in `iso.unbiased`, the model estimates are overwritten.

3. Duplicates are dropped so that the combinations of data quality covariates are unique.

The output can help to detect problematic estimates on certain data points so that adjustments can be made by the analyst if necessary. In the example above, the estimated bias and standard deviation for Indirect method and DHS-NS source were 0.06 and 0.09, respectively. These estimates were derived based on only three data points in this category which lie very close to the UN estimates (three of the brown dots in Figure 3.1 closest to the

black line). Even though conditionally on the model these estimates are correct, it might be overconfident to assume that the nationwide DHS estimates could reach such a high level of accuracy, and thus decrease the uncertainty around those points. Possible remedies could include adding more covariates to the model or removing questionable points from the dataset. However, we do not recommend doing this unless the analysts could understand the data very well, including the data collection, estimation processes, and could point out why specific dataset should be preferred.

Exploring TFR prediction

Plotting the posterior sample of projected TFR trajectories is done via the `tfr.trajectories.plot` function. The updated version of the package incorporates uncertainty about the past, if taken into account in the estimation and projection. For example, to plot the prediction of TFR for Burkina Faso contained in the `pred` object created in Section 3.3.3 or at the beginning of Section 3.3.4, use

```
R> tfr.trajectories.plot(pred, country = "Burkina Faso", nr.traj = 20,
+    pi = c(80, 95), uncertainty = TRUE)
R> tfr.trajectories.plot(pred, country = "Burkina Faso", nr.traj = 20,
+    pi = c(80, 95), uncertainty = FALSE)
```

Here, the parameter `uncertainty` is used to specify whether the uncertainty about the past TFR should be plotted together with the prediction. If `uncertainty` is `TRUE`, optional parameters `thin, burnin, col_unc` can be used to define the burn-in, thinning and the color for the past uncertainty plot.

Here, `pred` refers to a non-converged result. If we change the input `pred` to a converged ones, the plots would be as Figure 3.2. If the user selects `uncertainty = FALSE` for a simulation where past uncertainty was taken into account (similarly to the right panel of Figure 3.2), the past TFR used for the initialization of the model is shown as the observed

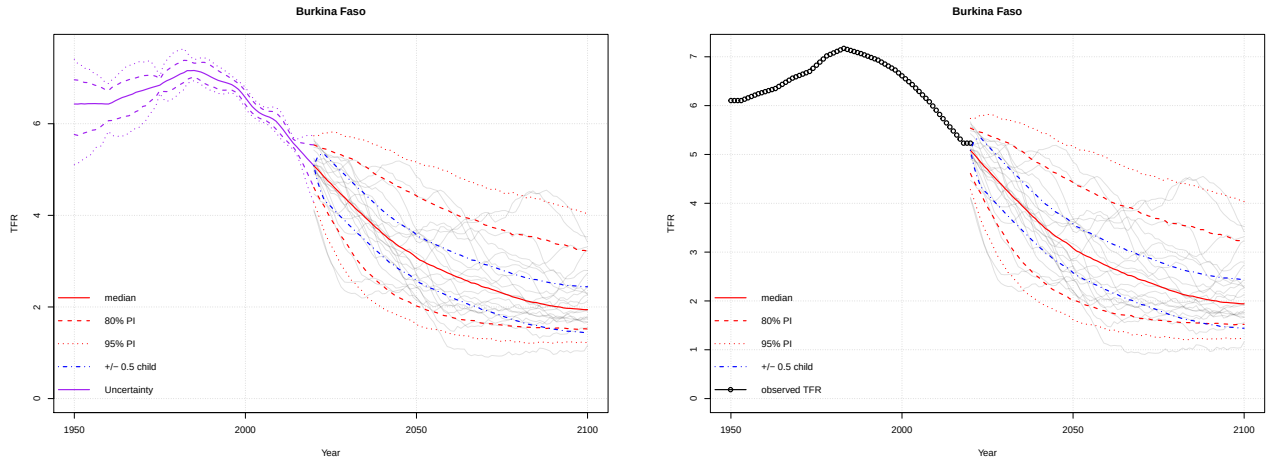


Figure 3.2: TFR prediction (from a converged simulation) for Burkina Faso with uncertainty about the past TFR (left panel) and without it (right panel). The black dots in the right panel represent the TFR used for initializing the simulation.

TFR. In this case, there could be a discontinuity between the last observed and the first projected time period.

These new arguments are also accepted by the `tfr.trajectories.plot.all` function which generates projection plots for all countries at once, as described by Ševčíková et al. [2011].

TFR predictions in a tabular format can be explored using the `tfr.trajectories.table` and `summary` functions which work the same way as in the previous versions of the package, except that in the former case, the output now contains uncertainty information about the past.

```
R> tfr.trajectories.table(pred, country = "Burkina Faso")

median    0.025      0.1      0.9    0.975 -0.5child +0.5child
1950 6.260554 5.796712 5.950538 6.628323 6.864889        NA        NA
1951 6.247358 5.836737 5.966940 6.608714 6.815591        NA        NA
1952 6.263510 5.837363 5.983308 6.601735 6.812113        NA        NA
```

```
1953 6.272365 5.876240 6.002074 6.602380 6.806366        NA        NA
1954 6.284774 5.874250 6.006758 6.621279 6.836863        NA        NA
...
2095 1.971179 1.273660 1.555806 3.469681 4.344101  1.471179  2.471179
2096 1.968939 1.293491 1.551197 3.458955 4.271590  1.468939  2.468939
2097 1.974058 1.307082 1.554862 3.372069 4.239252  1.474058  2.474058
2098 1.961362 1.334648 1.552755 3.367219 4.246914  1.461362  2.461362
2099 1.951146 1.324320 1.552732 3.366420 4.272629  1.451146  2.451146
2100 1.948056 1.337980 1.549511 3.329382 4.228294  1.448056  2.448056
```

R> summary(pred, country = "Burkina Faso")

```
Projections: 80 ( 2021 - 2100 )
Trajectories: 1000
Phase II burnin: 2100
Phase II thin: 9
Phase III burnin: 2100
Phase III thin: 9

Country: Burkina Faso

Projected TFR:
mean    SD 2.5%   5%  10%  25%  50%  75%  90%  95% 97.5%
2020 5.00 0.663 3.94 4.01 4.15 4.44 4.98 5.61 5.94 6.04  6.19
2021 4.90 0.741 3.71 3.79 3.96 4.28 4.89 5.56 5.94 6.07  6.21
2022 4.81 0.809 3.46 3.60 3.77 4.13 4.80 5.51 5.92 6.09  6.26
2023 4.71 0.869 3.22 3.40 3.59 4.00 4.69 5.44 5.90 6.13  6.31
...
```

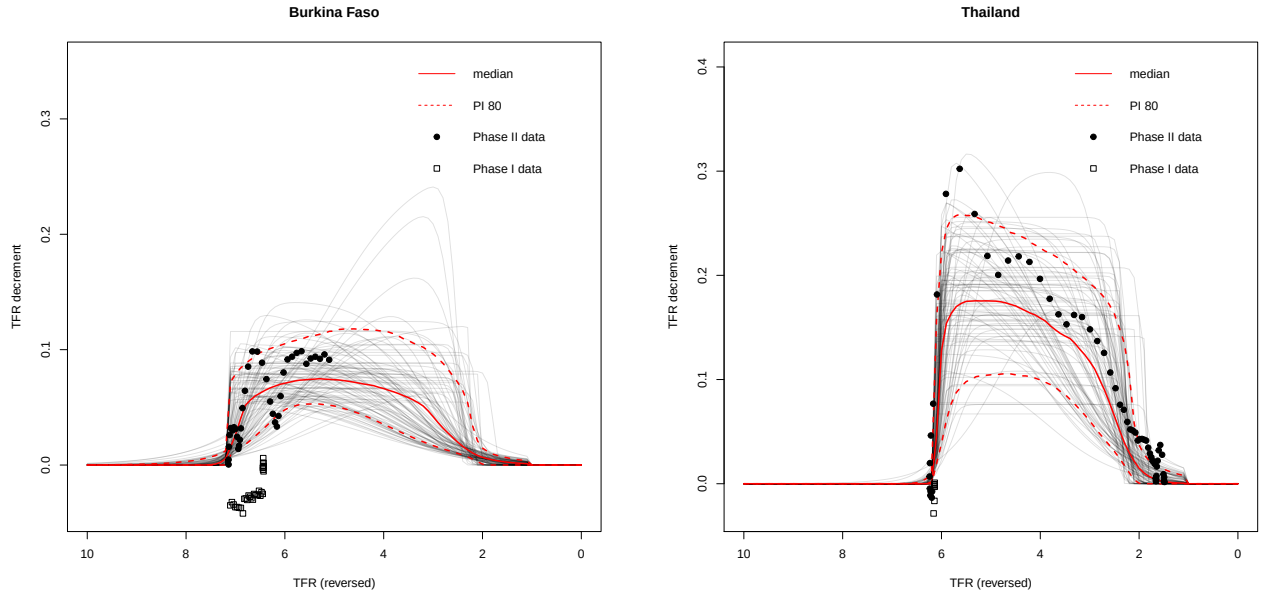


Figure 3.3: Estimated Double Logistic curves (from a converged simulation) for Burkina Faso (left panel) and Thailand (right panel). The data points (black dots and squares) are the estimated median decrements per year.

Exploring double logistic function

The double logistic function defined in (3.3) can be viewed using

```
R> DLcurve.plot(country = "Burkina Faso", mcmc.list = m, burnin = burnin,
+    pi = c(95, 80), nr.curves = 100)
```

Results can be seen in the left panel of Figure 3.3, while the right panel shows the result of the same call with `country = "Thailand"`.

If a simulation contains information about past uncertainty, then the Phase II and I data (black dots and squares) represent decrements of the estimated TFR median. In case of an annual simulation, these are annual decrements, otherwise these would correspond to five-year decrements. If the projections were produced without taking past uncertainty into account, then the data points represent the observed decrements.

This also applies to the `DLcurve.plot.all` function which plots the double logistic curves for all countries at once.

MCMC traces, density and diagnosis

To explore traces of the MCMC parameters, the existing functions `tfr.partraces.plot` (for country-independent parameters) and `tfr.partraces.cs.plot` (for country-specific parameters) can be used. Similarly, for density plots, `tfr.pardensity.plot` and `tfr.pardensity.cs.plot` are available.

As mentioned previously, there are two additional parameters in this version of the package, namely "rho_phase2", which is country-independent and defined in model (3.11), and "tfr" which is a country-specific parameter. These two parameters can be used within the aforementioned functions, like any other parameters .

For example, the trace plots and the density plots of ϕ and Nigeria's TFR estimate in year 1985 (as shown in Figures 3.4 and 3.5) can be visualized via

```
R> tfr.partraces.plot(m, par.names = "rho_phase2", nr.points = 200)
R> tfr.partraces.cs.plot(m, country = "Nigeria", par.names = "tfr_36",
+   nr.points = 200)
R> tfr.pardensity.plot(m, par.names = "rho_phase2", burnin = burnin)
R> tfr.pardensity.cs.plot(m, country = "Nigeria", par.names = "tfr_36",
+   burnin = burnin)
```

To check if the MCMCs have converged and adequately explored the parameter space, the `tfr.diagnose` function can be used; see Ševčíková et al. [2011] for more details. In the case of one-step estimation, the function checks parameters from Phase II as well as Phase III. In the case of two-step estimation, one would use `tfr.diagnose` for assessing the convergence of Phase II parameters, and `tfr3.diagnose` for assessing the convergence of Phase III parameters. Both functions accept a logical argument `express` which can disable or reduce the checking of country-specific parameters in order to speed up the process.

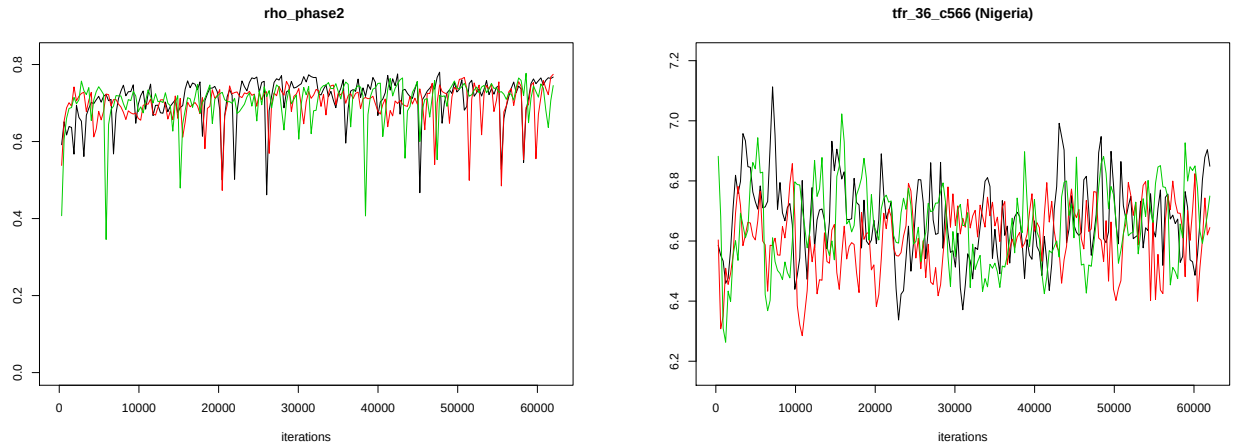


Figure 3.4: Trace plots for ϕ (left panel) and TFR of Nigeria in 1985 (right panel).

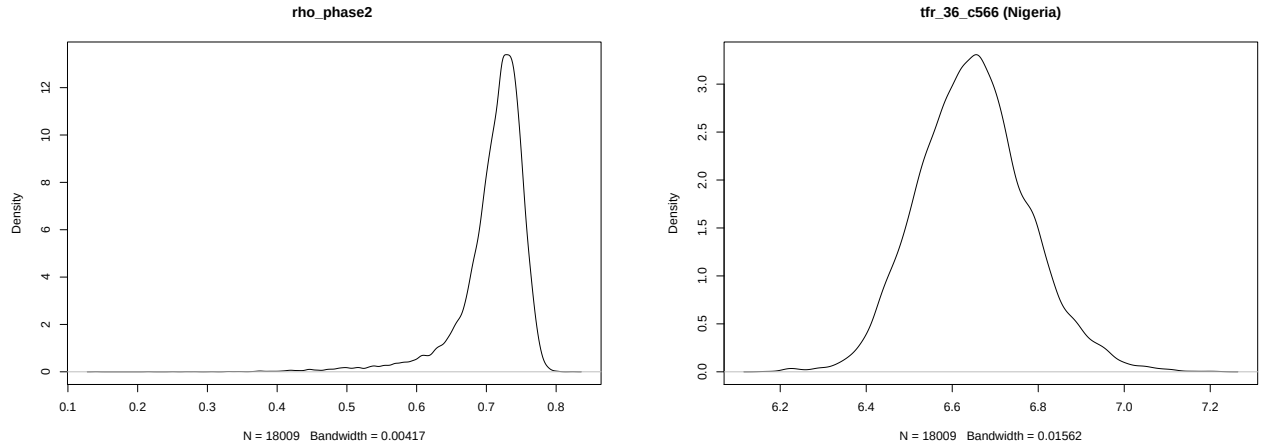


Figure 3.5: Density plots for ϕ (left panel) and TFR of Nigeria in 1985 (right panel).

If the estimation includes uncertainty about the past, the assessment of country-specific parameters include the "tfr" parameter for each country and time period, in our case more than 14200 "tfr" parameters. In practice, it is often impossible to achieve convergence for every single one of them. Thus, we introduced the rule of accepting the "tfr" parameters as having converged if 95% of them have converged.

To apply the convergence diagnostics to our simulation, one could do

```
R> tfr.diagnose(simu.dir.unc, thin = thin, burnin = burnin, express = TRUE)
```

As mentioned earlier, the MCMCs in our illustrative code examples have not been run for long enough to achieve full convergence. See Section 3.4.2 for alternative settings. Note that the toy simulation we proposed earlier cannot be checked for convergence, as there is a requirement of a minimum number of iterations per chain, which the toy simulation does not satisfy.

3.3.5 Estimating a small set of countries

The Bayesian framework we have shown so far is designed to estimate all countries of the world at once, where the historical experience of an individual country influences the distribution of its own parameters as well as of the world parameters, while using the same settings for all countries. However, this is not always practical for several reasons:

1. Analysts might want to experiment with settings for individual countries without waiting several hours for a simulation of the whole world to finish.

2. Different sets of covariates might be needed to estimate different countries.

3. Countries with unusual historical patterns or very small countries might be excluded from the simulation in order not to bias the world parameters. In practice, this will apply to countries with less than 100,000 people in the most recent observation.

It was the last reason, as well as the need to include aggregations in the estimation, that motivated us to implement the `run.tfr.mcmc.extra` function in the original version of the package. The idea is that while `run.tfr.mcmc` updates all parameters, the `run.tfr.mcmc.extra` function updates only the country-specific parameters of the specified countries, while re-using the existing distribution of the global parameters.

Since the function was designed for special cases of countries or aggregations, the original implementation allowed the user to process only the locations that have not been included in the world simulation. With the two additional use cases above, we have now relaxed that restriction and made it possible to rerun and overwrite existing estimations of country-specific parameters and past TFR estimates for individual countries, while allowing the user to change various estimation settings. However, several global settings are not subject to change, such as switching between annual and five-year estimation, or changing the `ar.phase2` argument.

Before we are going to describe an example use of this function, we need to clarify that the usage of this function should be based on experts who have clear understanding of the dataset they want to replace for specific countries. The example below is for illustration purpose.

Suppose that after running the simulation with the default data from the World Fertility Data, the user wishes to experiment with their own data that exclude Nigeria's questionable data points, namely the Indirect DHS-NS data points identified in Table 3.4 as having unreasonably low standard deviations and biases. Unlike in the main simulation, the experiment will not force the VR data of the United States to have zero bias and variance. For that purpose, we will extract data for Nigeria (code 566) and the USA (code 840) from the default raw dataset discussed in Section 3.3.1, remove the Indirect DHS-NS points for Nigeria and store them into a file called "raw_tfr_user.csv":

```
R> countries <- c(566, 840)
R> myrawTFR <- subset(rawTFR, country_code %in% countries)
R> myrawTFR <- subset(myrawTFR,
```

```
+   !(country_code == 566 & method == "Indirect" & source == "DHS-NS"))
R> write.csv(myrawTFR, file = "raw_tfr_user.csv", row.names = FALSE)
```

For experimentation with the `run.tfr.mcmc.extra` function, we recommend copying the main simulation into a different directory and applying the function to the copy. This is because the processing overwrites the existing estimation results, and thus there is no way back to the original results in case the experiments do not yield satisfactory outputs. Here we will append "_extra" to the directory name stored in `simu.dir.unc` and copy the content from `simu.dir.unc` into it. This step is equivalent to the command `"cp -r annual_unc annual_unc_extra"` on unix-based systems:

```
R> simu.dir.extra <- paste0(simu.dir.unc, "_extra")
R> dir.create(simu.dir.extra)
R> file.copy(list.files(simu.dir.unc, full.names = TRUE), simu.dir.extra,
+   recursive = TRUE)
```

To run the new estimation for the two selected countries, we can do

```
R> run.tfr.mcmc.extra(sim.dir = simu.dir.extra, countries = countries,
+    iter = total.iter, burnin = burnin, uncertainty = TRUE,
+    my.tfr.raw.file = "raw_tfr_user.csv", covariates = c("source", "method"))
```

We recommend using the same `total.iter` and `burnin` as in the main simulation.

To compare the new estimation results to those shown in Figure 3.1 we again use the `tfr.estimation.plot` function, now passing `simu.dir.extra` into the `sim.dir` argument. It can be seen in the left panel of Figure 3.6 that excluding the Indirect DHS-NS data points for Nigeria changed the estimates, especially for 1979. The uncertainty increased for the United States (right panel of Figure 3.6), since it was removed from the `iso.unbiased` set.

Finally, the option `uncertainty = TRUE` can be used even in two-step estimation where uncertainty about the past was not taken into account. This is possible because we do not

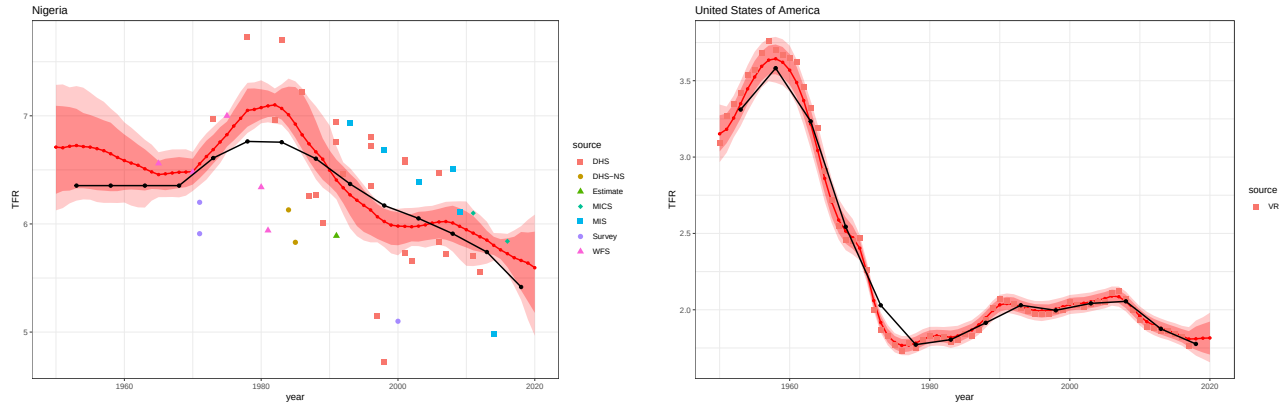


Figure 3.6: TFR estimation for Nigeria (left panel) and the United States (right panel), resulting from a non-converged simulation with modified data set.

expect the global parameters to be significantly different in the two situations (i.e. with and without uncertainty).

3.3.6 Structure of the output directory

Having a look at the simulation directory, here "annual_unc", one should see a structure similar to the following:

```
annual_unc
---- bayesTFR.mcmc.meta.rda
---- diagnostics
---- mc1
---- mc2
---- mc3
---- phaseIII
        ---- bayesTFR.mcmc.meta.rda
        ---- mc1
        ---- mc2
        ---- mc3
```

```
---- predictions
---- thinned_mcmc_9_2100
        ---- bayesTFR.mcmc.meta.rda
        ---- mc1
```

The directories "mc1", "mc2" and "mc3" on the first level are generated by the `run.tfr.mcmc` function and contain results from the three chains of the Phase II estimation. Each of the directories contains one text file per parameter. The names of the hyperparameters and their corresponding notation are the same as described in Table 1 in Ševčíková et al. [2011]. In addition, the parameter "rho_phase2" representing ϕ from Equation 3.11 is also stored as an hyperparameter if the Phase II-AR(1) is considered. The names of the files storing country-independent parameters consist of the parameter name and the suffix ".txt", while in the case of the files storing country-specific parameters the parameter name is followed by the suffix "_country*code*.txt".

If uncertainty is taken into account, the MCMC algorithm also generates estimates for the past TFR data. These samples are considered as country-specific parameters, called "tfr", and thus stored in files "tfr_country*code*.txt". They contain matrices of size the number of (thinned) iteration times the number of time periods. In the example above, the default starting year is 1950, and the present year is 2020, i.e. 71 years. Therefore, each file contains TFR estimates in 5100 rows and 71 columns.

The file "bayesTFR.mcmc.meta.rda" on the first level stores meta information about the Phase II estimation, which is contained in the `m$meta` object. If uncertainty is taken into account, the raw data used to obtain the estimates of TFR are stored as an additional element, called `raw_data.original`. A logical element `ar.phase2` indicates whether the autoregressive component of Phase II is considered in the estimation. In order to allow users to work with different subsets of countries with the same base of global estimates, information indicating whether the countries were processed separately has been also stored in the `meta` object. It is accessible via the `extra` element, created only if the `run.tfr.mcmc.extra`

function has been invoked and if `uncertainty` is `TRUE`. Here, `extra_iter` and `extra_thin` are used to retrieve the settings for specific countries. The raw data in this case are stored in a list called `raw_data_extra`. It is overwritten every time `run.tfr.mcmc.extra` is called for the same country.

The results of Phase III are stored in the directory "phaseIII". It has the same structure as described above. It is generated either by the `run.tfr.mcmc` function in case of a one-step estimation, or by the `run.tfr3.mcmc` function, in case of a two-step estimation. The meta file contains meta information related to the Phase III estimation. In the "mc$x$" directories, the names of the hyperparameters and their notations for Phase III are listed in Table 3.5. Similarly, the country-specific parameters and their notations are listed in Table 3.6. All files in this case contain one value per (thinned) iteration. Note that the country-specific parameters for Phase III are only estimated for countries which are already in Phase III, which is in our case 41 countries.

$\bar{\mu}$	$\bar{\rho}$	σ_μ	σ_ρ	σ_ε
`mu`	`rho`	`sigma.mu`	`sigma.rho`	`sigma.eps`

Table 3.5: Country-independent parameters for Phase III in model 3.7, with their corresponding names in the code. They can be obtained using `tfr3.parameter.names()`.

μ_c	ρ_c
`mu.c`	`rho.c`

Table 3.6: Country-specific paramters for Phase III in model 3.7, with their corresponding names in the code. They can be obtained using `tfr3.parameter.names.cs()`.

The "predictions" directory is created by the `pop.predict` function and it holds binary files, one per country, each containing the predicted TFR trajectories for that country.

Other convenience directories might have been created for speeding up processing. For example, the "thinned_mcmc_9_2100" directory was created by `pop.predict` to hold the

final chain for each parameter derived by applying the burnin, thinning and collapsing the three chains into one, in order to generate the predictions. Since we asked to generate 1,000 posterior TFR trajectories with burnin of 2,100 iterations, a thinning of 9 was applied to retrieve those trajectories: $3 \cdot (5,100 - 2,100)/9 = 1,000$. Thus, the parameter files in the "mc1" subdirectory here all contain 1,000 rows. Note that these values will differ when working with a toy simulation.

If functions for convergence diagnostics have been used, the simulation directory contains a folder "diagnostics" which holds results from these runs, one file per unique combination of thin and burnin.

3.4 Experiments

We have shown how the updated **bayesTFR** package can handle different versions of the TFR projection model. In this section, we will present results of experiments under different settings and discuss the implications of these settings. Based on those experiments we will give recommendations for a reasonable configuration of the model. Finally, we will discuss future directions in the development of the package.

3.4.1 Experiments with Settings

The new version of **bayesTFR** allows to handle different types of modeling needs, summarized in Table 3.2. An analyst can choose between a five-year and an annual model, as well as between accounting for past uncertainty or not. Flexibility is added by allowing the user to treat vital registration (VR) records for selected countries as unbiased, as well as using the autoregressive component in Phase II.

However, a question of consistency of results between the various settings may arise. For example, a forecast should not change dramatically when switching from five-year to annual data. Currently, there are no annual observations collected for all countries, and only a few countries (such as New Zealand) have good annual vital registration data, the only available annual observations. Thus, if past uncertainty is not taken into account the model would be

estimated on some version of interpolated data for most countries.

Countries in Phase III with good records

The first major difference can be seen for countries in Phase III, especially for countries with high quality VR records. We take Switzerland as an example. The left panel of Figure 3.7 shows TFR projections for a five-year model without accounting for past uncertainty (cell D in Table 3.2), while the right panel shows results from an annual model with uncertainty about the past (cell A in Table 3.2). It can be seen that the results on the right yield wider probability intervals, especially for the lower bound of the 95% quantile. For countries like Switzerland, the bias and uncertainty of past estimation is very low. Since the estimating process takes the linear intepolated TFR as the reference, the process can add extra bias to these data. Even though this is not large, the uncertainty propagated from the beginning of the forecast period could lead to a large difference.

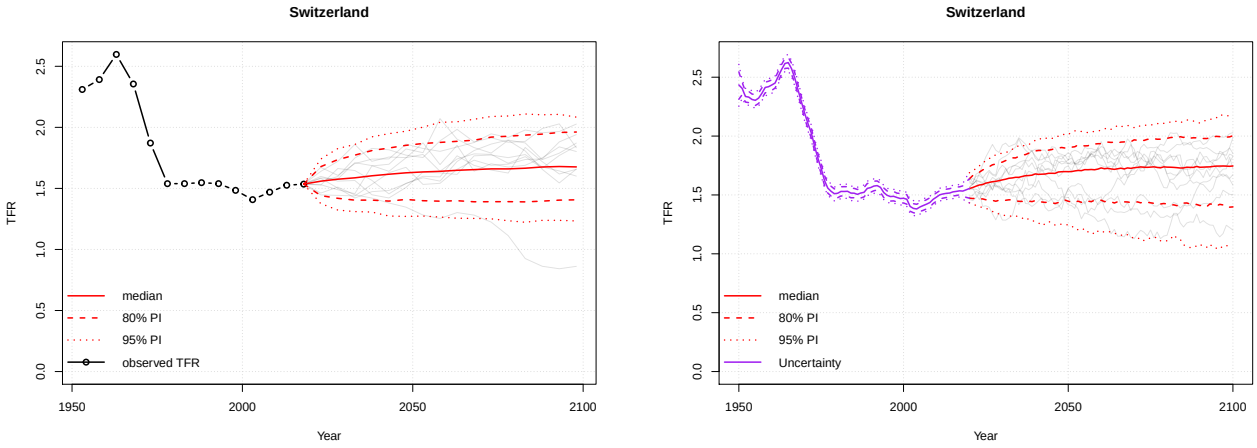


Figure 3.7: TFR predictions for Switzerland. Left panel: Original five-year model without accounting for past uncertainty. Right panel: Annual model with past uncertainty.

Now we consider the VR records for a set of selected countries (OECD and some developed countries as unbiased; the list can be found in the Appendix). The corresponding TFR projections for Switzerland are shown in the right panel of Figure 3.8.

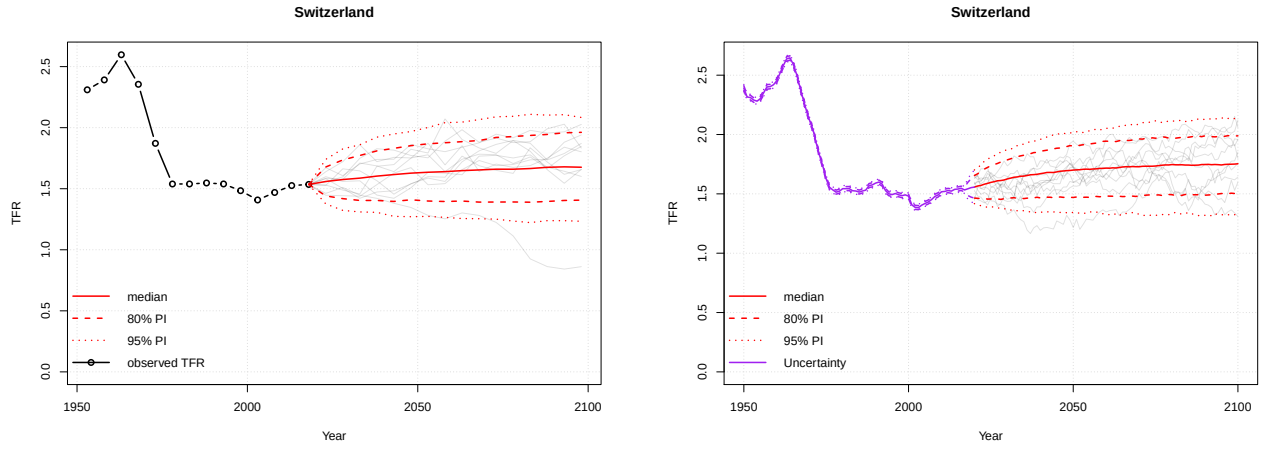


Figure 3.8: TFR prediction of Switzerland. Left panel: Original five-year model without accounting for past uncertainty. Right panel: Annual model with past uncertainty, with assuming VR records of selected countries (including Switzerland) as unbiased.

It can be seen that when compared to results from a five-year model (left panel), the differences between the two sets of projections are negligible. It is important especially for countries with perfect historical data, such as Switzerland, that similar results be obtained whether annual or five-year data are used.

Countries in Phase II

The second major difference relates to countries in Phase II, such as Nigeria. Figure 3.9 shows the difference between a projection resulting from a five-year model without accounting for past uncertainty (left panel) and from an annual model with uncertainty about the past without applying the Phase II-AR(1) component.

It can be seen that if we account for uncertainty and use annual data, the prediction shows a faster decline. Without performing an out-of-sample validation, it is impossible to say which of these projections is better. Nevertheless, a more detailed analysis revealed that the median of the residuals $\varepsilon_{c,t}$ for all countries in model (3.2) is highly autocorrelated. Figure 3.10 summarizes the estimates.

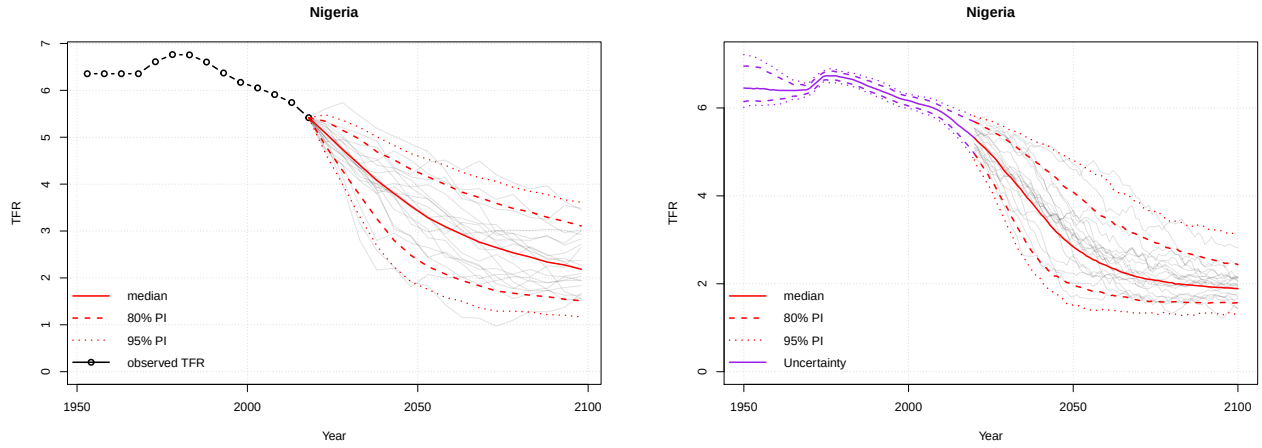


Figure 3.9: TFR prediction of Nigeria. Left panel: Original five-year model without accounting for past uncertainty. Right panel: Annual model with past uncertainty without Phase II-AR(1).

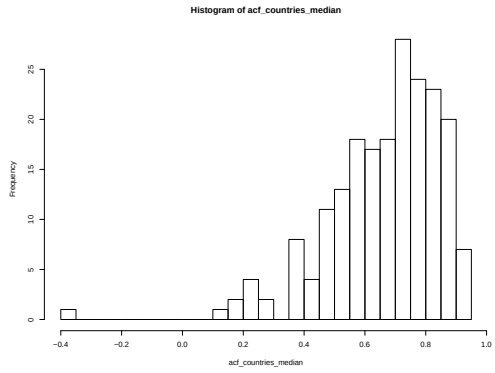


Figure 3.10: Histogram of autocorrelation for median Phase II residuals of all countries.

This suggests including the autocorrelation process in the modeling as defined in Equation (3.11). Figure 3.11 summarizes the differences. If we compare Figure (b) and (c), we found the decline has become slower, which is more in line with the five-year projections.

It can be seen however, that the starting point of the projections (year 2020) is now lower, and in fact it is significantly lower than the current UN estimates (last data point in estimation of (a) and (c)). The standard deviation of ε in model (3.11) is less than 0.02, if the autoregressive component is included. This could be problematic, given that for developed countries with low TFR and relatively stable societies, the standard deviation of annual TFR changes is about 0.04. This is likely a result of a possible smoothing of the data. To remedy that, we introduce a new lower bound on the σ_0 parameter (argument `sigma0.min` in `run.tfr.mcmc`) of 0.04, which becomes the new default. Figure 3.12 shows the relevant differences.

If the lower bound on σ_0 is applied, the prediction yields wider probability intervals as well as a higher median (top right panel), which better matches the five-year forecast. The estimation in this case (bottom right panel) also shows a better match with the UN estimates for most recent periods, which is another argument for using the new default for `sigma0.min`.

3.4.2 Recommendations

We have shown the flexibility of the new version of **bayesTFR** which can incorporate different variation of the TFR model as well as being compatible with the extant version of the model. As one of the key components in population projections currently adopted by the United Nations, this is a key step for migrating population projections from a five-year basis to an annual one. The package is designed to support UN analysts in this process, as well as to give other researchers and practitioners a tool to generate their own projections.

In addition to incorporating past uncertainty of TFR in the forecast, and performing annual-based projections, the package has introduced two other important components, namely the ability to specify vital registration data as unbiased, and the autoregressive component in Phase II. In Section 3.4.1, we have described the reasoning behind these two

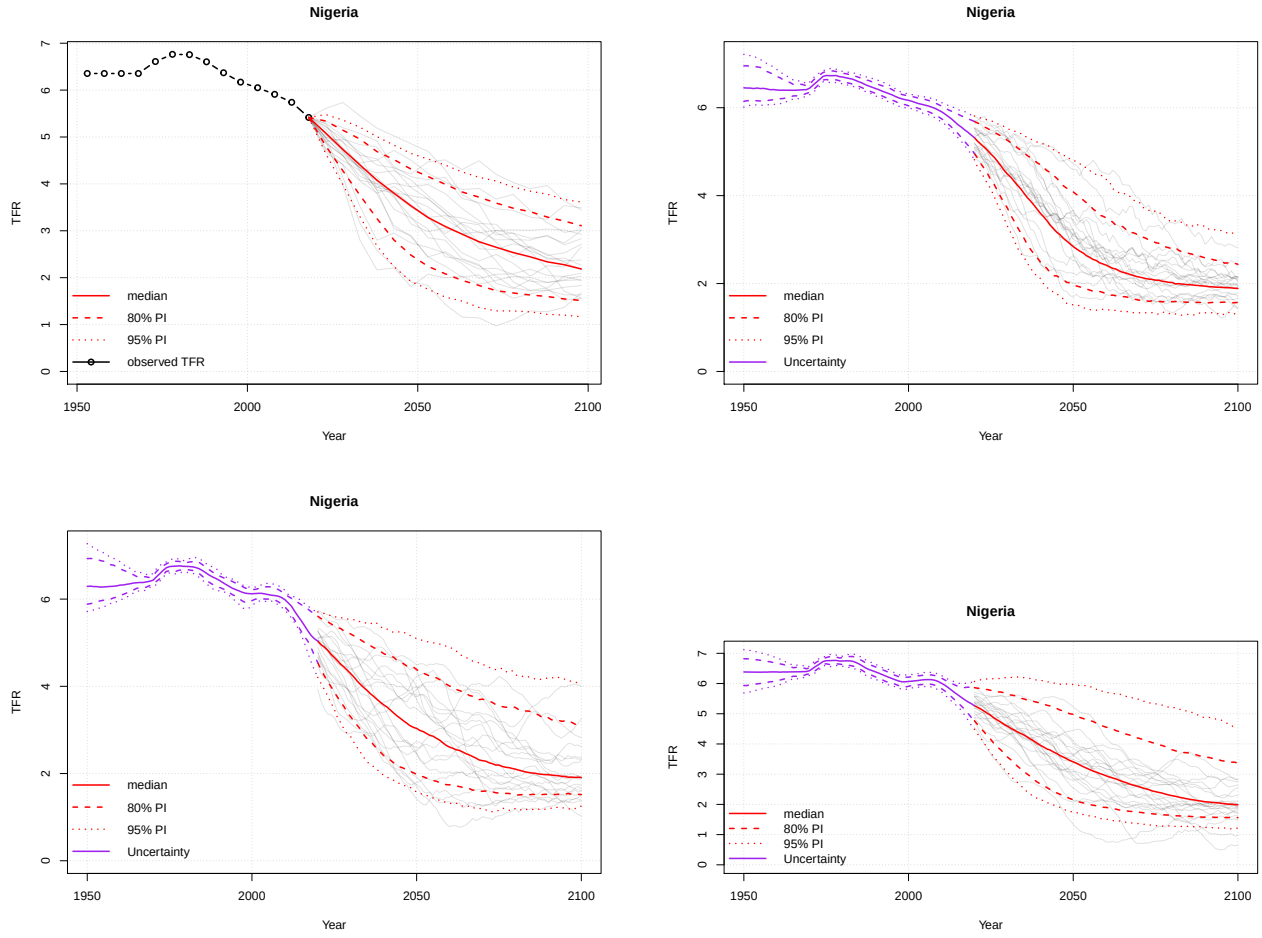


Figure 3.11: TFR prediction for Nigeria. Top left (a): 5 year version bayesTFR; top right(b): annual version bayesTFR without Phase II-AR(1); bottom left(c): annual version bayesTFR with Phase II-AR(1); bottom right(d): annual version bayesTFR with Phase II-AR(1), with new lower bound of $\sigma_{c,t}$.

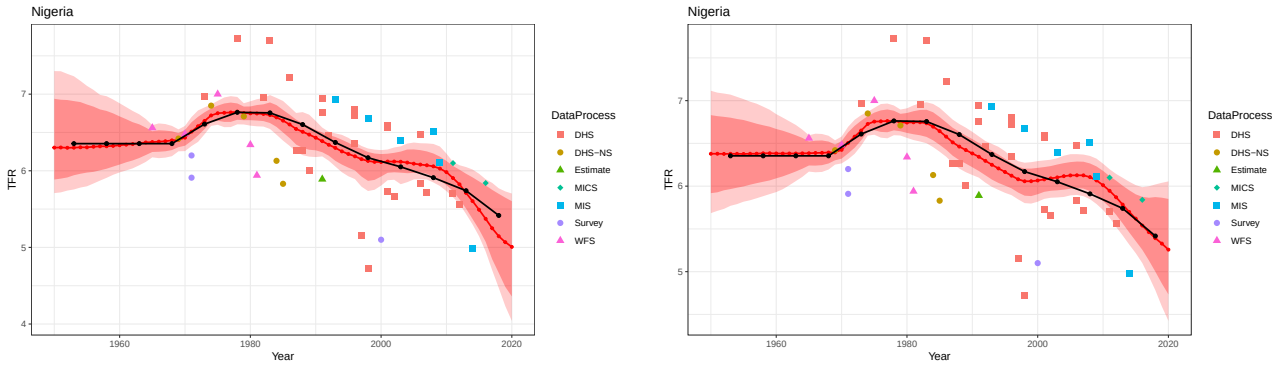


Figure 3.12: TFR prediction (top row) and estimation (bottom row) for Nigeria from an annual model with uncertainty with autoregressive component. Left column: without lower bound on σ_0. Right column: with `sigma0.min = 0.04`.

new options, as well as for setting a lower bound on the standard deviation.

Based on our experiments and analysis, when using the annual model with uncertainty about the past in a production-like setting, i.e. if full convergence of the MCMC algorithm is desired, we recommend the following settings:

```
R> annual <- TRUE
R> nr.chains <- 3
R> total.iter <- 62000
R> thin <- 10
R> burnin <- 2000
R> iso.goodvr <- c(36,40,56,124,203,208,246,250,276,300,352,372,380,392,
+    410,428,442,528,554,578,620,724,752,756,792,826,840)
R> m <- run.tfr.mcmc(output.dir = simu.dir.unc, nr.chains = nr.chains,
+    iter = total.iter, annual = annual, thin = thin, uncertainty = TRUE,
+    ar.phase2 = TRUE, iso.unbiased = iso.goodvr, parallel = TRUE)
R> pred <- tfr.predict(sim.dir = simu.dir.unc, end.year = 2100,
+    burnin = burnin, nr.traj = 1000, uncertainty = TRUE)
```

The ISO codes listed include most European countries, Australia, Japan, South Korea, New Zealand and the United States. These countries have a long history of vital registration with coverage rates often around 99%, indicating that their observations have been of high quality.

You should expect a full simulation with these settings to run for more than one day. Thus, we recommend processing it by a batch script in the background, so that it can be left unattended.

3.4.3 Discussion

In this chapter, we described the latest major update of the R package **bayesTFR**. This update significantly enriches the modeling framework considered in the existing version of the package, and gives analysts the flexibility to account for past TFR uncertainty, use annual data, and allow for an autoregressive model in Phase II. Moreover, by making use of the vectorization nature of R, the computational cost has been kept at a reasonable level while making the model more sophisticated. New functions for visualizing estimation results, as well as updated analysis tools will further support analysts in exploring the package outputs.

On the package development side, there are at least two major areas for future improvements. The first is modeling age-specific fertility rates with past uncertainty which is of interest to demographers. The second would be further vectorizing the MCMC process. If past uncertainty is included in the model, updating the estimates of TFR is the most time consuming part of the prcoess. Since we consider each past TFR per country and time period as a parameter, it adds over 14 000 parameters in the annual case. Thus, the speed of the Metropolis-Hastings step for updating TFR plays a big role in determining the overall speed of the method. If past uncertainty is not included, updates of country-specific parameters dominate the computing time, and thus are subject to further optimization.

On the modeling side, there are also two obvious directions for improvement. First, instead of modeling the bias and standard deviation based on linear regression for each country separately, these could be folded into the process, giving a fully united probabilistic

model. A pooled version could yield more robust estimates, especially given the small amount of data in some surveys. Another direction is related to the completeness of the VR data. The completeness of VR coverage is the most important factor for how precise the VR records are, and this is an important consideration for VR but not for other surveys. Due to the low bias of high quality vital registration systems, more research could be done on how to incorporate this information in the model.

Chapter 4

PROBABILISTIC FORECAST OF GLOBAL WARMING AND QUANTITATIVE ANALYSIS OF THE EFFECT OF PARIS AGREEMENT

4.1 Introduction

In December 2015, 195 parties, after decades of work, have finally signed the Paris Agreement within the United Nations Framework Convention of Climate Change [UNFCCC, 2015]. Major emitters, excluding Iran and Turkey, have agreed on strengthing the global response to the threat of climate change, in the context of sustainable development and efforts to eradicate poverty. Specifically, the Paris Agreement's long-term temperature goal is to keep the increase in global average temperature to well below 2°C above pre-industrial levels; and to pursue efforts to limit the increase to 1.5°C, recognizing that this would substantially reduce the risks and impacts of climate change.

As many researchers, journalists and government agents consider the agreement a historic milestone in the world's endeavour to tackle climate change, researches have been conducted on different aspects of this agreement had been conducted. For example, Pan et al. [2017] has explored the National Determined Contributions (NDCs) [United Nations Climate Change, 2018] of most countries for fairness and ambitions of their intended mitigation contributions. Jacquet and Jamieson [2016] has analyzed the overall impact of the Agreement, and described the agreement as soft but significant from political point of view. Moreover, Bodle et al. [2016] has discussed the "beyond-legal" nature of the Paris Agreement, while Spash [2016] has criticized that, by denying that tackling greenhouse gas emissions is incompatible with sustained economic growth, the Paris Agreement will change nothing. Moreover, many researchers tracked the performance of individual parties with signatures on the Agreement.

For example, Liobikienė and Butkus [2017] has tracked the performance of the European Union with deterministic methods, and Dong et al. [2018] has tracked the performance with deep neural networks methods for top 10 emitters.

Since the ultimate goal of all parties is to keep the increase in global average temperature to well below 2°C above pre-industrial levels, whether the Paris Agreement will be enough for this objective should be another important topic. Qualitative analysis, such as Bang et al. [2016], Young [2016], O'Brien [2018], Brown et al. [2019], provide some insights into key aspects and challenges concerning the agreement, and quantitative or partly-quantitative researches such as Rogelj et al. [2016], and Zhang et al. [2017], can help us understand how far the Paris Agreement is away from the ultimate goal of 2°C warming. Most researches are pessimistic about the realization of this goal, or at least they suggest that we need to do a lot more than the Paris Agreement for achieving the goal.

The Intergovernmental Panel on Climate Change (IPCC) has summarized most of these projections, in order to generate scenario-based climate change to 2100 [Intergovernmental Panel on Climate Change, 2014]. These forecast are based on assumptions of future input of socio-economic factors, such as emissions, and in the IPCC report, these factors have been summarized and concentrated to 4 representative scenarios, namely Representative Concentration Pathways. For emissions, these in turn are forecast using a version of the Kaya identity [Kaya et al., 1997], which expresses carbon emissions as a product of population, GDP per capita, and carbon intensity, the latter defined as the amount of carbon used to produce a given amount of GDP. The IPCC produced projections for each of four possible scenarios for future population, economic growth, and carbon intensity. The scenarios are explicitly not forecasts, and the scenario-based approach has been criticized as lacking validity by Moss and Schneider [2000], who called instead for a probabilistic statistically-based approach.

Raftery et al. [2017] developed a fully statistical probabilistic model for forecasting future carbon emissions, and hence future global temperature increase. They used the UN's then newly probabilistic projections of world population by country to 2100 [United Nations,

2015b]. Using data from 1960 to 2010, they developed a joint Bayesian hierarchical model for economic growth and carbon intensity, and obtained a resulting probabilistic forecast of future carbon emissions to 2100 using the Kaya identity. They translated this to global mean temperature increase using a relationship developed by the Intergovernmental Panel on Climate Change [2014]. They concluded that the probability of global mean temperature increase over pre-industrial levels being less than 2°C is only 5%, assuming a continuation of current trends.

This raises the question of what would need to be done to meet the goal of the Paris Agreement of keeping the increase to 2°C, or ideally to 1.5°C [UNFCCC, 2015]. To attempt to answer this, we updated our data to 2015, and developed a better method for translating carbon emissions to temperature change, as described in Section 4.2. This is based on the Coupled Model Intercomparison Project Phase 5 [Hurrell et al., 2011] (CMIP 5) ensemble of climate models, but takes better account of bias and measurement error in the models in the ensemble.

With this tool, we can address questions such as the following. Besides generating probabilistic forecast of the global mean temperature by the end of this century, we can also ascertain the probability that parties met their promises in the Paris Agreement. Moreover, with our tool, we can forecast the climate change probabilistically assuming the commitments on the Paris Agreement will be met, and also can analyze whether the goal set in the Paris Agreement is enough for the 2°C warming. Moreover, we can quantitatively assess the impact of the withdrawal of the USA from the agreement. With our analysis, there will be 26% probability that we could control the global warming by 2°C until 2100, given all parties can do what the promised and can continue these efforts until the end of this century, which is far better than the 5% with the current trend. We still need to do more, but implementing the Paris Agreement will be a good starting point.

This chapter is organized as follows. The data and the model specifications are described in Section 4.2. The model results, including the updated emission forecasts, evaluation of general circulation models and the probabilistic forecast of climate change are reported in

Section 4.3. In Section 4.4, we report the method's performance with validation studies. We then provide the detailed analysis of the Paris Agreement's effect on the emissions and climate change in Section 4.5. We conclude with a discussion in Section 4.6.

4.2 Methodology

4.2.1 Model Overview

In this chapter, we built the model in three steps. In the first step, we build a model for probabilistic forecast of CO_2 emissions. We used annual data on population, GDP and carbon emissions for each year from 1960 to 2015 for 161 countries containing over 99% of the world's population. We then build Bayesian time series models on CMIP5 model forecasts and historical simulations, together with the actual historical temperature anomalies, in order to obtain the estimates of the uncertainty and bias of CMIP 5 models. In the last step, we take the probabilistic forecast of CO_2 emissions from the first step as input, and link the cumulative emissions to the CMIP 5 models forecasts.

4.2.2 Kaya's Identity and CO_2 Emission Forecast

Data

In order to generate reasonable forecast of CO_2 emissions for all countries and regions, we use a simple form of the Kaya identity for each countries. This requires us to use the population data, GDP per capita data and the carbon intensity data.

For population, we used the UN's 2019 estimates of population for all countries from 1950 to 2015 [United Nations, 2019b]. We produced probabilistic projections for all countries with the model used by the UN for its probabilistic projections [Raftery et al., 2012].

GDP per capita data came from the Maddison Project, 2018 version [Bolt et al., 2018], using data from 1960 to 2015. This uses purchasing power parity (PPP) rather than market exchange rates, and provides two sets of GDP data, cgdppc for real GDP per capita in 2011US$ with multiple benchmarks, and rgdpnapc for real GDP per capita in 2011US$ with

a 2011 benchmark. According to the documentation, rgdpnapc is suitable for cross-country growth comparisons, while cgdppc is more suitable for cross-country income comparisons. Since we are trying to forecast GDP per capita, we use rgdpnapc for GDP.

CO2 emissions data came from the Global Carbon Budget [Le Quéré et al., 2018]. We used data from 1960 - 2015.

Kaya's Identity Model

In this chapter, we build models on population, GDP per capita, carbon intensity, and CO_2 emissions of each country. CO_2 emissions could be split up into: Population, GDP per capita, and carbon intensity (modeled as emissions per GDP), and:

$$\mathrm{CO_2} = \mathrm{Population} \times \frac{\mathrm{GDP}}{\mathrm{Population}} \times \frac{\mathrm{CO_2}}{\mathrm{GDP}}$$

For population, we used the UN's official population projections for all countries, which are probabilistic and also based on Bayesian hierarchical models for fertility and mortality. Currently, we are using the UN's official 2019 version, and generating forecast from 2015 to 2100.

For GDP per capita, we use a Bayesian hierarchical model for all countries by assuming there is a world frontier (the United States), towards which all countries may converge. The GDP per capita for the world frontier is modeled with a random walk with a constant positive drift, while the gap between the world frontier and all other countries GDP per capita is modeled with an autoregressive model. Specifically, we model GDP per capita of the world frontier (the United States) as:

$$F_t = F_{t-1} + \gamma + \gamma_{\mathrm{pre1973}} \cdot \mathbb{I}[t < 1973] + \varepsilon_t^f$$

and the GDP per capita gap between any other countries and the United States' as:

$$F_t - G_{c,t} = \phi_c(F_{t-1} - G_{c,t-1}) + \varepsilon^g_{c,t} \tag{4.1}$$

For carbon intensity, we model carbon intensity using a Bayesian hierarchical model using post-peak data, based on the observation that most countries have reached and passed a peak intensity. For each country, the intensity is modeled as a linear trend plus an autoregressive random process. Specifically, we model post-peak carbon intensity (CO_2 emission per unit of GDP) as:

$$\tau_{c,t} = \eta(t - \bar{t}) + \beta\tau_{c,t-1} - \delta_c + \varepsilon^\tau_{c,t} \tag{4.2}$$

In order to combine population, GDP per capita and carbon intensity together, an examination of model errors were conducted, and indicated no signifcant correlation between the model errors in population and the other two components. Thus, we build a joint Bayesian hierarchical model for GDP per capita and carbon intensity with correlation in model errors, and build models for population according to Raftery et al. [2017].

4.2.3 Global Mean Temperature Forecast

Model Overview

We use the existing General Circulation Models (GCMs), which represent physical processes in the atmosphere, ocean, cryosphere and land surface and are the most advanced tools currently available for simulating the response of the global climate system to increasing greenhouse gas concentrations.

The Coupled Model Intercomparison Project Phase 5 (CMIP 5, Hurrell et al. [2011]), is a standard experimental protocol for studying the output of coupled ocean-atmosphere GCMs. Since GCMs are based on representative concentration pathways (RCPs), forecasts from CMIP 5 models are also based on scenarios, which will be different from our probabilistic

forecasts.

Instead of running each GCM for each input trajectory of emission forecasts for probabilistic forecasts, which would not be feasible, we used the existing forecasts with different scenarios of emission forecasts and developed statistical models of the relationship between CO_2 emissions and model forecasts in CMIP 5 models. We found that the forecasts of each CMIP 5 model match linearly with the cumulative emissions for different RCP scenarios, with different correlation and scale. We took our CO_2 emission forecast as input, and used the linear relationship between global mean temperature and the cumulative emissions to generate probabilistic forecasts of future global mean temperature anomalies.

Data

In order to conduct the analysis discussed further on in this section, we used the CMIP 5 models forecast data as well as the historical surface temperature anomalies. For historical temperature measures, we used the HadCRUT4 database [Morice et al., 2012], a gridded dataset of historical surface temperature anomalies relative to a 1961-1990 reference period, used by the IPCC [Intergovernmental Panel on Climate Change, 2014]. Data are available for January 1850 onwards and are updated monthly. We based our forecast on the CMIP 5 model data [Taylor et al., 2012]. Each experiment on CMIP 5 models includes historical simulations back to 1860, and also provides estimates of future climate changes, either near-term until 2035 or long-term until 2100 or even 2300, under different scenarios.

Model

We generated forecasts based on each CMIP 5 model, and then combined these forecasts as an ensemble. For each model, our forecast is based on two parts. The first part takes CO_2 emissions trajectories as input, and uses the linear relationship between cumulative CO_2 emissions and global mean temperature predictions to generate the CMIP 5 model temperature forecast. The second part takes the uncertainty and bias of the CMIP 5 model

forecast directly, based on the difference between historical simulations and the historical temperature anomalies.

For each CMIP 5 model, the historical simulations are generated similarly under different scenarios. We denote the historical simulations by $x_{i,t}$, where i is the index for specific model and t is for year. We take the HadCRUT 4 observed temperature anomalies as the observations of the global surface-mean temperature, and denote them by $\tilde{y}_t$. Then we denote the difference between the CMIP model backcasts by $y_{i,t} = x_{i,t} - \tilde{y}_t$. We denote the true temperature anomalies by $\tilde{z}_t$. Then $z_{i,t} = x_{i,t} - \tilde{z}_t$ is the difference between the unobserved true global surface-mean temperature anomaly and the backcasts of the CMIP models backcasts. Analysis of the autocorrelation and partial autocorrelation functions of the $z_{i,t}$ time series shows that it is well represented for a first-order autoregressive, or AR(1) model. This leads us to specify the following model:

$$y_{i,t} = z_{i,t} + \delta_t \tag{4.3}$$

$$z_{i,t} = \rho_i z_{i,t-1} + \varepsilon_{i,t} \tag{4.4}$$

Here, $\delta_t = \tilde{y}_t - \tilde{z}_t$ are the measurement error of HadCRUT 4 observations, and $\varepsilon_{i,t}$ are i.i.d. white noise processes. Without further information, the distribution of δ_t and $\varepsilon_{i,t}$ are set as:

$$\delta_t \sim \mathcal{N}(0, V_t^2) \tag{4.5}$$

$$\varepsilon_{i,t} \sim \mathcal{N}(0, W_i^2) \tag{4.6}$$

where V_t^2 is the variance of the measurement error of HadCRUT 4 observations, and is provided in HadCRUT 4 dataset. δ_t is considered as time-independent, and δ_t and $\varepsilon_{i,t}$ are considered uncorrelated.

The forecast anomalies are linearly related to the cumulative emissions. Therefore, we collected the input cumulative CO_2 emissions, denoted by $c_{j,t}$ for all scenarios and year, and

the forecast anomalies from each model $x_{i,j,t}$, where i indexes models, j indexes scenarios and t indexes years. Then we model the historical simulations as:

$$x_{i,j,t} = b_i c_{j,t} + e_{i,j,t}. \tag{4.7}$$

For 36 out of 39 CMIP 5 models, R^2s of the regressions are higher than 0.9, which we could find more details in Section 4.3.3. Therefore, we could use model (4.7) to convert our CO_2 emission forecasts into the anomaly forecast. This part will also be elaborated in Section 4.2.5.

4.2.4 Estimation

The estimation process for the temperature model has three parts: estimating the model for CO_2 emissions, estimating the dynamic model for the bias and measurement error of the CMIP 5 models given by equation (4.3), and estimating the model of the connection between CO_2 emissions and CMIP 5 model forecasts given by equation (4.7).

For the model of CO_2 emissions, including the population forecasts and the joint model of GDP per capita and carbon intensity, we fitted our model using Markov Chain Monte Carlo (MCMC) sampling, as implemented in the JAGS package in the R programming language. Five chains were used and each chain was run for 100,000 iterations after 5,000 burn-in iterations, and the samples were thinned by 20. Trace plots and standard diagnostics indicated that the number of iterations was big enough.

For the dynamic model for the bias and measurement error of the CMIP 5 models (4.3), we also fitted our model with MCMC sampling, implemented in the JAGS R package. For this part, 3 chains were used and each chain was run for 10,000 iterations after 1,000 burn-in iterations. Lastly, the models of the connection between emissions and CMIP 5 forecasts were estimated by linear regression.

4.2.5 Model Forecast

Forecasting procedures are conducted with the following steps.

First we made probabilistic forecasts of the CO_2 emissions for all countries and regions, by making forecasts of population for all countries and forecasts of GDP per capita and carbon intensity jointly. We then drew samples of future population, and sampled jointly from the posterior predictive distribution of GDP per capita and intensity for all future years and countries. We then multiplied them together to obtain posterior trajectories of CO_2 emissions. This was repeated 1,000 times to obtain 1,000 posterior samples of future CO_2 emissions for all countries and time periods.

We then made forecasts of the CMIP 5 models based on our forecasted CO_2 emissions. For each trajectory, we calculated the global cumulative emissions from 2010, to be indicated by c_t. Then, for each CMIP 5 model, the model forecast of global mean temperature $x_{i,t}$ is calculated as $b_i c_t$, where b_i are estimated with the linear regression of the model (4.7).

We then forecasted the bias and uncertainty of the historical CMIP 5 estimates using equation (4.3). For each model i, we drew 1,000 trajectories from the posterior predictive distribution $z_{i,t}$ up to year 2100. Finally, we added the $x_{i,t}$ and $z_{i,t}$ forecasts for each trajectory and each CMIP model, and we repeated the steps 1,000 times to obtain 1,000 posterior samples of global mean temperature forecast up to year 2100.

The last step is to combine the forecasts from all CMIP 5 models. We assigned equal weights to the 39 CMIP 5 models. We sampled uniformly from the CMIP models to obtain model i, and given the sampled model i, we sampled one trajectory of $x_{i,t} + z_{i,t}$ from 1,000 posterior samples. We repeated these steps 1,000 times to obtain the final forecast distribution of the temperature anomalies.

4.3 Model Results

4.3.1 CO_2 Emission Forecast

We trained the model descibed in Section 4.2.2. Then the joint model is used to forecast the CO_2 emissions for more than 160 countries with at least 20 years of historical data of CO_2 emissions.

The resulting forecast of global CO_2 emissions is shown in Figure 4.1, along with the forecasts of the IPCC's four main scenariosIntergovernmental Panel on Climate Change [2014]. Adding the additional five years of population, economic and emissions data led to a decline in the median forecast for global annual emissions in 2100 to 34 Gt CO_2, or 8 Gt CO_2t lower than the previous forecast in Raftery et al. [2017]. This reflects slower growth in emissions in 2010–2015 than previously. Specifically, the yearly CO_2 emissions of the United States decreased by 5% from 2000 to 2010, but in 2010 to 2015 the speed of CO_2 emission decrease almost doubled by decreasing another 5%. The CO_2 emission of other major country China, which is the only country emissions currently larger than those the United States, increases by 158% in the first decade of the 21st century, but only increases by 15.7% from 2010 to 2015.

4.3.2 Estimation of CMIP 5 Model Bias

Past Data

As described in Section 4.2.3, after we got the probabilistic forecast of the CO_2 emissions, the next step for generating probabilistic forecast of global mean temperature anomaly should be quantifying the bias and uncertainty of the CMIP 5 models' forecasts. Thus, we analyzed the data of past CMIP 5 model simulations. The data is summarized in Figure 4.2

There are clear cuts between the first lag term and later ones of partial auto-correlation function (PACF) for the bias of CMIP 5 model simulation (compared with the HadCrut4 observations). These results could be found in Table 4.1. Results suggested that the error

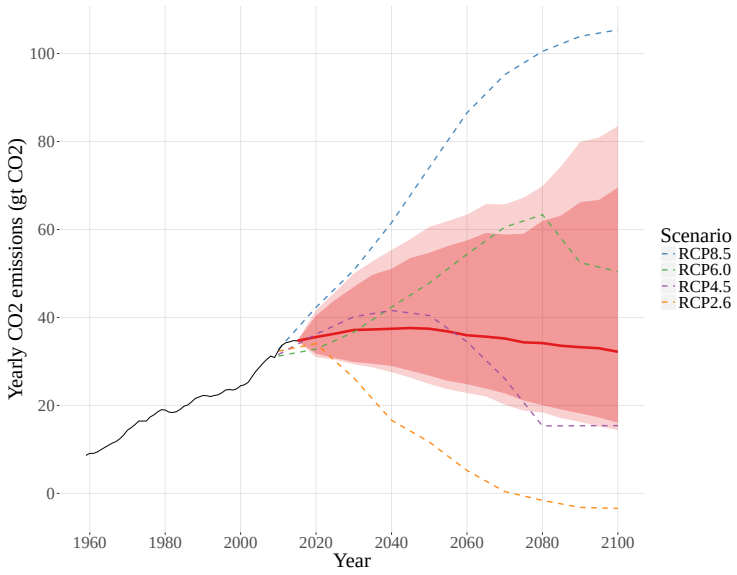


Figure 4.1: CO_2 Emissions are forecasted differently, based on the new data. The forecast median of yearly emission at 2100 is now 34 Giga tons around the world

should be modeled an $AR(1)$ process and suggested the model (4.4).

Table 4.1: Partial auto-correlation function of simulation errors. Among 39 models, all lag 1 PACFs are significant, while there are only 2 Lag 2 ones are significant, and 2 lag 3 ones are significant with 95% confidence interval. Since there are only 5% of models suggesting extra AR terms in the model, we choose to use AR(1) as a generic model for all CMIP5 models.

model	Lag(1)	Lag(2)	Lag(3)
ACCESS1-0	0.55*	0.07	0.09
ACCESS1-3	0.66*	0.01	0.26*
bcc-csm1-1-m	0.60*	0.16	0.12
BNU-ESM	0.62*	0.16	0.17
CanESM2	0.55*	-0.02	0.17

Estimated Historical True Anomaly

We estimated the model in Section 4.2.3 with MCMC process. We estimate the model with 1,000 burn-ins, 10,000 iterations with 3 chains, and the model converged well. The

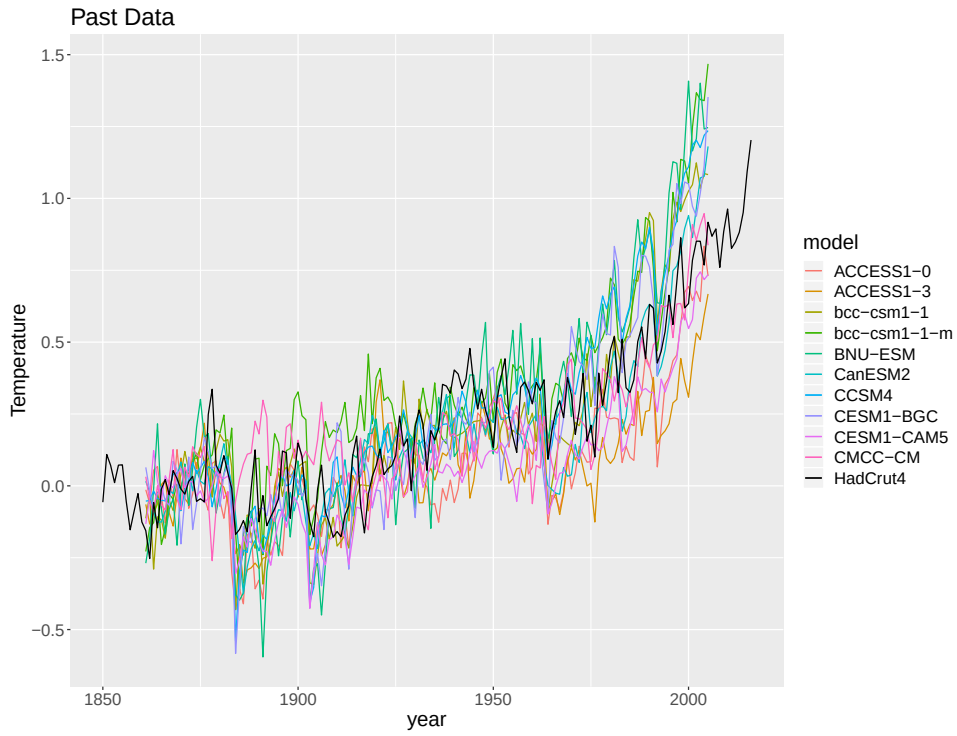


Figure 4.2: CMIP 5 models. Both model simulation and the real data are adjusted such that the mean anomaly between 1861-1880 for each model and real data is 0. The black line represents the real data, while all colored lines are for CMIP 5 model simulations. In total there are 39 CMIP 5 models, but we only include 10 models in this plot.

historical true anomalies are estimated and summarized in Figure 4.3:

With this model, we determined that bias of historical simulations is well modeled by an AR(1) model. As indicated by the red shaded areas in Figure 4.3, the uncertainty is wider in earlier periods, but is tighter in later periods, based on the HadCrut 4 data.

4.3.3 Global Mean Temperature Forecast

Global Mean Temperature Forecast with Individual CMIP Models

The next steps for probabilistic forecasting the global mean temperature, according to the description in Section 4.2.5, are to forecast the future errors of each CMIP model forecasts,

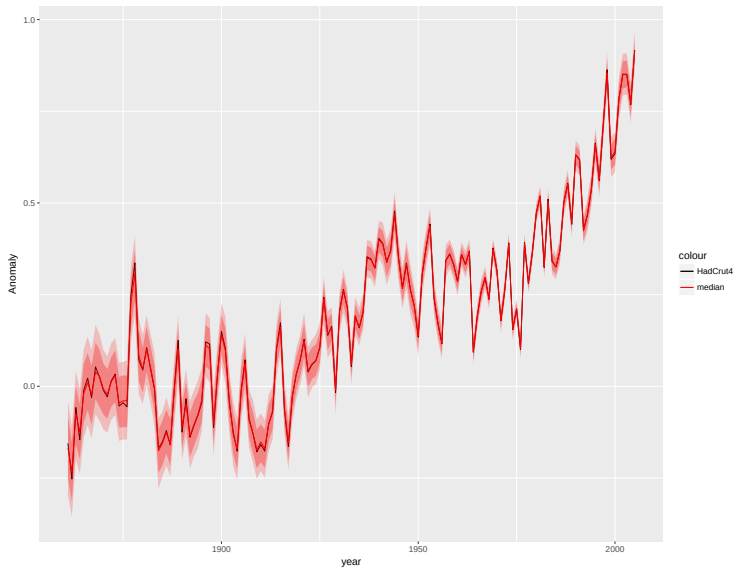


Figure 4.3: Estimation of historical anomalies. The black line represents HadCrut4 observations, while the red line and the shaded areas represents estimated median and 90%, 95% estimation intervals.

and to link the cumulative CO_2 emissions with each CMIP model forecasts. Finally, we add these two parts together to generate probabilistic forecast for each model.

In this section, we will take one model "ACCESS1-0" as an example. Forecasts with all models are presented in the supplementary materials, and the assembled forecasts are presented later on this section.

First consider the estimation of model (4.7) for this example. With the cumulative emission calculated with representative concentration pathways, Table 4.2 summarizes estimation of this lienar model.

Table 4.2: Estimation of Model 4.7 for ACCESS1-0, CO_2 with unit of 1,000 Giga-tons.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.473	0.015	31.537	0
Cumulative CO_2	0.626	0.005	119.130	0

The associated R^2 is 0.98, which suggests an almost perfect linear relationship between

cumulative CO_2 emissions and temperature anomalies.

For each iteration, we would forecast $z_{i,t}$ and $x_{i,t}$ independently. For $z_{i,t}$, we sample one trajectory $\tilde{z}_t$ of past estimation as well as ρ_i, W_i, and make forecast of future $z_{i,t}$ following the model (4.3). For $x_{i,t}$, we sample one trajectory of cumulative emissions c_t, and apply the estimated linear relationship in Table 4.2 with model (4.7). The forecast of $x_{i,t}$ and $z_{i,t}$ are shown in Figure 4.4. In the beginning of the forecast, the uncertainty is dominated by the random distortion of yearly temperature. As time increases, the uncertainty of temperature change is then dominated by the CMIP model forecasts, which is due to the uncertainty in CO_2 emissions forecast.

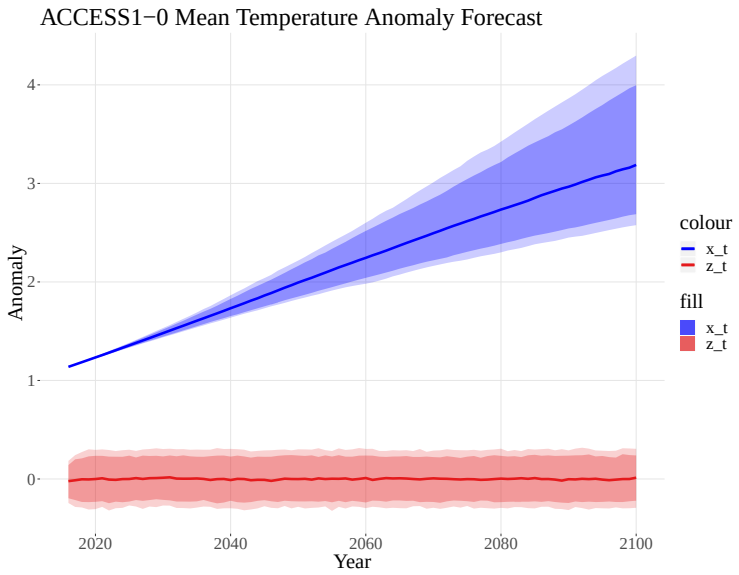


Figure 4.4: Forecast for $x_{i,t}$ and $z_{i,t}$ for ACCESS1-0 model. The solid lines represent predicted median while the shaded area represents 90% and 95% predictive interval.

Finally, subtracting of forecast $x_{i,t}$ and $z_{i,t}$ yields the temperature anomaly forecasts in Figure 4.5. We could figure out that the anomalies will reach 3.24°C, with 90% predictive interval as $[2.48°C, 4.05°C]$.

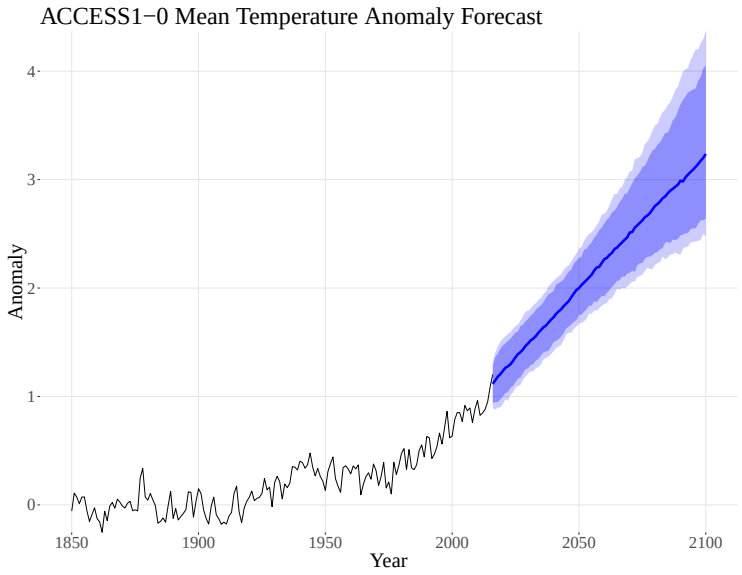


Figure 4.5: Forecast temperature with ACCESS1-0.

Assembled Forecast

Once we have the forecasts of global mean temperature for all individual CMIP models, finally we assembled them together by assigning equal weights on each models. Figure 4.6 shows the updated probabilistic forecast of global mean temperature increase from 2015 to 2100 based on current trends. The median temperature anomaly will become 1.9° degree around 2050, with 90% predictive interval as 1.3° to 2.5° degree. Moreover, The median forecast for 2100 is 2.8°C, with likely range (90% prediction interval) of [2.1, 3.9]°C. The median is 0.4°C lower than that of Raftery et al. [2017], and the upper bound is 1.0°C lower, while the lower bound is 0.1°C higher, with the tighter interval reflecting the additional five years of data and the improved model. The resulting forecast suggested that with the current trend, the global warming would be far greater than the 2°C.

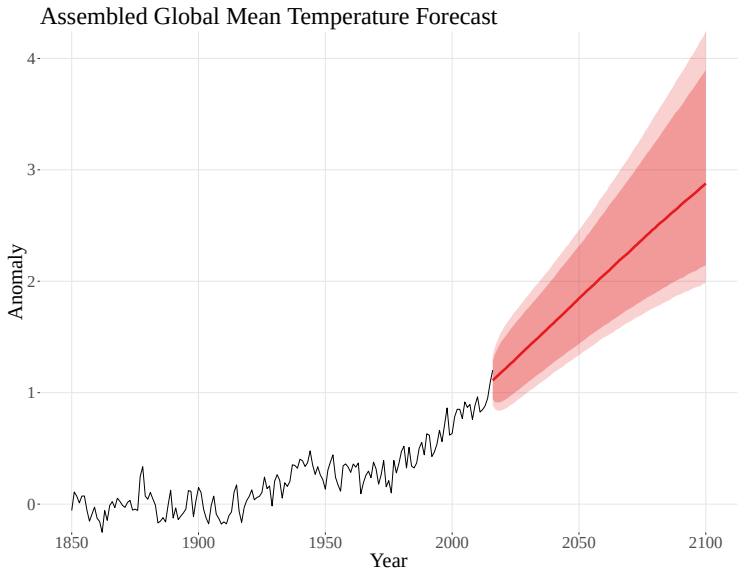


Figure 4.6: Probabilistic forecast of the global temperature anomaly to 2100. The black line represents the historical HadCRUT4 observations, while the red lines and the shaded area represents the forecast median, 90% and 95% prediction interval.

4.4 Model Validation

The previous work showed that the carbon emissions forecasting model validated well in terms of out-of-sample forecasts [Raftery et al., 2017]. It remains to assess the temperature forecasting model.

4.4.1 Study Design

We splitted our data into training (from 1960-2005) and validation sets (from 2006-2015). The inputs to the validation study consist of all GDP data, population data, carbon emission data, historical temperature data and the CMIP 5 model estimations and forecasts.

For each model output of CMIP 5 projects, we can estimate the temperature from 1860 to 2005, and, after 2005, we can extrapolate it. With our suggested methodology, we could not generate forecasts before 2005, because CMIP 5 models do not extrapolate before 2005, since in CMIP5 model simulations, RCP inputs before 2005 are considered accurate, and

same for all 4 scenarios, but inputs after 2005 are considered as forecast and thus different. Therefore, we would not know the CMIP5 “forecast" would be given CO_2 emissions different from RCP scenario inputs in the past. Thus, we could only conduct validation study with training data ending in 2005.

We conducted the out-of-sample validation study as follows:

- We trained the IPAT models with population, GDP and carbon emissions for all countries and regions with data before 2005.

- We estimated the bias and measurement error variance for all CMIP 5 models with data prior to 2005.

- We used the results from the estimated IPAT models to make the forecast of CO_2 emissions from 2006 to 2015.

- Then, we use the forecasted CO_2 emissions to generate the anomaly forecast from 2006 to 2015 for each CMIP model.

- We ensemble all anomaly forecasts to obtain the probabilistic forecast of anomaly from 2006 to 2015.

4.4.2 Validation Performance

Figure 4.7 shows the results of our model validation exercise. The mean absolute error on the test set is 0.107, and the observations fitted comfortably within the prediction intervals. The uncertainty is mainly due to the estimated measurement error variance for all CMIP 5 model forecasts. All of the observed anomalies in 2006 to 2015 are within 90% predictive intervals, and the 50% predictive interval covers 5 out of 10 observed anomalies, indicating that the interval is well calibrated.

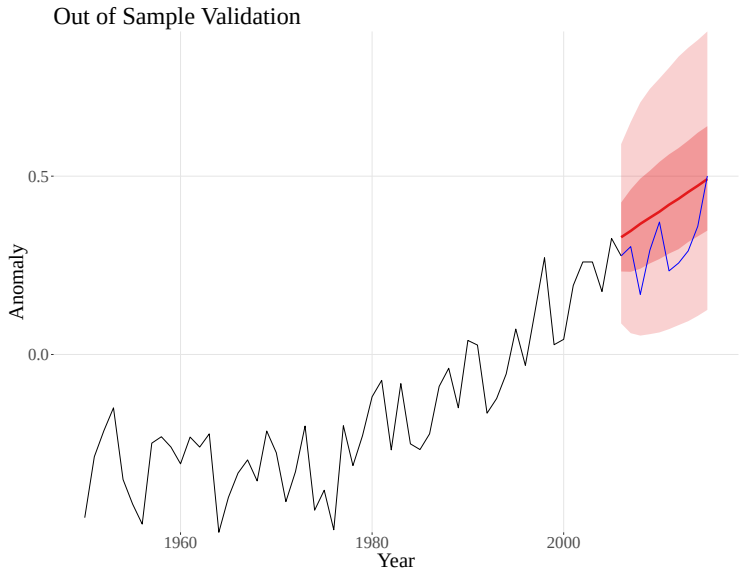


Figure 4.7: Out of sample validation plots. The black line is the historical anomaly observed in HadCRUT4 data base. The red line is the forecasted median, and the dark and light shaded area represents 50% and 90% predictive interval. The blue curve is the observed anomaly in 2006 to 2015.

4.5 Analysis of Paris Agreement

4.5.1 Overview

The Paris Agreement builds upon the United Nations Framework Convention and is an attempt to bring all nations into a common cause to undertake ambitious efforts to combat climate change and adapt to its effects, with enhanced support to assist developing countries to do so. The Agreement requires all Parties to put forward their best efforts through nationally determined contributions and to strengthen these efforts in the years ahead.

A major component of the Paris Agreement is the Nationally Determined Contributions (NDCs), that were promised by 185 of the 197 signatory countries [United Nations Climate Change, 2018]. A further 12 countries, such as the Philippines, submitted Intended NDCs, but have not yet formally ratified the Paris deal, and so they have not yet submitted NDCs. For these countries, we have taken their NDCs to be the same as their intended NDCs. In

the NDCs, countries made promises in cutting emissions directly, cutting emissions with respect to the unit GDP emissions, and making energy-related efforts such as increasing the percentage of energy usage from renewable sources. We excluded 50 of the 197 countries from our analysis because their promises of cuts in carbon emissions or intensity were unclear. For example, the United Arab Emirates promised to increase the share of “clean energy" in the energy mix to 24% by 2021, but it is unclear how that would actually affect their carbon emissions.

Since this is the first global response to the threat of the climate change, the quantitative analysis of the agreement is of great interest. In this chapter, we address several quantitative questions with respect to the Paris Agreement, including the following.

- What is the probability that major countries can achieve what they put in their NDCs?
- With the current analysis, we know that with the current trend, we will not keep the global temperature rise this century below 2°C. Assume all countries achieve their promises, what will be the global emission by the end of this century, and thus the climate change? Moreover, since the United States intends to withdraw from the Paris Agreement, we analyzed the effect on the global climate change of the United States separately.
- If we are still away from the objective (2 degree Celsius warming), how much more need to be done to achieve the objective?

In this Section, we address these three questions in the following sub-sections.

4.5.2 Summary of National Determined Contributions

As explained in Section 4.5.1, we will only analyze countries promising reductions on emission directly. These can be classified into direct emission cuts such as the United States, intensity cuts such as China, and emission cuts compared with the *Business-as-Usual* (BAU) scenario such as Afghanistan.

Of the 147 countries under analysis, 81, including 28 combined as the Europe Union, promised either direct reductions in emissions, such as the USA, or cuts in carbon intensity, such as China. The remaining 66 countries promised emissions cuts relative to the BAU scenario; in practice this often means limiting the increase in emissions rather than decreasing them. For example, Afghanistan reported their emissions as 28.8 Mt of CO_2 in 2005 and forecast their emissions to be 35.5 and 48.9 Mt in 2020 and 2030 respectively under the BAU scenario. They committed to a reduction of 13.6% relative to BAU for the year 2030, corresponding to 42.7 Mt, an *increase* of 20% over 2020. Most of the NDCs have 2030 as their target date, but some, such as the USA and Brazil, refer to 2025.

It's straightforward to set the goal for countries promising direct cuts or intensity cuts. For countries promising emission cuts under the BAU scenario, we found referenced year emissions according to their NDCs. Then, we took the ratio of the emission between the referenced year and the target year as their actual commitments (which could be increases in emissions). For countries without referenced year in their NDCs, we took the percentage reduction in emission compared with BAU levels as their promised intensity cuts.

Some countries promised a reduction within a specified range, and for these we took the lower bound reduction as their commitment. For countries promising extra reductions dependent on international supports, we assume that only the unconditioned part would be achieved. These are the most conservative assumptions and since the political negotiation between countries are extremely difficult, we believe this is the best we can expect.

4.5.3 Probability

We first address the question, what is the probability that each country will meet its NDC, given current trends? For addressing this question, we forecast the CO_2 emissions up to the target year, and calculate the proportion of trajectories below the target. This probability is shown in Figure 4.8. For most of the major emitters, the probabilities are low, such as the USA (2%), China (16%) and Japan (10%), Germany (13%) and France (2%). For some countries, however, such as Russia (93%), they are much higher.

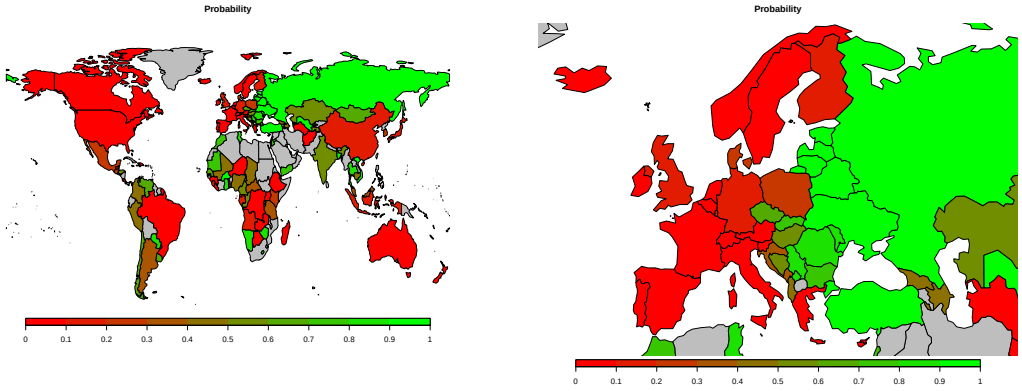


Figure 4.8: Probability that countries achieve their Paris Agreement Goals according to their nationally determined contributions (NDCs). **a.** All countries. **b.** European countries. The probabilities vary widely between countries, from values near 0 to values near 1. However, the probabilities are low for most major emitters (USA, China, European Union, Japan).

4.5.4 Climate Change with NDCs Achieved

Next we ask, what is the probability that warming will be kept to 2°C if all countries *do* meet their NDCs? We summarized the NDCs and the intended NDCs submitted to the Paris Agreement and calculating the target year emissions or intensity. The NDCs generally refer to 2030, with a few countries referring to 2025, while the 2°C target relates to 2100. Therefore, the answer to the question depends on how countries emission cuts are achieved, and on what happens after the NDCs are met, i.e. between 2030 (or 2025) and 2100. For years before they promised to achieve the emission cuts, we assume the intensity will have the same amount of extra reduction on every year. For years after their promised targeting year, we consider two scenarios for changing their emission forecasts. In one scenario, which we call the "Adjusted" scenario, countries revert to their pre-2015 trend after the NDCs are met, in most cases improving their carbon intensity levels, but at a slower pace than between 2015 and 2030. In the other scenario, which we call the "Continued" scenario, countries continue to improve their carbon intensity at the same rate until 2100 after they meet the NDC.

For two different scenarios, we adjust our forecast trajectories in order to make all fore-

cast trajectories to met their promises, and aggregate these forecasts to check the outcome of climate change. Specifically, for trajectory showing lower emission (or intensity) than promised, this trajectory will not be modified. However, if posterior trajectories show higher emission (or intensity) than promised, we modify that trajectory under different scenarios.

Specifically, for countries promising emission cuts in intensity such as China and India, suppose the promised intensity is J_T in year $2015+T$, and the forecasted intensity of that trajectory is $I_T(> J_T)$, then it means that we need extra annual intensity cut by level $a = 1 - \left(\frac{J_T}{I_T}\right)^{1/T}$ from 2016 to the target year. This will lead to the GDP per capita gap changing annually by a factor of $(1-a)^{\rho\sigma_{\text{GDP}}/\sigma_\tau}$ in expectation, where the ρ is the estimated correlation between intensity and GDP per capita, and σ_{GDP} and σ_τ are the estimated standard deviation of the random noise in GDP per capita and carbon intensity model (4.1, 4.2).

For countries promising emission cut directly, such as European countries, suppose the promised emission is D_T in year $2015+T$, and the forecasted intensity of that trajectory is $E_T(> D_T)$. We then assume the required extra intensity cut is a per year, so that at the target year $2015+T$, the intensity will be multiplied by a factor of $(1-a)^T$, while the frontier gap of GDP per capita ($F_t - G_{c,t}$ in equation (4.1)) will by multiplied by a factor of $(1-a)^{T\rho\sigma_{\text{GDP}}/\sigma_\tau}$. Since our goal is to make them meet their promises, we can determine the level a needed to get the required cut by solving

$$D_T = E_T(1-a)^{T-T\rho\sigma_{\text{GDP}}/\sigma_\tau} \tag{4.8}$$

For years between $2015+T$ to 2100, we will forecast each countries intensity and GDP per capita under two scenarios. Under “Adjusted" scenario, the intensity and emission forecast for each trajectory between $2015+T$ to 2100 is multiplied by the ratio of promised and forecasted intensity and GDP per capita at the target year $2015+T$. Under “Continued" scenario, for each year $2015+t$ between $2015+T$ to 2100, assuming the extra annual intensity cut is by level a, then the intensity forecast is multiplied by the factor of $(1-a)^t$, and the

forecast frontier gap of GDP per capita is multiplied by the factor of $(1-a)^{t\rho\sigma_{\text{GDP}}/\sigma_\tau}$.

With the rules above, the adjusted forecasts of two major countries, the United States and China are presented under different scenarios. The United States CO_2 emissions are the top among country making commitments in cutting total emissions, while China is the top among countries making commitments in intensity. Here we assume that the United States will still follow the Paris Agreement, and compare the resulting emissions or intensity with the one following the current trend, named as “None" scenario.Their resulting emissions are summarized in Figures 4.9 and 4.10.

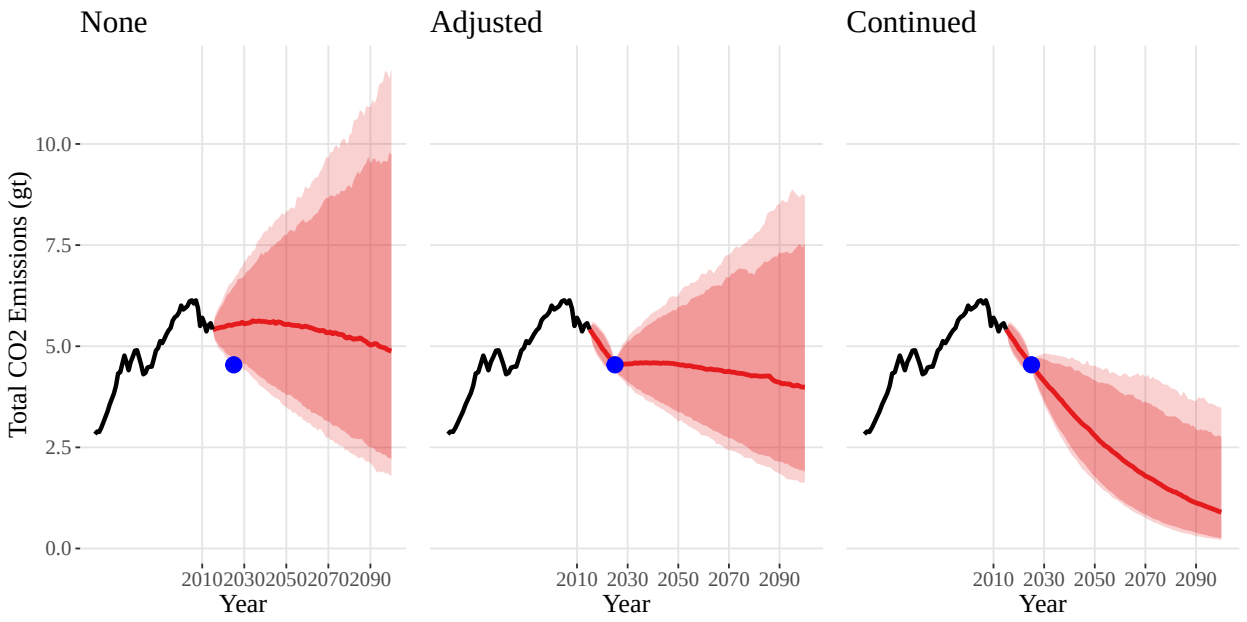


Figure 4.9: Emission forecast of the United States, under scenarios of current trend (left), *Adjusted* Scenario which assumes promises on Paris agreement is met and policies will not be continued (middle), and *Continued* Scenario whose policies will be continued (right). The blue dots are for the target, which is 26% less than the yearly emission in 2005.

In 2017, President Trump announced that the USA would withdraw from the Paris Agreement. We therefore consider a fourth scenario, under which the USA does not meet its NDC and continues emissions in line with current trends, while all other countries make additional efforts and meet their NDCs. The probabilistic forecasts of global mean temperature under

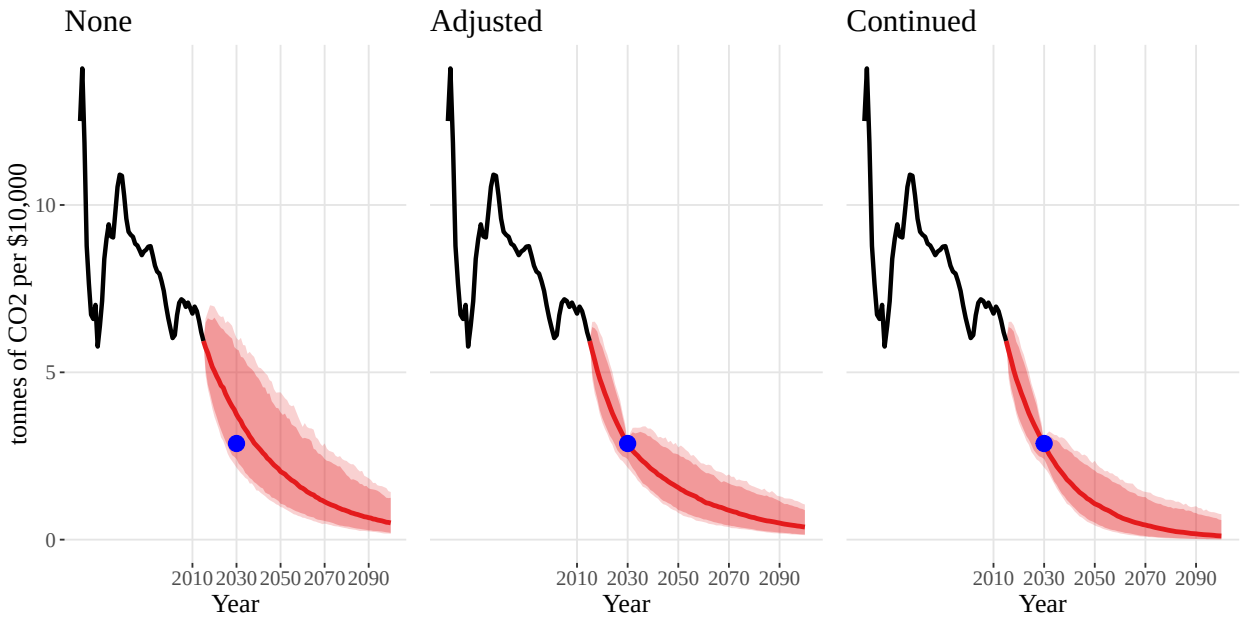


Figure 4.10: Intensity forecast of the China, under scenarios of current trend (left), *Adjusted* Scenario which assumes promises on Paris agreement is met and policies will not be continued (middle), and *Continued* Scenario whose policies will be continued (right). The blue dots are for the target, which is 60% less than the intensity in 2005.

all four scenarios are shown in Figure 4.11.

We find that on current trends, but without additional efforts to meet the NDCs, the median forecast of global mean temperature increase is 2.8°C with likely interval (90% prediction interval) [2.1, 3.9]°C. If all countries meet their NDCs, but revert to current trends thereafter, the median forecast declines by 0.2°C to 2.6°C, with likely interval [2.0, 3.4]°C. If all countries meet their NDCs and continue to reduce carbon emissions at the same rate thereafter, the median forecast declines by a further 0.3°C, to 2.3°C, with likely interval [1.8, 2.9]°C. The probability of staying below 2°C is 5% under the "None" scenario, 12% under the "Ajusted" scenario, and 26% under the "Continued" scenario.

If the USA continues on its current trend rather than meeting its NDC, the median forecast of cumulative global carbon emissions would be about 10% (220 Gt CO_2) higher than under the "Continued" scenario. The median temperature forecast would then rise to

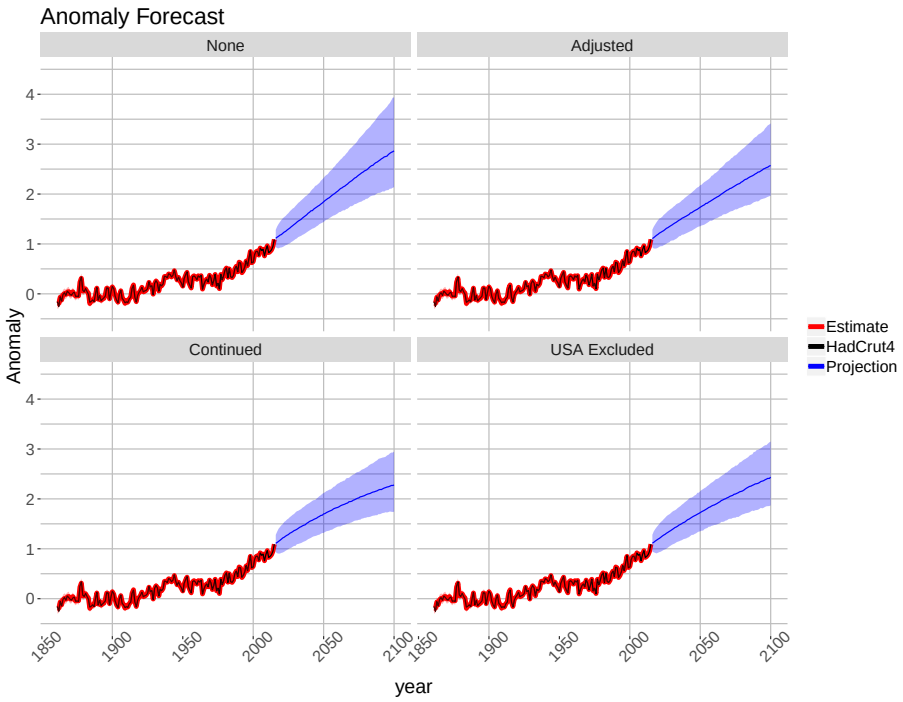


Figure 4.11: Global Mean Temperature Forecast under different scenarios. Here "None" represents no extra efforts, "Adjusted" stands for Paris Agreement Met and Policy stopped. The "Continued" scenario is for the forecast assuming Paris Agreement will be met and the policies are continued. Lastly, the "USA Excluded" scenario is assuming that all countries except for the United States would make their promises and would also continue the policy until the end of this century.

2.4°C with likely interval [1.9, 3.1]°C, and the probability of staying below 2°C would go down from 26% to 18%. Under all the scenarios, the probability of staying below 1.5°C is less than 2%.

4.5.5 Extra Effort to 2° C warming

Our results suggest that even if all countries meet their promises under the Paris Agreement and continue to reduce emissions at the same rate thereafter, it is unlikely that warming would stay under 2°C, a conclusion also reached by other authors using different approaches[Rogelj et al., 2016, Pan et al., 2017]. We therefore ask more precisely, what further

reductions would be needed to ensure this? Or, to put it another way, by how much would the NDCs need to be increased?

In order to answer this question, we define the problem as setting the objective to be achieve less than 2°C of global warming as measured in two ways, either by median or by 95% quantile. Since the global warming at 2100 depends on the cumulative emission of this century and the CMIP 5 model biases, we will control the extra reduction in cumulative emissions with bisection methods[Weisstein Eric, 2020] to optimize this.

Our median forecast of cumulative carbon emissions by 2100 is 3,108 Gt CO_2 without the Paris Agreement, and 2,083 Gt CO_2 under the "Continued" scenario. We find that to have a 50% chance of limiting warming to 2°C, cumulative emissions would need to be reduced further to 1,579 Gt CO_2. Assuming a constant rate of annual decline in emissions, this would require that the annual rate of decline would need to be increase by 1% to reach the NDCs, and 1.8% to have a 50% chance of staying under 2°C.

Similarly, to have a 95% chance of limiting warming to 2°C, cumulative emissions would need to be reduced further to 839 Gt CO_2. Moreover, to have a 50% or 95% chance of limiting the warming by 1.5°C, cumulative emissions would need to be reduced to 471 Gt and 129 Gt CO_2.

Then, we now describe a method for calculating the increases in the NDCs (and the intended NDCs) needed to meet the targets, given the required cumulative emissions to 2100.

In the absence of additional efforts, the median forecast is for total global annual emissions to remain roughly constant from 2016 to 2100, at around $A = 3108/85 = 36.6$ Gt CO_2 per year (see Figure 4.1). Suppose that to achieve a given climate target with a given probability would require keeping cumulative emissions in 2016–2100 to at most X Gt CO_2. Let a be the rate of decline in annual global emissions needed to achieve this, such that if E_t is global annual emissions in year $2015+t$, then $E_t = E_0 e^{-at}$. Then $100 \times a$ is essentially the percentage annual decline in global emissions when a is close to 0.

To find a, note that cumulative emissions from 2016 to year $(2015 + T)$ from a starting

point of 1 Gt CO_2 per year in 2016, is

$$C(a,T) = \int_0^T e^{-at}dt = (1 - e^{-aT})/a. \tag{4.9}$$

Then a is the solution of the nonlinear equation $C(a,T)A = X$, and this can be found using a numerical univariate root-finding method.

Let a_P be the value of a needed for meeting the Paris Agreement NDCs, assuming that for countries for which the NDCs refer to years before 2030 (such as the USA), the declines in annual emissions continue beyond the NDC target date to 2030 at the same average annual rate. In that case, $X = 2083$, and solving the equation $C(a,T)A = X$ yields $a_P = 0.0101$, or an annual rate of decline in emissions of just over 1%.

Each target and probability corresponds to a different value of X, and the corresponding value of a can be calculated in the same way as for the NDCs. For example, to have a 50% chance of staying below 2°C in 2100 requires $X = 1579$, which corresponds to $a = 0.0182$, or an annual rate of decline about 80% higher than needed to meet the Paris NDCs.

We can calculate the corresponding needed increase in NDCs as follows, taking Germany as an example. The NDC for Germany is to reduce carbon emissions by 40% from 1990 to 2030. Germany's carbon emissions in 1990 were 1052 Mt CO_2, and 795 Mt CO_2 in 2015. Thus the NDC for Germany corresponds to a target of $1052 \times 0.6 = 630$ Mt CO_2 in 2030. This requires an annual rate of decline of $1 - \left(\frac{630}{795}\right)^{\frac{1}{15}} = 0.0154$ from 2015 to 2030. To stay below 2°C in 2100 requires a rate of decline that is $0.0182/0.0101 = 1.802$ times higher, or 0.0278. This leads to a revised target for 2030 of $795 \times (1 - 0.0278)^{15} = 521$. This is a reduction of 50% over the 1990 level, which is 25% more than the NDC level of a 40% reduction. Thus we say that to stay below 2°C in 2100, Germany would need to increase its NDC by 25%.

The calculation is slightly different for countries whose NDCs are expressed in terms of carbon intensity rather than carbon emissions. We assume that GDP is measured in current values in local currency, and we use the numbers reported by the World Bank. We will

take China as an example. China's NDC is to reduce carbon intensity by 60% from 2005 to 2030. China's carbon emissions in Mt CO_2 were 5771 in 2005 and 9717 in 2015. Its GDP was 18.73T yuan in 2005 and 68.60T yuan in 2015. Thus its carbon intensity was 308.1 in 2005 and 141.7 in 2015. We then do the same calculation as we did for Germany, but for carbon intensity instead of carbon emissions. The result is that China's NDC should become a reduction of 64.2% instead of 60%, an increase of 7% in the promised reduction.

Based on the previous method, the needed increase in the NDCs would vary by country, depending on their promises and progress to date. For the six largest emitters, the needed increases in the NDCs would be 7% for China, 38% for the USA, 55% for India, 49% for Japan, and 25% for Germany in order to have an even chance of staying below 2°C.

Note that this would require global progress towards net zero emissions, but it would not require global annual emissions to reach net zero before 2100, although it would likely involve individual countries doing so. Under this scenario, emissions would need to decline by about 80% relative to their median forecast in the absence of additional efforts (which is roughly equal to the current level), giving global annual emissions of about 7.8 Gt CO_2 by 2100.

Similar calculations indicate that to make it likely (90% probability) to stay below 2°C of warming by 2100, rather than just an even chance, would require more than quadrupling the annual rate of decline in emissions. This would require reaching close to global net zero emissions (10% of the current level) by 2070. Many individual countries would need to reach net zero earlier to achieve this goal.

To have an even chance of staying below 1.5°C would require multiplying the annual rate of decline by about 8, reaching close to global net zero emissions by 2045. To make it likely to stay below 1.5°C would require multiplying the annual rate of decline by almost 30, and reaching close to global net zero by 2023. It is not too surprising that staying below 1.5°C would be so difficult, given that there is already estimated committed warming of 1.3°C [Davis et al., 2010, Mauritsen and Pincus, 2017, Brown et al., 2019, Tong et al., 2019].

4.6 Discussion

In this chapter, we have proposed a new method for probabilistic forecast of the climate change. This method combined the probabilistic CO_2 emissions forecast in Raftery et al. [2017], with the well-studied, scenario-based CMIP 5 forecast. With the proposed method, we generate probabilistic forecasts based on well-founded atmospheric science models as well as statistical models. Moreover, in this chapter, we analyzed the outcome of the Paris Agreement quantitatively.

Based on our analysis, we are not optimistic about keeping the global temperature rise below 2°C above pre-industrial levels. With our analysis, there will be about 5% chance that at the end of the century, the global temperature rise will below 2 degrees Celsius, with the median rise at 2.88 degrees. This means that without global response to the current situation, it will be almost impossible to achieve the target.

Unfortunately there are problems associated with the Paris Agreement, currently the only global response to the threat of climate change. The first problem is the feasibility of countries fulfilling their promises. With the current trend, major countries and regions, including the United States, China, Japan and the European Union, are all far from their promises. With their current trend, the probabilities that they can achieve their goal are all less than 10%. Moreover, even assuming all countries could do what they promised in their NDCs, we are still not enough to keep the global temperature rise below 2°C. With our analysis, even with the most rigorous scenario, we still need to reduce an extra 60% of the cumulative emissions such that we will have a 50% chance of keeping the warming below 2°C.

Note that these calculations refer to global mean temperature only; the same global mean temperature can yield different spatial temperature distributions [Seneviratne et al., 2018]. Also, we have assumed that changes in emissions will be at a roughly constant rate, while in fact of course, the rate of change could change substantially over time [Rogelj et al., 2019a,b]. However, the historical data on which our model is built do indicate that changes in carbon

intensity have tended to be relatively constant over time.

One possible limitation of our model is that we do not forecast negative emissions because the model is on the logarithmic scale. While this does not pose a problem in terms of fitting past data or most plausible future prospects, a possible way to achieve the most ambitious goals, such as staying below 1.5°C, would be to compensate for the use of fossil fuels by massive reforestation, possibly leading to negative emissions. It has been estimated that the Annex I countries (the industrialized countries, including OECD countries and EIT parties [OECD, 2020]) could claim a net carbon offset as high 0.2 GtC per year in this way [Yamagata and Alexandrov, 2001]. This amount is small relative to the forecast yearly emissions, and we could not find other evidence of a likely large effect of forestation, other than in a few regions, such as some areas of China [Huang et al., 2012]. Indeed, in the Amazon rainforest, deforestation is not even decelerating [Shukla et al., 1990, Binswanger, 1991, Boekhout van Solinge, 2014, Fearnside, 2017]. Thus there is little evidence suggesting that this would be a major force leading to negative emissions.

One issue is that for some countries the NDCs are vague, for example giving only a range of dates for meeting the NDCs, not saying whether the promised cuts are in absolute emissions or carbon intensity, or making the commitments conditional on vaguely specified levels of international support. In these cases we assumed that the NDCs were for the smallest cuts consistent with the documents. For example, if a range of dates was given for reaching a target, we took the date closest to 2030. If it was not specified, we assumed that the promised cut was in carbon intensity, not absolute emissions. If commitments were conditional on international support, we took the NDC to be the lowest level of cut promised unconditional on international support.

There were also data issues that we resolved in a similar manner. For example, it has been argued that China's historical emissions have been about 10% lower than assumed here due to different estimates of emissions factors [Liu et al., 2016], but we chose to use the estimate from the Global Carbon Budget [Le Quéré et al., 2018], to keep all forecasts comparable.

The present approach is probabilistic, in contrast with that of the [Intergovernmental

Panel on Climate Change, 2014], who used four main deterministic scenarios for emissions, RCPs. There are problems interpreting the RCPs, and from the early days of the IPCC, some members called for a statistical forecasting approach such as the one we have used here, in preference to scenarios [Moss and Schneider, 2000]. One reason for using deterministic scenarios may have been that official population forecasts, an important input to the emissions forecasts, were then available only as deterministic scenarios. In 2015, however, the UN made its official population projections for all countries probabilistic for the first time [United Nations, 2015b, Raftery et al., 2012], making it possible to produce probabilistic emissions forecasts soon thereafter [Raftery et al., 2017]. The resulting emission forecast range in [Raftery et al., 2017] is tighter than the range of the RCPs, but consistent with the two middle RCPs.

In general when we have had to make assumptions not clearly dictated by the data, we have tried to err, on the side of being conservative, in the sense of favoring lower rather than higher emissions and hence temperature. For example, while our forecasts of GDP have performed well in out-of-sample validation assessments [Raftery et al., 2017], some other probabilistic forecasts of global long-term economic growth are higher [Müller and Watson, 2016, Startz, 2020].

Chapter 5
DISCUSSION AND FUTURE WORK

This chapter summarizes the main contributions of the dissertation and the future extensions of the research. Additional discussions can be found in the Sections 2.5, 3.4.3 and 4.6.

5.1 Contributions

I have developed a method for probabilistic estimation of past total fertility rates, and have made projections of total fertility rates with uncertainty of past estimations. This method is considered as a remedy for the existing problem of the probabilistic forecast currently adopted by the United Nations [United Nations, 2015b, 2019b]. The validation results have shown that the existing method will have a lower coverage when there are corrections on TFR estimations in the official estimation across version, and the new method can fix this discrepencies by providing wider intervals in forecasting for countries with less accurate historical data. This new method works for major countries, and calls for improvements in the mechanisms and practice of data collection in those countries, which today have to rely on surveys in the absence of reliable vital registration.

Another main contribution of my work is extending the total fertility rate models into annual estimates. This is a major contribution to the United Nations' probabilistic population framework, as we not only take the auto-correlation pattern of TFR transition into consideration in the model, but also make the model generic and ready to use by updating them in the R package **bayesTFR**. We have removed the limit of covariates in use, allowed for different combinations of settings in the model, and have made the visualization, analysis easy to get. Furthermore, other functionalities, such as special treatment for countries with

long history of Vital Registration systems, are also provided. With the updated tool, the analysis on different features for different countries could be conducted in order to generate better annual TFR estimates.

We extend our work into the application of probabilistic forecast of population. Specifically, we contribute in carbon emissions and climate change. In Raftery et al. [2017], we have proposed and implemented the method for probabilistic forecast of future CO_2 emissions. This is a vital input in forecasting future climate change, as we discuss in this dissertation in Section 1.2.2 that the gap between actual forecast of climate change and the climate model was not bridged before. We have built another model for connecting the probabilistic forecast of CO_2 emissions to the global mean temperature. This work extends the scenario-based description of future climate change by generating probabilistic projections, both in point and in interval.

A key outcome of this work is the collection of the countries' intended contributions to fight against climate change in the Paris Agreement. This is the first time that we could understand the nature of the Paris Agreement and its effect quantitatively, including the probability of countries' achieving their targets, the outcome of the agreement and the extra efforts needed for global climate targets in IPCC [Intergovernmental Panel on Climate Change, 2014]. We combine this work with the global mean temperature forecast, in order to let the audience have a clear picture for how far we are from Sustainable Developments and help the decision makers in making plans quantitatively fighting for the climate change.

5.2 Future Work

5.2.1 TFR Estimation and Projection

In the modeling process of TFR estimation, we have discussed the potential for building a complete Bayesian hierarchical model for the raw TFR estimations from the World Fertility Data. There are clear incentives in doing so. The estimation process will be generic instead of having a different process in estimating the bias and standard deviation before the MCMC

process. Additionally, building a generic model will make changes in the distributional assumptions of the data easier. There are data points (such as the data point in Argentina in Figure 2.4) that clearly fit t distributions better than normal distributions. Though in the experiments and validation study, it turns out in this case that a generic normal assumptions could generate reasonable estimates as well as good intervals, these data points are suggesting different operations. From a theoretic statistical perspective, we could frame our model by setting different distributional assumptions on different sources of the data for different countries, but this is beyond the scope of the current approach and if we want to extend it, it must be conducted very carefully. Combining them will require a complete understanding of data generating process in the data base, as simply building a model without understanding the nature of demographic data will lead to unreasonable estimates.

In the persective of computational statistics, we have implemented the method in the package **bayesTFR**. In the developing process, we have figured out that in R, vectorization in computations would significantly improving the computing speed, which is one of the major concern for developing more sophiscated model. We have optimized the MCMC steps for TFR estimation, but for the country-specific parameters, we have kept the updating procedure sequencially. This is not the major concern for our purpose, as the time used in TFR updates is longer than the country-specific parameters. However, this is a direction for updating the **bayesTFR** itself, as for users who want to keep the original framework of estimation, vectorizing the updates of country-specific parameter could be beneficial for speeding up their analysis.

For demographic audiences, there would be more potentials for extending the work. Including noisy features in forecasting total fertility rates is typically not a good idea for accuracy in statistical forecast, but many factors could be considered as additional features in accelerating or decelerating the process of fertility transitions. For example, other members of our research group has led to the publication of research in relationship between education and fertility transition [Liu and Raftery, 2020a]. Similarly, the relationship between smoking [Li and Raftery, 2020], HIV [Godwin and Raftery, 2017] and the life expectancy are consid-

ered important complements for the life expectancy and mortality rate projections. There could be many other social-economic factors that could be taken into consideration, at least for past estimation purpose.

In the discussion section in Chapter 2, we have pointed out that China was excluded in the framework for solving the problem of the underreporting, which is a special data issues for one of the most important countries in demographic science. As we described and as far as I understand, scientists at the UN Population Division have been developing such a method and are currently preparing a paper to describe it. This could be important for demographic audiences and it is an extension of the current framework.

Fertility transition is one of the most important topics in demographic researches. With the estimation of past TFR, we could get the probabilistic statements of fertility transition for countries in the scope. This could lead to a summary report of countries' starting and ending years of fertility transitions, which could be important to certaint demographic researchers.

Finally, uncertainty in age-specific fertility rates is not modeled explicitly. We generate future age-specific fertility rates forecast by taking the predicted total fertility rates trajectories and convert them to age-specific ones. This method only takes the uncertainty of total fertility rates into account, but the uncertainty specifically to age-specific fertility rates are omitted. One possible extension for this purpose could be similar to Lee-Carter method, or first-order approximation of age-specific fertility rate.

5.2.2 Climate forecasting

From the statistical point of view, there are definitely other alternatives for ensembling models from CMIP model set. For example, Bayesian model averaging [Madigan et al., 1996, Hoeting et al., 1999, Raftery et al., 2005] could be a straightforward extension. Instead of having equal weights on different climate models ensembled in CMIP 5 model sets, we could consider including weighted average of them based on their forecasting performance with the Bayesian model averaging schema. The reason why we didn't do this is because we are not confident in having a thorough understanding of all these models, including their benefits

and drawbacks. Thus, evaluating the performance simply using statistical way without considering the underlying scientific nature could be misleading. Since we are focusing on long-term forecast, a short period in evaluating model performance may not be enough, and by far as I know, there are corrections in CMIP models for the past trend. If we don't exclude these effects in model selection, a simple statistical way for ensembling them could lead to bad forecast or could under-calibrate the forecasting intervals.

Currently, the analysis of the Paris Agreement is conducted based on National Determined Contributions of all countries. As we discussed in the paper, many countries have promised on areas that are not related to CO_2 emissions directly, and thus we excluded them in analysis. However, these part could contribute to global emissions, and the contribution should also be evaluated in the future. It is the fifth year after signing the Paris agreement, and we could have access to more data in evaluating the trajectories of each country to their promises.

We have built a complete probabilistic model for CO_2 emission forecast and the corresponding global mean temperature forecast. The modeling procedure could be considered as a framework in generating similar climate change forecasts, as long as the relationship of corresponding features in forecast and the input features is achievable. By making forecast on RCP-related components, we could bridge between the climate change forecast and the social-economic factors' forecast, which I believe could be a method to extend to many other aspects of research in climate change, such as sea level rises.

Finally, regional forecasts could also be achievable. The existing pattern scaling technique [Tebaldi and Arblaster, 2014] could be a possible tool for completing the task. Based on our preliminary analysis of the historical reanalysis data [Hersbach et al., 2020], 60 percent of total variation of regional temperature across the globe in the last 40 years could be summarized by the first two principle components. This implies that there is potentials in connecting the global mean temperature forecasts to regional temperature. For this purpose, Markov Random Fields could be possible technique in use, and it would raise great interest to climate scientists.

BIBLIOGRAPHY

Guy J. Abel, Bilal Barakat, K. C. Samir, and Wolfgang Lutz. Meeting the sustainable development goals leads to lower world population growth. *Proceedings of the National Academy of Sciences*, 113(50):14294–14299, 2016.

Maarten Alders, Nico Keilman, and Harri Cruijsen. Assumptions for long-term stochastic population forecasts in 18 European countries. *European Journal of Population/Revue Européenne de Démographie*, 23:33–69, 2007.

Juha Alho and Bruce Spencer. *Statistical Demography and Forecasting*. Springer Science & Business Media, 2006.

Juha Alho, Maarten Alders, Harri Cruijsen, Nico Keilman, Timo Nikander, and Dinh Quang Pham. New forecast: Population decline postponed in Europe. *Statistical Journal of the United Nations Economic Commission for Europe*, 23:1–10, 2006.

Juha Alho, Svend E. Hougaard Jensen, and Jukka Lassila. *Uncertain Demographics and Fiscal Sustainability*. Cambridge University Press, Cambridge, U.K., 2008.

Leontine Alkema, Adrian E. Raftery, Patrick Gerland, Samuel J. Clark, François Pelletier, Thomas Buettner, and Gerhard K. Heilig. Probabilistic projections of the total fertility rate for all countries. *Demography*, 48:815–839, 2011.

Leontine Alkema, Adrian E. Raftery, Patrick Gerland, Samuel J. Clark, and François Pelletier. Estimating trends in the total fertility rate with uncertainty using imperfect data: Examples from West Africa. *Demographic Research*, 26:331–362, 2012.

Guri Bang, Jon Hovi, and Tora Skodvin. The Paris Agreement: Short-term and long-term effectiveness. *Politics and Governance*, 4(3):209–218, 2016.

Hans P. Binswanger. Brazilian policies that encourage deforestation in the Amazon. *World Development*, 19(7):821–829, 1991.

Ralph Bodle, Lena Donat, and Matthias Duwe. The Paris Agreement: analysis, assessment and outlook. *CCLR*, page 5, 2016.

Tim Boekhout van Solinge. Researching illegal logging and deforestation. *Journal of Crime, Criminal Law and Criminal Justice*, 3(2):35–48, 2014.

Christina Bohk-Ewald, Peng Li, and Mikko Myrskylä. Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences*, 115(37):9187–9192, 2018.

Jutta Bolt, Inklaar Robert, Herman de Jong, and Jan Luiten van Zanden. Maddison project database, version 2018, "rebasing 'maddison': new income comparisons and the shape of long-run economic development". 2018. URL `http://www.ggdc.net/maddison`.

John Bongaarts and John Casterline. Fertility transition: is sub-Saharan Africa different? *Population and Development Review*, 38(Suppl 1):153, 2013.

H Booth, S Pennec, and R. J. Hyndman. Stochastic population forecasting using functional data methods: The case of France. In *Annual Meeting of the International Union for the Scientific Study of Population, Marrakech, Morocco*, 2009.

William Brass. *Uses of Census or Survey Data for the Estimation of Vital Rates*. United Nations, New York, 1964.

William Brass. *Demography of Tropical Africa*. Princeton University Press, Princeton, N.J., 2015.

Stephen P. Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.

Calum Brown, Peter Alexander, Almut Arneth, Ian Holman, and Mark Rounsevell. Achievement of Paris climate goals unlikely due to time lags in the land system. *Nature Climate Change*, 9(3):203–208, 2019.

Yong Cai. An assessment of China's fertility level using the variable-r method. *Demography*, 45:271–281, 2008.

Edwin Cannan. The probability of cessation of growth of population in England and Wales during the next century. *Economic Journal*, 5:506–515, 1895.

P. C. Roger Cheng and Eric S. Lin. Completing incomplete cohort fertility schedules. *Demographic Research*, 23:223–256, 2010.

Ansley J. Coale and Donald R. McNeil. The distribution by age of the frequency of first marriage in a female cohort. *Journal of the American Statistical Association*, 67(340): 743–749, 1972.

Steven J. Davis, Ken Caldeira, and H Damon. Matthews. Future CO_2 emissions and climate change from existing energy infrastructure. *Science*, 329:1330–1333, 2010.

Cong Dong, Xiucheng Dong, Qingzhe Jiang, Kangyin Dong, and Guixian Liu. What is the probability of achieving the carbon dioxide emission targets of the Paris Agreement? evidence from the top ten emitters. *Science of the Total Environment*, 622:1294–1303, 2018.

Dalkhat M. Ediev. Comparative importance of the fertility model, the total fertility, the mean age and the standard deviation of age at childbearing in population projections. Paper presented to the International Population Conference, Busan, Korea. https://www.iussp.org/en/event/17/programme/paper/3054, 2013.

M. D. R. Evans. American fertility patterns: A comparison of white and nonwhite cohorts born 1903-56. *Population and Development Review*, pages 267–293, 1986.

Philip Fearnside. Business as usual: a resurgence of deforestation in the Brazilian Amazon. *Yale Environ*, 360:1–6, 2017.

Federal Republic of Nigeria National Population Commission and ICF Macro. Nigeria demographic and health survey 2008. *National Population Commission and ICF Macro*, 2009.

Bailey K. Fosdick and Adrian E. Raftery. Regional probabilistic fertility forecasting by modeling between-country correlations. *Demographic Research*, 30:1011–1034, 2014.

Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472, 1992.

Patrick Gerland, Adrian E. Raftery, Hana Ševčíková, Nan Li, Danan Gu, Thomas Spoorenberg, Leontine Alkema, Bailey K. Fosdick, Jennifer L. Chunn, Nevena Lalic, Guiomar Bay, Thomas Buettner, Gerhard K. Heilig, and John Wilmoth. World population stabilization unlikely this century. *Science*, 346:234–237, 2014.

Patrick Gerland, Ann Biddlecom, and Vladimíra Kantorová. Patterns of fertility decline and the impact of alternative scenarios of future fertility change in sub-Saharan Africa. *Population and Development Review*, 43:21–38, 2017.

Jessica Godwin and Adrian E. Raftery. Bayesian projection of life expectancy accounting for the HIV/AIDS epidemic. *Demographic research*, 37:1549, 2017.

Daniel M. Goodkind. China's missing children: the 2000 census underreporting surprise. *Population Studies*, 58:281–295, 2004.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 2020.

Jennifer A. Hoeting, David Madigan, Adrian E. Raftery, and Chris T. Volinsky. Bayesian model averaging: a tutorial. *Statistical Science*, pages 382–401, 1999.

Lin Huang, Jiyuan Liu, Quanqin Shao, and Xinliang Xu. Carbon sequestration by forestation across china: Past, present, and future. *Renewable and Sustainable Energy Reviews*, 16 (2):1291–1299, 2012.

J Hurrell, M Visbeck, and A Pirani. WCRP Coupled Model Intercomparison Project–Phase 5, special issue of the CLIVAR exchanges newsletter, 2011.

Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. WMO/UNEP, 2014.

Jennifer Jacquet and Dale Jamieson. Soft but significant power in the Paris Agreement. *Nature Climate Change*, 6(7):643–646, 2016.

Yoichi Kaya, Keiichi Yokobori, et al. *Environment, energy, and economy: strategies for sustainability*. United Nations University Press Tokyo, 1997.

Endale Kebede, Anne Goujon, and Wolfgang Lutz. Stalls in Africa's fertility decline partly result from disruptions in female education. *Proceedings of the National Academy of Sciences*, 116(8):2891–2896, 2019.

Nathan. Keyfitz. The limits of population forecasting. *Population and Development Review*, 7:579–593, 1981.

Corinne Le Quéré, Robbie M. Andrew, Pierre Friedlingstein, Stephen Sitch, Judith Hauck, Julia Pongratz, Penelope A. Pickers, Jan Ivar Korsbakken, Glen P. Peters, Josep G.

Canadell, et al. Global carbon budget 2018. *Earth System Science Data*, 10(4):2141–2194, 2018.

Ronald D. Lee. Modeling and forecasting the time series of us fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting*, 9:187–202, 1993.

Ronald D. Lee and Shripad Tuljapurkar. Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association*, 89:1175–1189, 1994.

Yicheng Li and Adrian E. Raftery. Estimating and forecasting the smoking-attributable mortality fraction for both genders jointly in over 60 countries. *The Annals of Applied Statistics*, 2020.

Genovaitė Liobikienė and Mindaugas Butkus. The European Union possibilities to achieve targets of Europe 2020 and Paris Agreement climate policy. *Renewable Energy*, 106:298–309, 2017.

Daphne H. Liu and Adrian E. Raftery. How do education and family planning accelerate fertility decline? *Population and Development Review*, 2020a.

Peiran Liu and Adrian E. Raftery. Accounting for uncertainty about past values in probabilistic projections of the total fertility rate for most countries. *Annals of Applied Statistics*, 14(2):685–705, 2020b.

Zhu Liu, Steven J Davis, Kuishuang Feng, Klaus Hubacek, Sai Liang, Laura Diaz Anadon, Bin Chen, Jingru Liu, Jinyue Yan, and Dabo Guan. Targeted opportunities to address the climate–trade dilemma in China. *Nature Climate Change*, 6(2):201–206, 2016.

Robert E. Lucas. Some macroeconomics for the 21st century. *Journal of Economic Perspectives*, 14(1):159–168, 2000.

Wolfgang Lutz and K. C. Samir. Dimensions of global population projections: What do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B*, 365:2779–2791, 2010.

David Madigan, Adrian E. Raftery, Chris T. Volinsky, and Jennifer Hoeting. Bayesian model averaging. In *Proceedings of the AAAI Workshop on Integrating Multiple Learned Models, Portland, OR*, pages 77–83, 1996.

Thorsten Mauritsen and Robert Pincus. Committed warming inferred from observations. *Nature Climate Change*, 2017.

M. Giovanna Merli and Adrian E. Raftery. Are births underreported in rural China? manipulation of statistical records in response to China's population policies. *Demography*, 37:109–126, 2000.

Colin P. Morice, John J. Kennedy, Nick A. Rayner, and Phil D. Jones. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, 117 (D8), 2012.

Richard H. Moss and Stephen H. Schneider. Towards consistent assessment and reporting of uncertainties in the IPCC TAR: Initial recommendations for discussion by authors. In R. Pachauri and T. Taniguchi, editors, *Cross-Cutting Issues in the IPCC Third Assessment Report*. Cambridge University Press, Cambridge, U.K., 2000.

Ulrich K. Müller and Mark W. Watson. Measuring uncertainty about long-run predictions. *Review of Economic Studies*, 83:1711–1740, 2016.

Christopher J. L. Murray, Charlton S. K. H. Callender, Xie Rachel Kulikoff, Vinay Srinivasan, Degu Abate, Kalkidan Hassen Abate, Solomon M. Abay, Nooshin Abbasi, Hedayat Abbastabar, Jemal Abdela, et al. Population and fertility by age and sex for 195 countries

and territories, 1950–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1995–2051, 2018.

Mikko Myrskylä, Joshua R. Goldstein, and Yen-hsin A. Cheng. New cohort fertility forecasts for the developed world: Rises, falls, and reversals. *Population and Development Review*, 39(1):31–56, 2013.

National Research Council and Committee on Population and others. *Beyond Six Billion: Forecasting the World's Population*. National Academies Press, 2000.

Radford M. Neal. Slice sampling. *Annals of Statistics*, 31:705–741, 2003.

OECD. List of annex I countries, 2020. URL `https://www.oecd.org/env/cc/listofannexicountries.htm`.

Karen O'Brien. Is the 1.5 C target possible? exploring the three spheres of transformation. *Current Opinion in Environmental Sustainability*, 31:153–160, 2018.

Xunzhang Pan, Michel den Elzen, Niklas Höhne, Fei Teng, and Lining Wang. Exploring fair and ambitious mitigation contributions under the Paris Agreement goals. *Environmental Science and Policy*, 74:49–56, 2017.

Paraskevi Peristera and Anastasia Kostaki. Modeling fertility in modern populations. *Demographic Research*, 16:141–194, 2007.

Dudley L. Poston Jr and Baochang Gu. Socioeconomic development, family planning, and fertility in China. *Demography*, pages 531–551, 1987.

Samuel H. Preston, Patrick Heuveline, and Michel Guillot. *Demography: Measuring and Modeling Population Processes*. Blackwell, Malden, Mass., 2000.

Thomas W. Pullum, Bruno Schoumaker, Stan Becker, and Sarah E. K. Bradley. An assessment of DHS estimates of fertility and under-five mortality. In *International Population*

Conference of the International Union for the Scientific Study of Population (IUSSP), Session 132: Data Quality in Demographic Surveys, August, volume 28, 2013.

Adrian E. Raftery and Steven M. Lewis. Implementing MCMC. In *Practical Markov Chain Monte Carlo*, pages 115–130. Chapman and Hall, London, 1996.

Adrian E. Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133 (5):1155–1174, 2005.

Adrian E. Raftery, Nan Li, Hana Ševčíková, Patrick Gerland, and Gerhard K. Heilg. Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109:13915–13921, 2012.

Adrian E. Raftery, Jennifer L. Chunn, Patrick Gerland, and Hana Ševčíková. Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50:777–801, 2013.

Adrian E. Raftery, Leontine Alkema, and Patrick Gerland. Bayesian population projections for the United Nations. *Statistical Science*, 29:58–68, 2014a.

Adrian E. Raftery, Nevana Lalic, and Patrick Gerland. Joint probabilistic projection of female and male life expectancy. *Demographic Research*, 30:795–822, 2014b.

Adrian E. Raftery, Alec Zimmer, Dargan M. W. Frierson, Richard Startz, and Peiran Liu. Less than 2°C warming by 2100 unlikely. *Nature Climate Change*, 7(9):637, 2017.

Robert D. Retherford, Minja Kim Choe, Jiajian Chen, Li Xiru, and Cui Hongyan. How far has fertility in China really declined? *Population and Development Review*, 31:57–84, 2005.

Joeri Rogelj, Michel Den Elzen, Niklas Höhne, Taryn Fransen, Hanna Fekete, Harald Winkler, Roberto Schaeffer, Fu Sha, Keywan Riahi, and Malte Meinshausen. Paris Agreement

climate proposals need a boost to keep warming well below 2 C. *Nature*, 534(7609):631–639, 2016.

Joeri Rogelj, Piers M Forster, Elmar Kriegler, Christopher J Smith, and Roland Séférian. Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature*, 571:335–342, 2019a.

Joeri Rogelj, Daniel Huppmann, Volker Krey, Keywan Riahi, Leon Clarke, Matthew Gidden, Zebedee Nicholls, and Malte Meinshausen. A new scenario logic for the Paris Agreement long-term temperature goal. *Nature*, 573:357–363, 2019b.

Carl Schmertmann, Emilio Zagheni, Joshua R. Goldstein, and Mikko Myrskylä. Bayesian forecasting of cohort fertility. *Journal of the American Statistical Association*, 109:500–513, 2014.

Bruno Schoumaker. Reconstructing fertility trends in sub-Saharan Africa by combining multiple surveys affected by data quality problems. In *Proceedings of the 2010 Annual Meeting of the Population Association of America*, 2010.

Bruno Schoumaker. Omissions of births in DHS birth histories in sub-Saharan Africa: Measurement and determinants. In *Proceedings of the 2011 Annual Meeting of the Population Association of America*, volume 31, 2011.

Bruno Schoumaker. *Quality and Consistency of DHS Fertility Estimates, 1990 to 2012*. ICF International Rockville, 2014.

Sonia I. Seneviratne, Joeri Rogelj, Roland Séférian, Richard Wartenburger, Myles R. Allen, Michelle Cain, Richard J. Millar, Kristie L. Ebi, Neville Ellis, Ove Hoegh-Guldberg, et al. The many possible climates from the Paris Agreement's aim of 1.5°C warming. *Nature*, 558:41–49, 2018.

Hana Ševčíková and Adrian E. Raftery. bayesPop: probabilistic population projections. *Journal of statistical software*, 75, 2016.

Hana Ševčíková, Leontine Alkema, and Adrian E. Raftery. bayesTFR: An R package for probabilistic projections of the total fertility rate. *Journal of Statistical Software*, 43(1):1, 2011.

Hana Ševčíková, Nan Li, Vladimíra Kantorová, Patrick Gerland, and Adrian E. Raftery. Age-specific mortality and fertility rates for probabilistic population projections. In *Dynamic demographic analysis*, pages 285–310. Springer, 2016.

Hana Ševčíková, Adrian E. Raftery, and Patrick Gerland. Probabilistic projection of subnational total fertility rates. *Demographic research*, 38:1843, 2018.

David Shapiro and Tesfayi Gebreselassie. Fertility transition in sub-Saharan Africa: falling and stalling. *African Population Studies*, 23(1), 2008.

David Shapiro and Andrew Hinde. On the pace of fertility decline in sub-Saharan Africa. *Demographic Research*, 37:1327–1338, 2017.

Jagadish Shukla, Carlos Nobre, and Piers Sellers. Amazon deforestation and climate change. *Science*, 247(4948):1322–1325, 1990.

Clive L. Spash. This changes nothing: The Paris Agreement to ignore reality. *Globalizations*, 13(6):928–933, 2016.

Richard Startz. The next hundred years of growth and convergence. *Journal of Applied Econometrics*, 35:99–113, 2020.

Michael A. Stoto. The accuracy of population projections. *Journal of the American Statistical Association*, 78:13–20, 1983.

Karl E. Taylor, Ronald J. Stouffer, and Gerald A. Meehl. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498, 2012.

Claudia Tebaldi and Julie M. Arblaster. Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Climatic Change*, 122(3):459–471, 2014.

Dan Tong, Qiang Zhang, Yixuan Zheng, Ken Caldeira, Christine Shearer, Chaopeng Hong, Yue Qin, and Steven J. Davis. Committed emissions from existing energy infrastructure jeopardize 1.5°C climate target. *Nature*, 572:373–377, 2019.

UNFCCC. Adoption of the Paris Agreement. Report NO. FCCC/CP/2015/L.9/Rev.1, 2015. URL `http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf`.

UNICEF. UNICEF annual report 2015. *United Nations*, 2016.

United Nations. *World Population Prospects: The 2008 Revision*. United Nations, New York, 2008.

United Nations. *World Population Prospects: The 2012 Revision*. United Nations, New York, 2012.

United Nations. World fertility data. `http://www.un.org/en/development/desa/population/publications/dataset/fertility/wfd2015.shtml`, 2015a.

United Nations. *World Population Prospects: The 2015 Revision*. United Nations, New York, 2015b.

United Nations. *World Population Prospects: The 2017 Revision*. United Nations, New York, 2017.

United Nations. World fertility data 2019. `https://www.un.org/en/development/desa/population/publications/dataset/fertility/wfd2019.asp`, 2019a.

United Nations. *World Population Prospects: The 2019 Revision*. United Nations, New York, 2019b.

United Nations. The sustainable development goals report 2019. `https://unstats.un.org/sdgs/report/2019`, 2019c.

United Nations Climate Change. National Determined Contributions. https://www4.unfccc.int/sites/NDCStaging/Pages/Home.aspx, 2018. [Online; accessed 19-Sept-2019].

Population Division U.S. Census Bureau. Annual estimates of the resident population: April 1, 2010 to July 1, 2017. *U.S. Census Bureau*, 2017.

Detlef P. Van Vuuren, Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson, Kathy Hibbard, George C. Hurtt, Tom Kram, Volker Krey, Jean-Francois Lamarque, et al. The representative concentration pathways: an overview. *Climatic change*, 109(1-2): 5, 2011.

Hana Ševčíková, Nan Li, Vladimira Kantorová, Patrick Gerland, and Adrian E. Raftery. Age-specific mortality and fertility rates for probabilistic population projections. In R. Schoen, editor, *Dynamic Demographic Analysis*, chapter 15, pages 285–310. Springer, 2016.

W. Weisstein Eric. Bisection, 2020. URL https://mathworld.wolfram.com/Bisection.html. From MathWorld–A Wolfram Web Resource.

Mark C. Wheldon. Personal communication, 2019.

Pascal K. Whelpton. Population of the United States, 1925–1975. *American Journal of Sociology*, 31:253–270, 1928.

Pascal K. Whelpton. An empirical method for calculating future population. *Journal of the American Statistical Association*, 31:457–473, 1936.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL https://ggplot2.tidyverse.org.

Yu Xie. Demography: Past, present, and future. *Journal of the American Statistical Association*, 95(450):670–673, 2000.

Yoshiki Yamagata and Georgii A Alexandrov. Would forestation alleviate the burden of emission reduction? an assessment of the future carbon sink from ard activities. *Climate Policy*, 1(1):27–40, 2001.

Zeng Yi. Is fertility in China in 1991–92 far below replacement level? *Population Studies*, 50:27–34, 1996.

Oran R. Young. The Paris Agreement: destined to succeed or doomed to fail? *Politics and Governance*, 4(3):124–132, 2016.

Zhenwu Zhai and Wei Chen. Research of China's total fertility rates in late 1990s (in Chinese). *Demographic Research (in Chinese)*, 31(1):19–31, 2007.

Yongxiang Zhang, Qingchen Chao, Qiuhong Zheng, and Lei Huang. The withdrawal of the US from the Paris Agreement and its impact on global climate change governance. *Advances in Climate Change Research*, 8(4):213–219, 2017.

Appendix A

APPENDICES TO CHAPTER 2

A.1 Model Specification

Here we provide a full description of the Bayesian hierarchical model, which was summarized in the main text:

$$
\begin{aligned}
\text{Level 1: } & y_{c,t,s}|f_{c,t} \sim \mathcal{N}(f_{c,t}+\delta_{c,s}, \rho^2_{c,s})\,, \\
& \mathbb{E}[\delta_{c,s}] = \boldsymbol{x}_{c,s}\boldsymbol{\beta}\,, \\
& \mathbb{E}[\rho_{c,s}] = \boldsymbol{x}_{c,s}\boldsymbol{\gamma}\,, \\
& f_{c,t} = \frac{1}{5}[(t_{l+t}-t)f_{c,t_l} + (t-t_l)f_{c,t_{l+5}}] \text{ for } t \in [t_l, t_{l+5}]\,; \\
\text{Level 2: } & \text{Phase I: } f_{c,t} = f_{c,t-5} + \varepsilon_{c,t}\,, \\
& \text{Phase II: } f_{c,t} = f_{c,t-5} - g(f_{c,t-5}|\theta_c) + \varepsilon_{c,t}\,, \\
& \text{Phase III: } f_{c,t} = \mu_c + \rho_c(f_{c,t-5}-\mu_c) + \varepsilon_{c,t}\,, \\
& \boldsymbol{\theta}_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c) \\
& \varepsilon_{c,t} \sim \mathcal{N}(0, \sigma^2_{c,t})\,, \\
& g(f_{c,t}|\boldsymbol{\theta}_c) = -\frac{d_c}{1+\exp\left(-\frac{2\ln(9)}{\Delta_{c1}}\left(f_{c,t}-\sum_i \Delta_{ci}+0.5\Delta_{c1}\right)\right)} \\
& \qquad\qquad + \frac{d_c}{1+\exp\left(-\frac{2\ln(9)}{\Delta_{c3}}\left(f_{c,t}-\Delta_{c4}-0.5\Delta_{c3}\right)\right)}
\end{aligned}
$$

The country-specific variance, $\sigma_{c,t}$, varies according to the phase and the current fertility

level, as follows:

$$\sigma_{c,t} = \begin{cases} s_\tau & \text{If } t \text{ is in Phase I.} \\ c_{1975}(t)\left(\sigma_0 + (f_{c,t} - S)(-aI_{f_{c,t}>S} + bI_{f_{c,t}<S})\right) & \text{If } t \text{ is in Phase II.} \\ \sigma_\epsilon & \text{If } t \text{ is in Phase III.} \end{cases}$$

$$c_{1975}(t) = \begin{cases} c & t \leq 1975 \\ 1 & t > 1975 \end{cases}$$

The country-level parameters, $\{U_c, \rho_c, \mu_c, \gamma_{ci}, \Delta_{c4}, d_c\}$, are specified as follows:

$$\text{Level 3: } U_c \begin{cases} = f_{c,\tau} & \tau_c \leq 1950 \\ \sim U(\min\{5.5, \max_t\{f_{c,t}\}\}, 8.8) & \tau_c < 1950 \end{cases}$$

$$\phi_c = \log\left(\frac{d_c - 0.25}{2.5 - d_c}\right),$$

$$\phi_c \sim \mathcal{N}(\chi, \psi^2),$$

$$\Delta'_{c4} = \log\left(\frac{\Delta_{c4} - 1}{2.5 - \Delta_{c4}}\right),$$

$$\Delta'_{c4} \sim \mathcal{N}(\Delta_4, \delta_4^2),$$

$$p_{ci} = \frac{\Delta_{ci}}{U_c - \Delta_{c4}} \text{ for } i = 1, 2, 3,$$

$$p_{ci} = \frac{\exp(\gamma_{ci})}{\sum_j \exp(\gamma_{cj})},$$

$$\gamma_{ci} \sim \mathcal{N}(\alpha_i, \delta_i^2),$$

$$\mu_c \sim \mathcal{N}(\bar{\mu}, \sigma_\mu^2),$$

$$\rho_c \sim \mathcal{N}(\bar{\rho}, \sigma_\rho^2);$$

where τ_c is the starting year of phase II for country c.

The hyperparameters are $\{s_\tau, \sigma_0, a, b, S, c, \sigma_\epsilon, \chi, \psi, \Delta_4, \delta_4, \boldsymbol{\alpha}, \boldsymbol{\delta}, \bar{\mu}, \sigma_\mu, \bar{\rho}, \sigma_\rho\}$. Some of these refer to Level 2 and some to Level 3. The prior distribution of these hyperparameters is as

follows:

$$
\begin{aligned}
\text{Level 4: } 1/s_\tau^2 &\sim \text{Gamma}(1, 0.4^2)\,, \\
\sigma_0 &\sim U[0.01, 0.6]\,, \\
a &\sim U[0, 0.2]\,, \\
b &\sim U[0, 0.2]\,, \\
S &\sim U[3.5, 6.5]\,, \\
c &\sim U[0.8, 2]\,, \\
\sigma_\epsilon &\sim U[0, 0.5]\,, \\
\chi &\sim \mathcal{N}(-1.5, 0.6^2)\,, \\
1/\psi^2 &\sim \text{Gamma}(1, 0.6^2)\,, \\
\Delta_4 &\sim \mathcal{N}(0.3, 1)\,, \\
1/\delta_i^2 &\sim \text{Gamma}(1, 1) \text{ for } i = 1, 2, 3, 4\,, \\
\alpha_1 &\sim \mathcal{N}(-1, 1)\,, \\
\alpha_2 &\sim \mathcal{N}(0.5, 1)\,, \\
\alpha_3 &\sim \mathcal{N}(1.5, 1)\,, \\
\bar{\mu} &\sim U[0, 2.1]\,, \\
\sigma_\mu &\sim U[0, 0.318]\,, \\
\bar{\rho} &\sim U[0, 1]\,, \\
\sigma_\rho &\sim U[0, 0.289]
\end{aligned}
$$

A.2 MCMC Diagnostics

The model was estimated using Markov chain Monte Carlo, as described in the main text. We used 3 chains of 12000 iterations each, thinned by 10 with burn-in of 2000. We diagnosed convergence by visual inspection of trace plots, and we also used more formal convergence diagnostics.

The convergence statistics of Gelman and Rubin [1992] and Brooks and Gelman [1998] for the total fertility rates of Nigeria are shown in Table A.1 as an example. Conventionally, these statistics indicate adequate convergence if they are close to 1, with 1.1 conventionally viewed as a threshold for acceptability. This is the case here, and indeed they are very close to 1, indicating satisfactory convergence by this criterion. This was true for all the model parameters and all the unobserved TFR values of all countries except for a very few number of countries including Djibouti, Mayotte, Saudi Arabia and South Sudan.

Table A.1: Gelman-Rubin and Brooks-Gelman Diagnostics for country-specific parameters for Nigeria. The point estimate of the multivariate potential scale reduction factor is 1.01926.

	Point estimates	Upper Confidence Interval
$f_{Nigeria,1953}$	1.008	1.020
$f_{Nigeria,1958}$	1.008	1.026
$f_{Nigeria,1963}$	1.008	1.027
$f_{Nigeria,1968}$	1.011	1.041
$f_{Nigeria,1973}$	1.003	1.008
$f_{Nigeria,1978}$	1.007	1.023
$f_{Nigeria,1983}$	1.009	1.032
$f_{Nigeria,1988}$	1.005	1.018
$f_{Nigeria,1993}$	1.007	1.024
$f_{Nigeria,1998}$	1.003	1.012
$f_{Nigeria,2003}$	1.000	1.000
$f_{Nigeria,2008}$	1.002	1.006
$f_{Nigeria,2013}$	1.002	1.006

The diagnostic of Raftery and Lewis [1996] is designed to indicate whether the MCMC chain has been run for enough iterations to estimate posterior quantiles of interest with

enough precision. We computed this diagnostic for all parameters and all unobserved TFR values for the 0.025 and 0.975 quantiles, requiring that the estimated quantile fall within 0.0125 of the true quantile in each case with probability at least 0.95. If the chain is completely independent, this yields a required number of iterations equal to 600. The largest number of iterations required was less than 5,000 for all the roughly 3,600 quantities being estimated, again indicating that the chain had been run long enough. The results for Nigerian TFR are again shown in Table A.2 as an example, for the 0.025 quantile. In all cases, the number of iterations needed was well below the 10,000 that were actually used, indicating that the number of iterations used was adequate.

Table A.2: Raftery-Lewis Diagostics for TFR Values of Nigeria

	Burn-in	Total Iteration	Dependence Factor
$f_{Nigeria,1953}$	21	3566	5.94
$f_{Nigeria,1958}$	23	3952	6.59
$f_{Nigeria,1963}$	22	3751	6.25
$f_{Nigeria,1968}$	13	2321	3.87
$f_{Nigeria,1973}$	14	2491	4.15
$f_{Nigeria,1978}$	20	3423	5.70
$f_{Nigeria,1983}$	18	3213	5.36
$f_{Nigeria,1988}$	16	2693	4.49
$f_{Nigeria,1993}$	14	2396	3.99
$f_{Nigeria,1998}$	13	2321	3.87
$f_{Nigeria,2003}$	18	3052	5.09
$f_{Nigeria,2008}$	15	2662	4.44
$f_{Nigeria,2013}$	18	3178	5.30

A.3 Out of sample validation Details

We are also interested in comparing the current model directly with the model proposed in [Alkema et al., 2011]. In the out of sample validation settings of [Alkema et al., 2011], we should use WPP 2008 all the time, train the model with data before 1995, and make forecast for next three periods (1995 - 2000, 2000 - 2005, 2005 - 2010). All other settings are the same in Section 3 of the main paper. We summarize the out of sample validation performance with similar criterion.

Table A.3: proportion of left-out UN estimates that fall above or below their 80% and 95% projection intervals splitted by periods.

	95% PI			80% PI		
	Below	Above	Coverage	Below	Above	Coverage
1995-2000	3.0%	1.5%	94.5%	13.9%	4.5%	81.6%
2000-2005	3.5%	2.5%	94.0%	12.9%	4.0%	83.1%
2005-2010	1.0%	3.0%	96.0%	7.5%	7.5%	85.0%

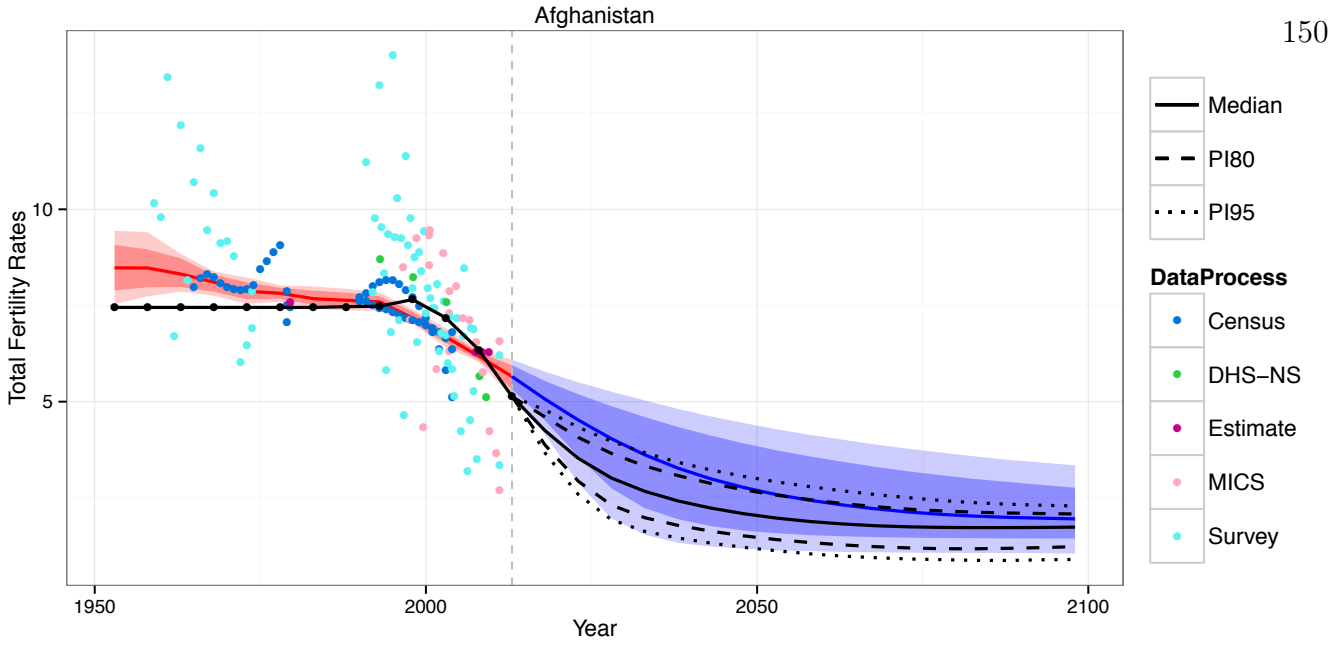
The overall coverage rates are 95.3% and 83.2% for 95% and 80% predictive intervals

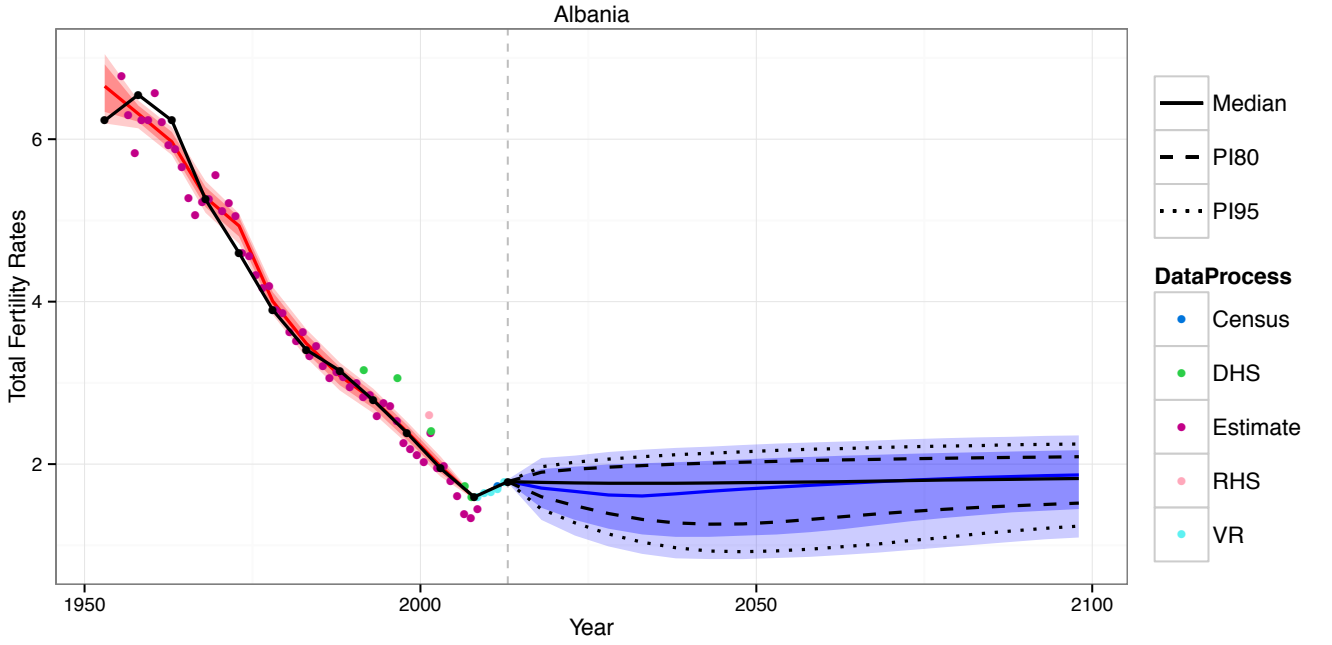
respectively. The coverage rate is very stable, and the 95% predictive intervals have achieved very precise coverage, which is similar to those out of sample validation in the main paper. The 80% out of sample interval is slightly wider than the nominal value. Since we just have one realization of all 201 countries and regions for out of sample validation, the coverage rate is satisfactory. We tried to compare this result in [Alkema et al., 2011], which is 91.5% and 77% for 95% and 80% predictive intervals respectively. Moreover, since in [Alkema et al., 2011], the phase III was trained without using Bayesian models, and the long term mean of phase III for all countries was set to be 2.1, which has been changed in the current model in BayesTFR, we tried to reproduce the same out of sample validation study with hierarchical model and different long term mean. We could figure out that the coverage rate is 92% and 77.5% for 95% and 80% predictive intervals respectively.

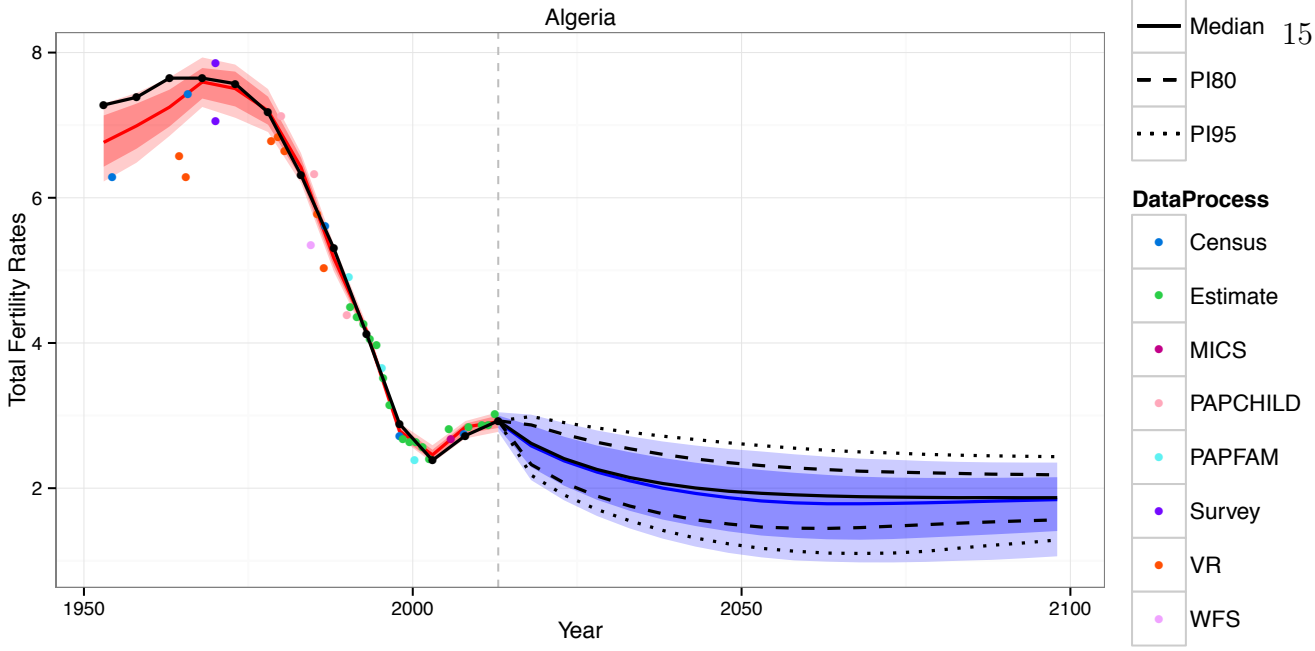
Thus, we could figure out that after we change the model into the setting of those in paper [Alkema et al., 2011], we could find that the overall coverage rate is slightly closer to the nominal rate. Thus, we could argue that our assumption that with accounting for uncertainty of past TFR estimation, the uncertainty in predictive interval is wider as we expected. If we are considering a single version of wpp package, our model could generate predictive intervals that are closer to the correct value, but the improvement is small.

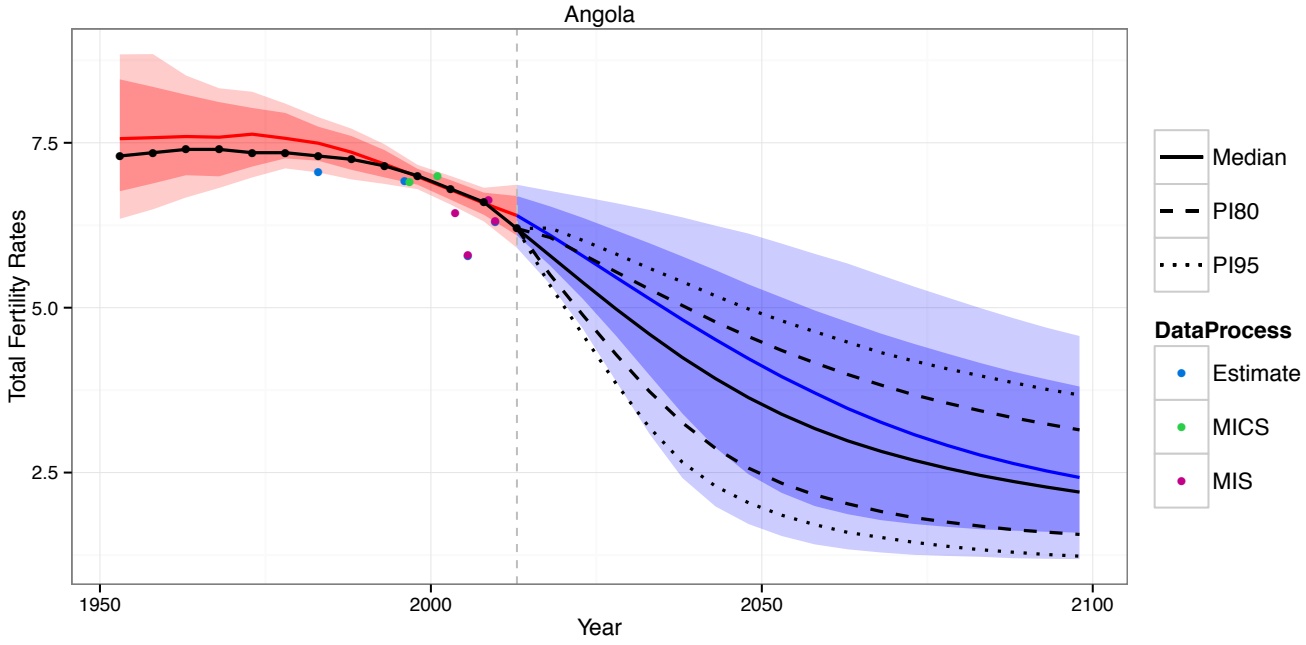
A.4 Estimation and Projection for All Countries

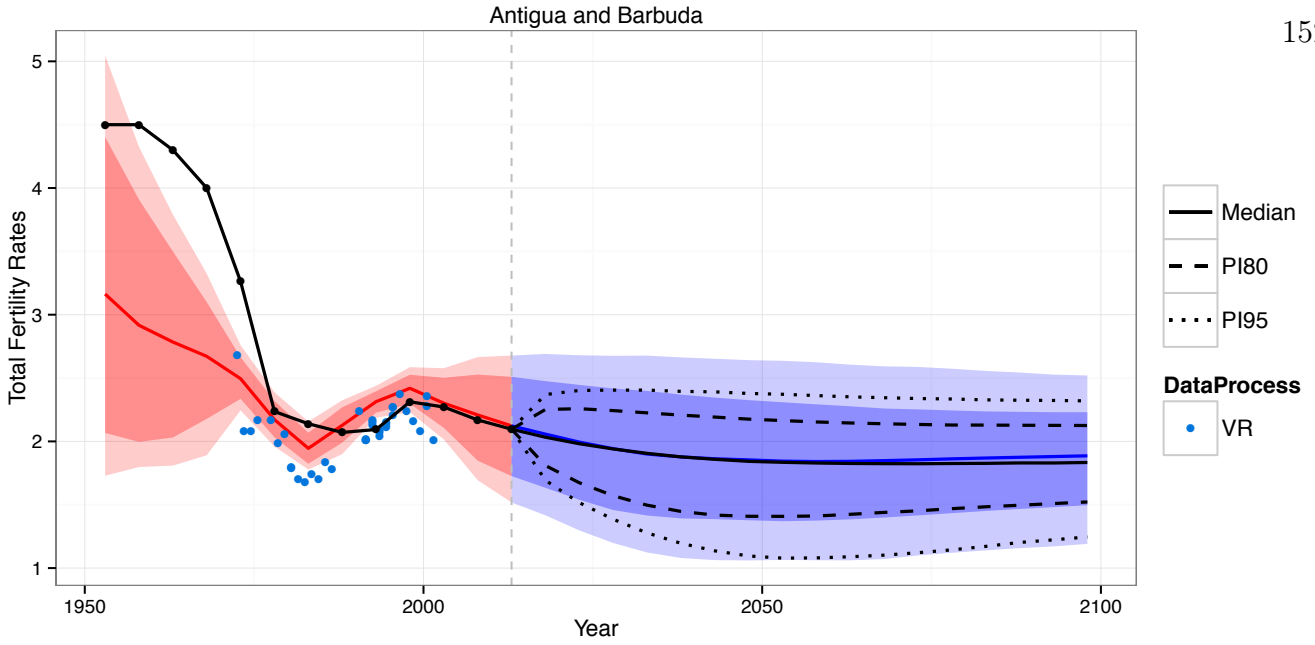
We give the results of estimation of past TFR from 1950 to 2015, and probabilistic projections of future TFR from 2015 to 2100 for all countries with current populations over 100,000 except for Antigua and Barbuda and Seychelles (for about 90,000).

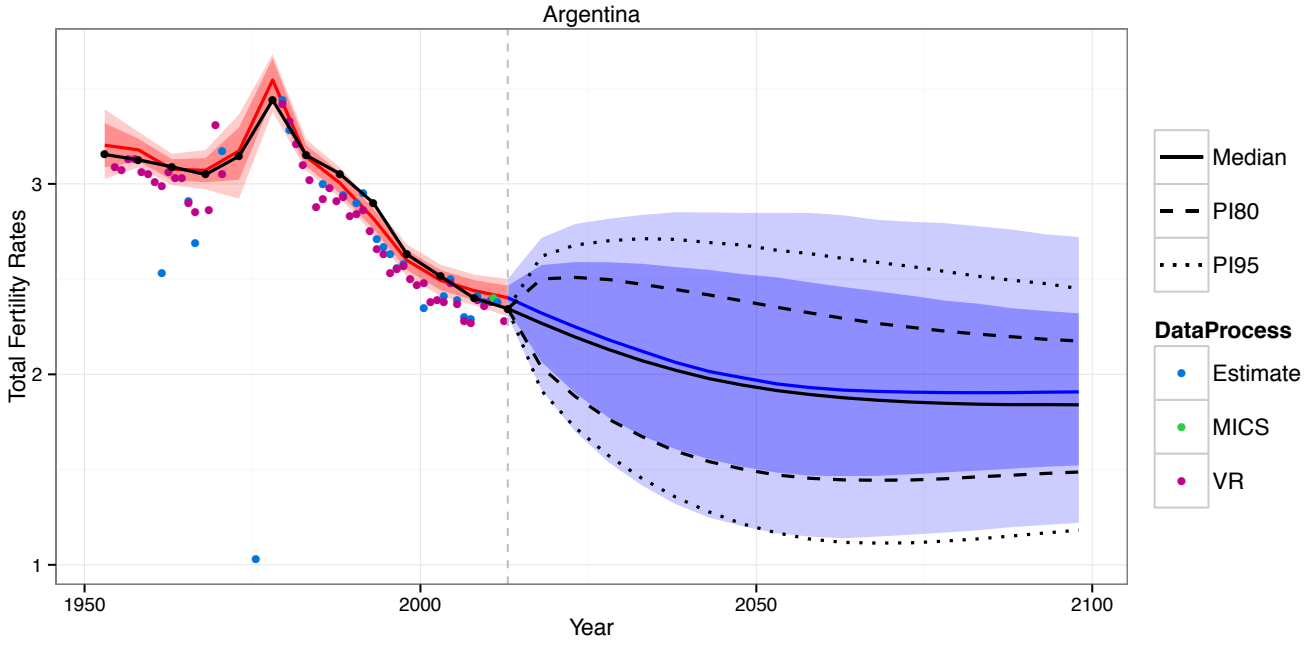

Afghanistan
Total Fertility Rates
10
5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS−NS
Estimate
MICS
Survey

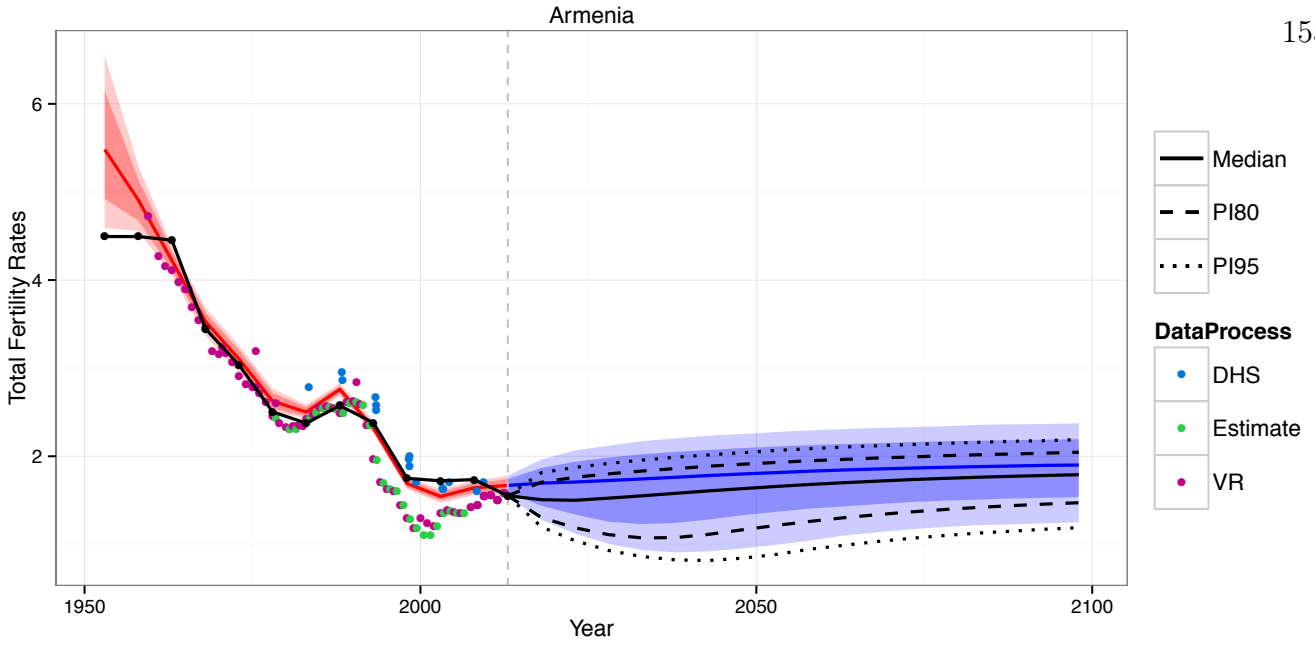

Albania
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
RHS
VR

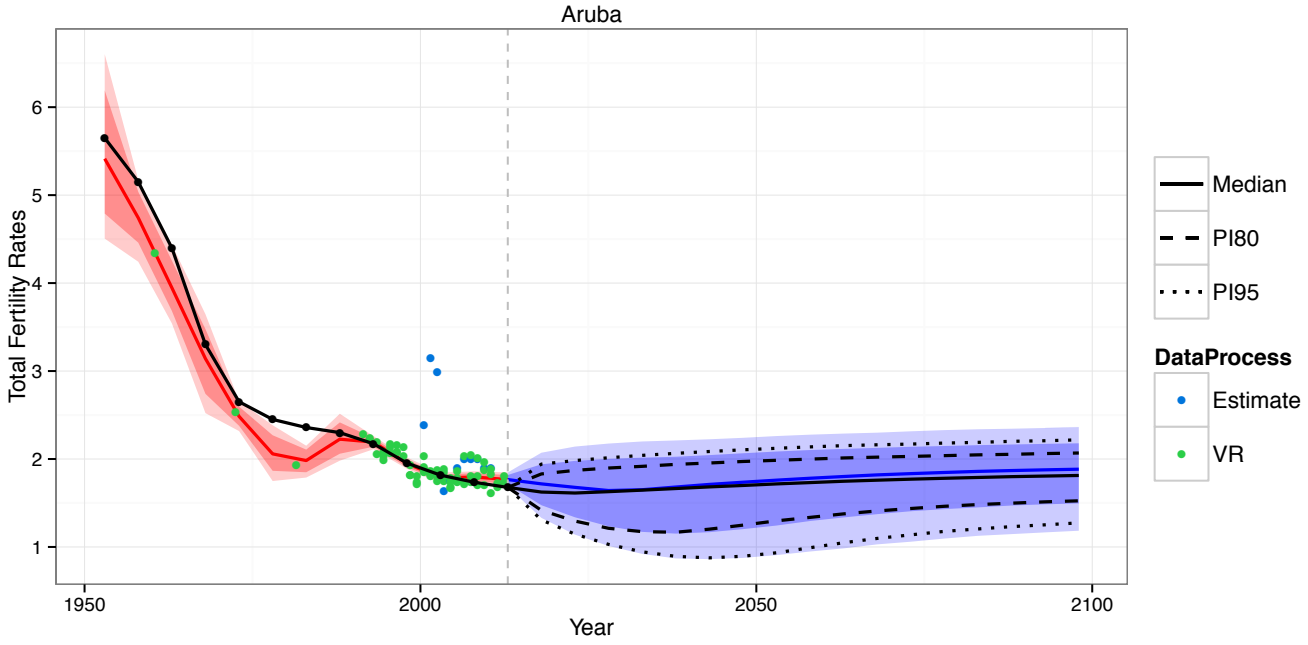

Algeria
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
MICS
PAPCHILD
PAPFAM
Survey
VR
WFS

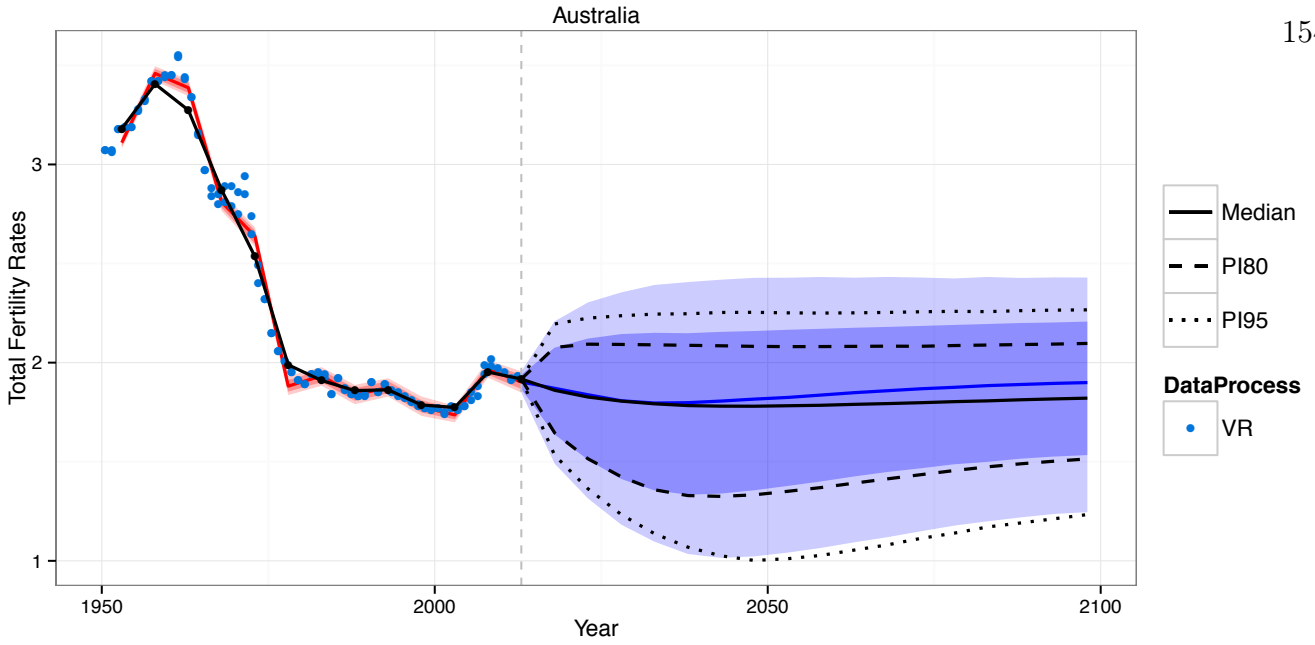

Angola
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
MICS
MIS

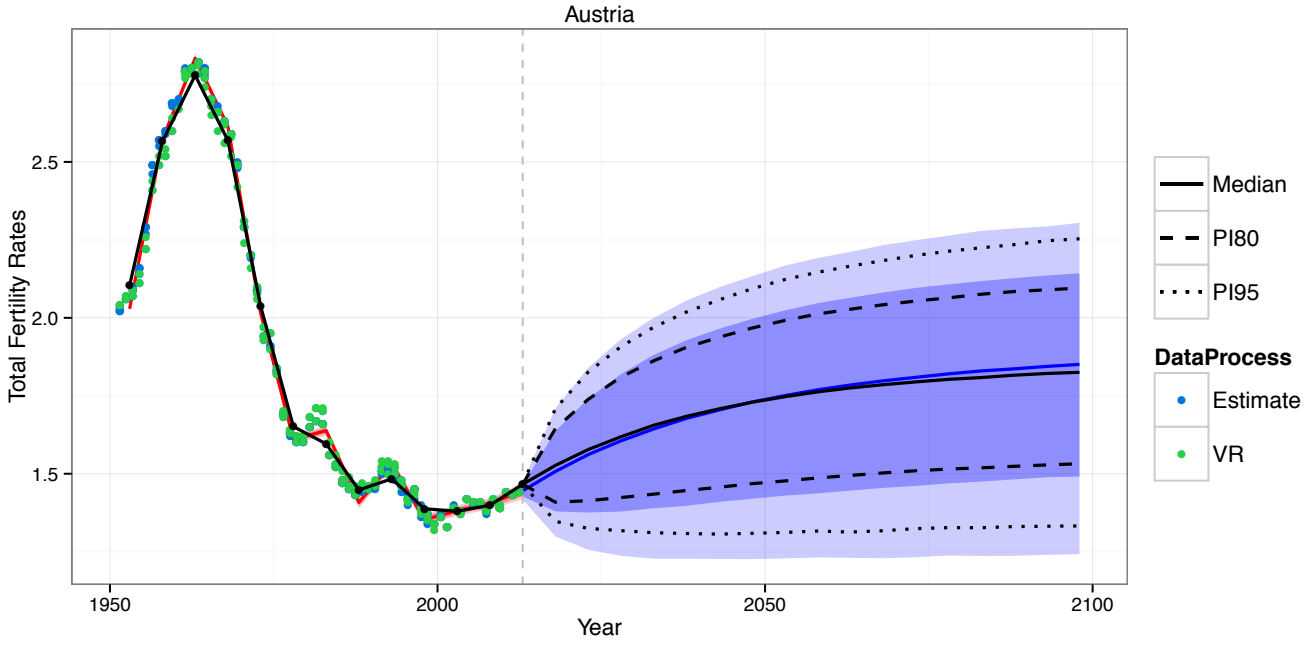

Antigua and Barbuda
Total Fertility Rates
Year
1950
2000
2050
2100
1
2
3
4
5
Median
PI80
PI95
DataProcess
VR

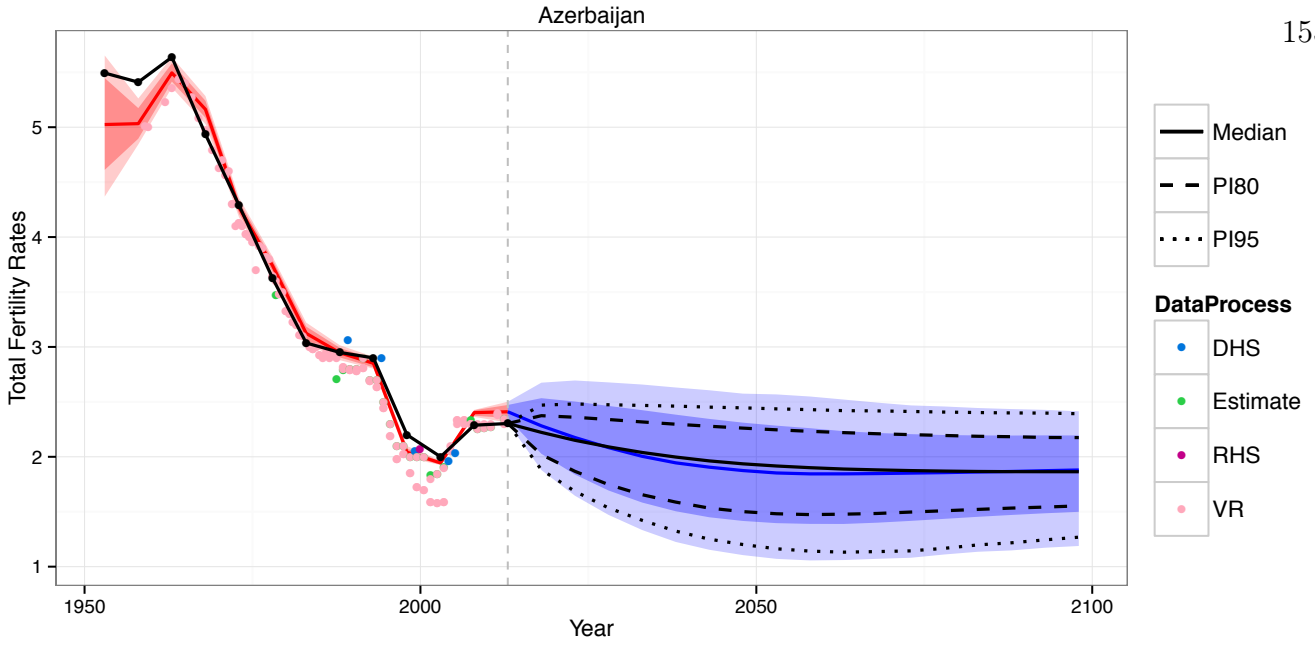

Argentina
Total Fertility Rates
Year
1950
2000
2050
2100
1
2
3
Median
PI80
PI95
DataProcess
Estimate
MICS
VR

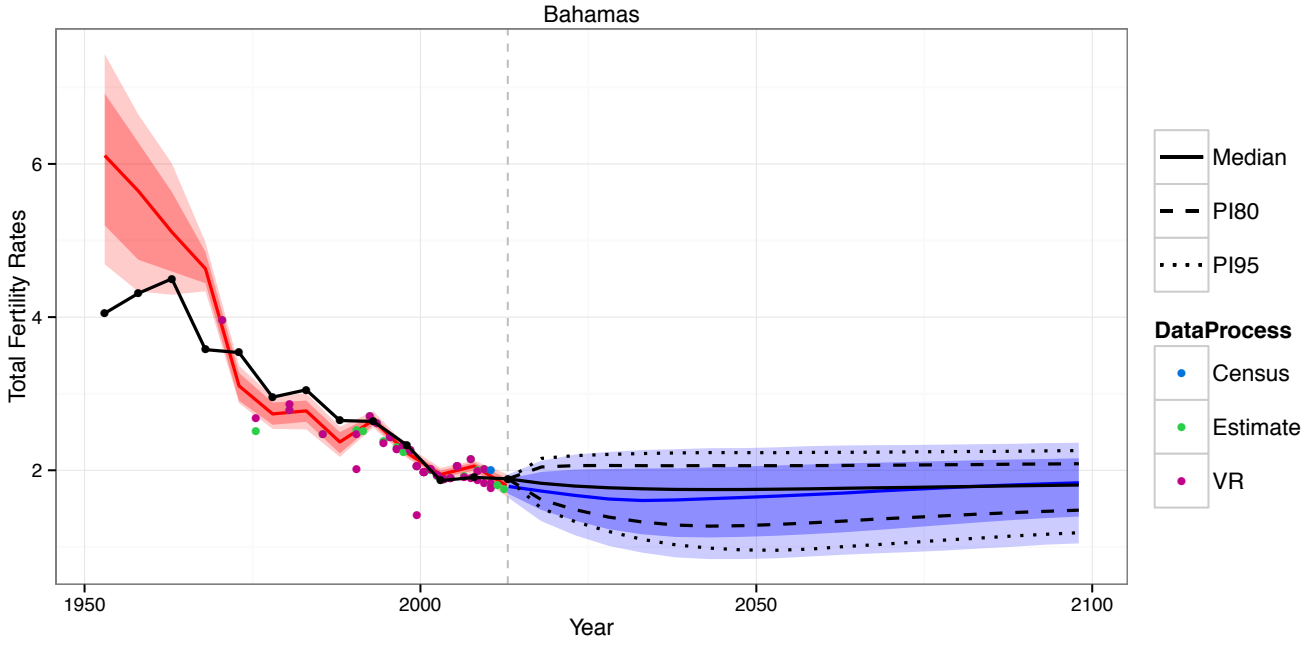

Armenia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
VR

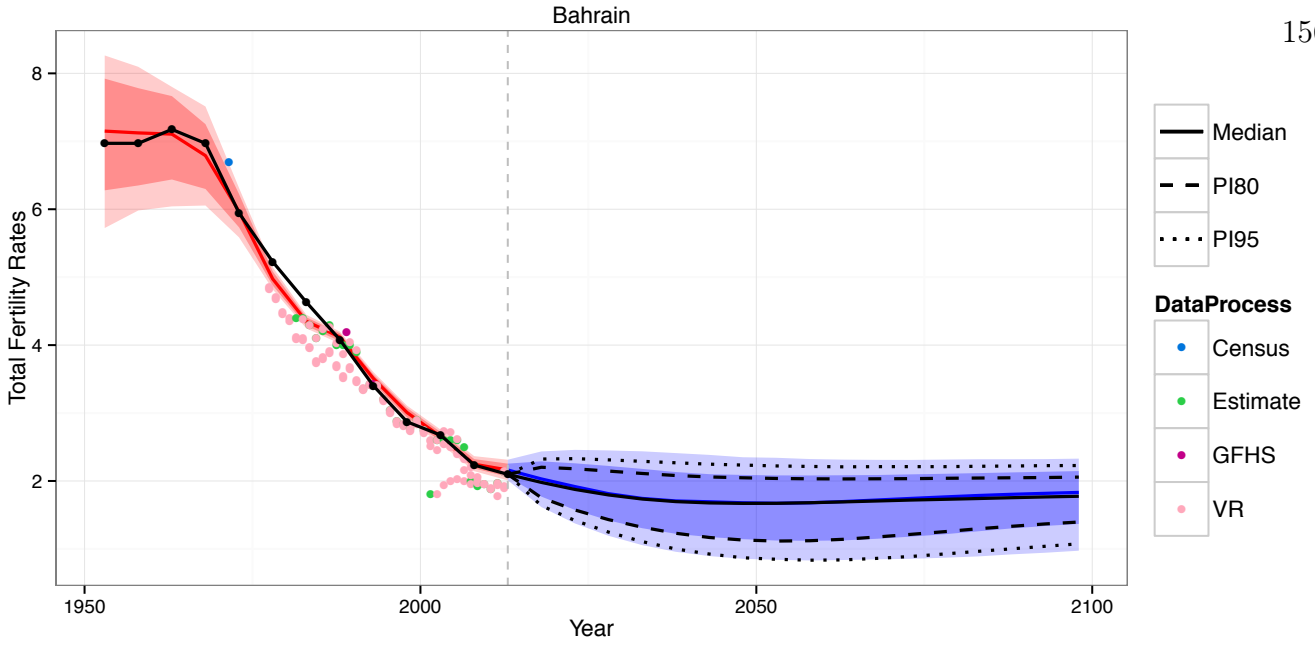

Aruba
Total Fertility Rates
6
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

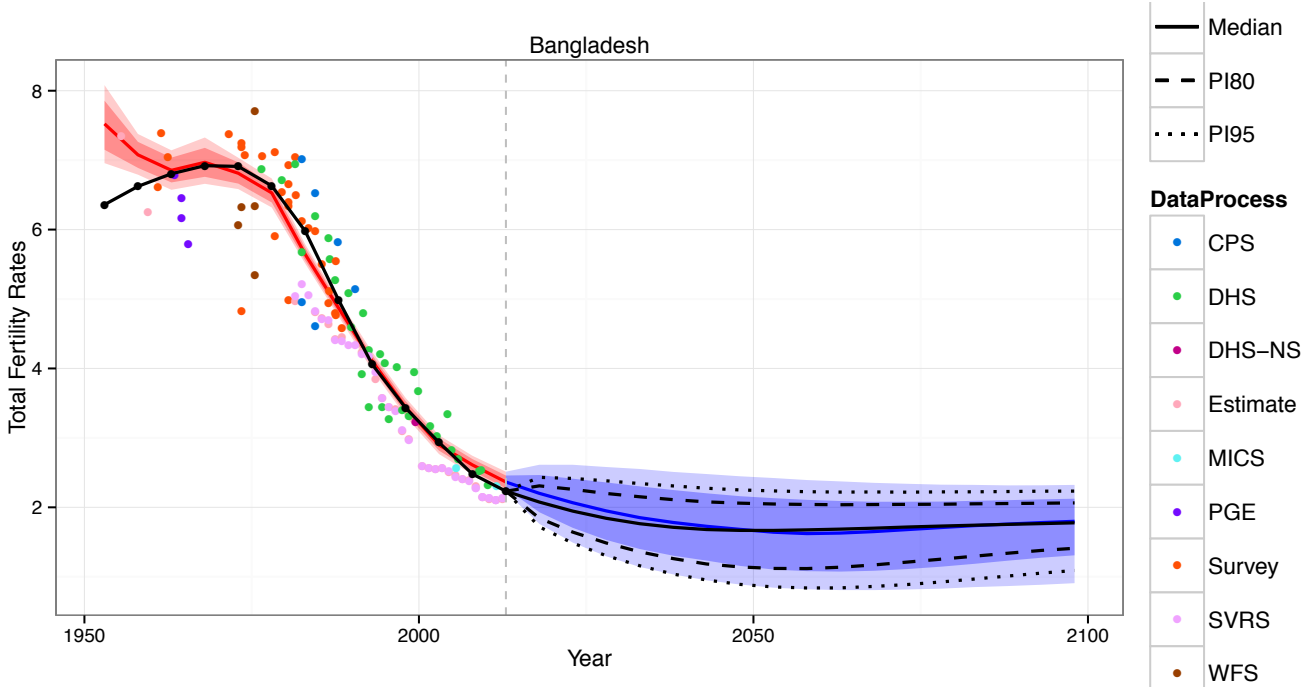

Australia
Total Fertility Rates
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
VR

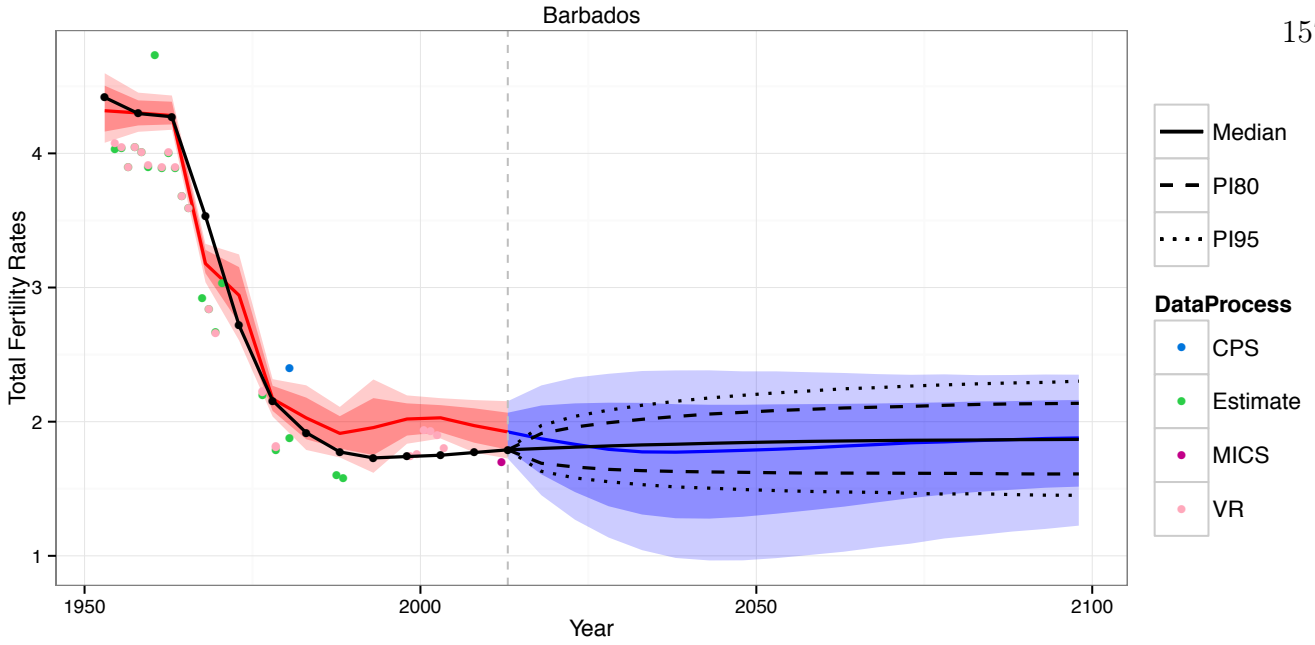

Austria
Total Fertility Rates
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

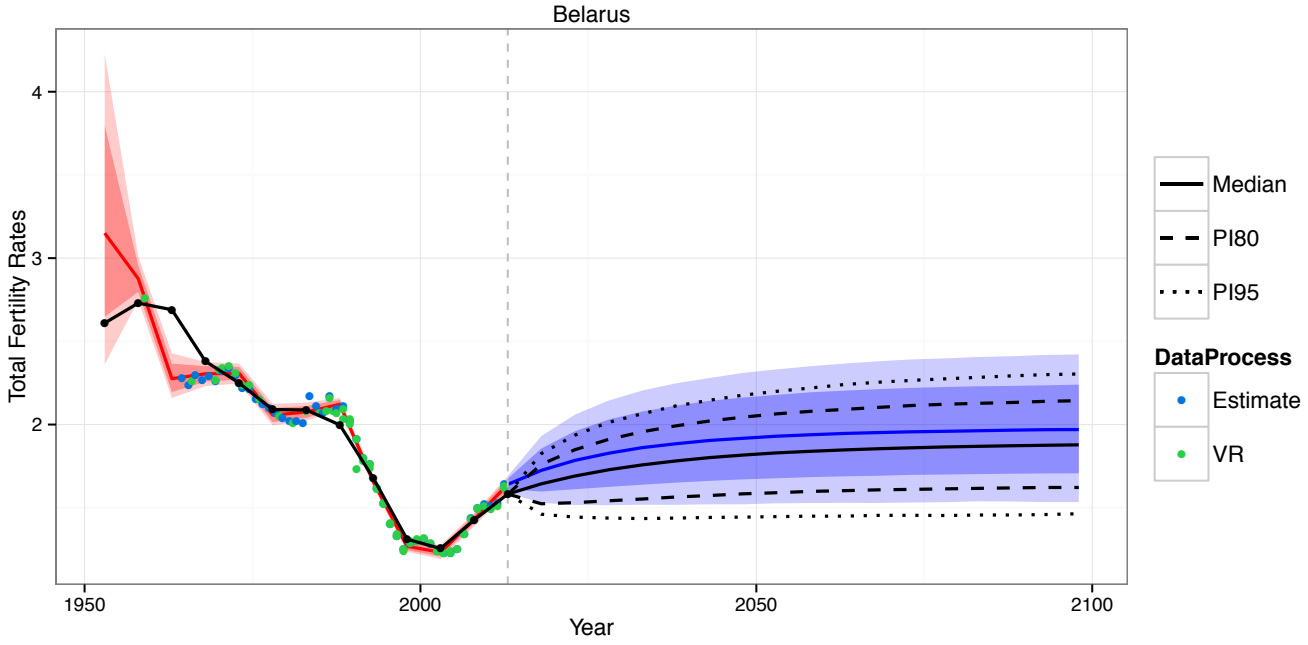

Azerbaijan
Total Fertility Rates
1
2
3
4
5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
RHS
VR

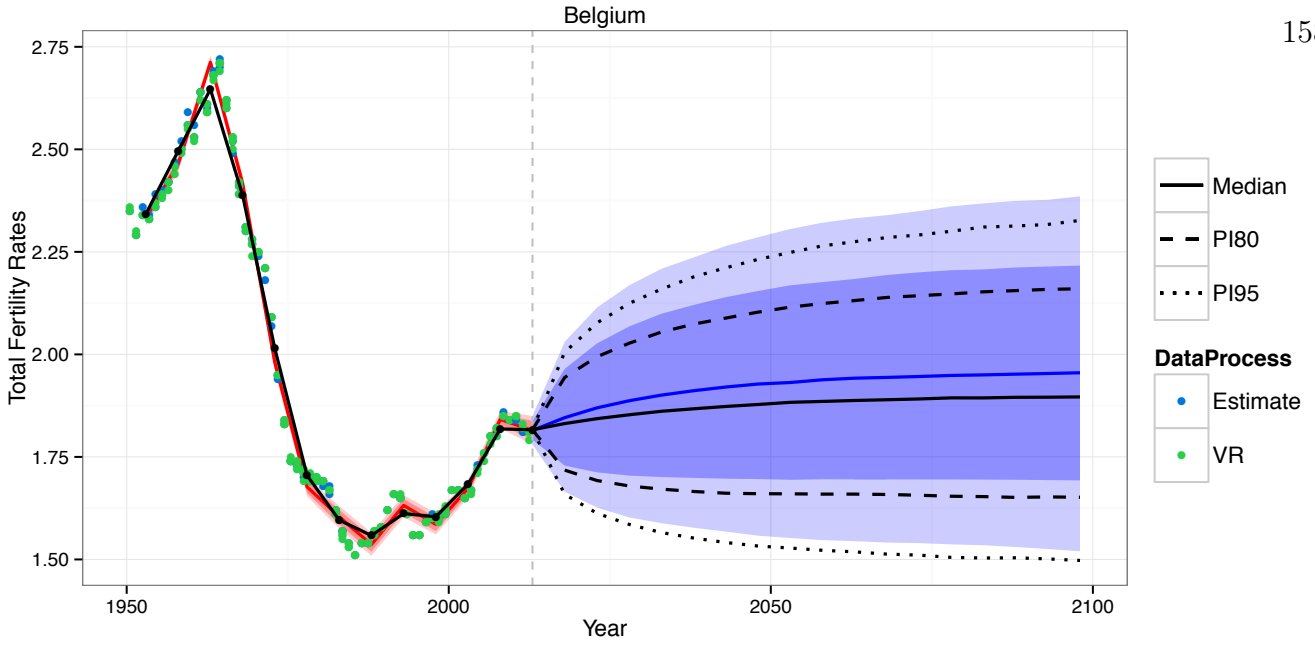

Bahamas
Total Fertility Rates
2
4
6
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

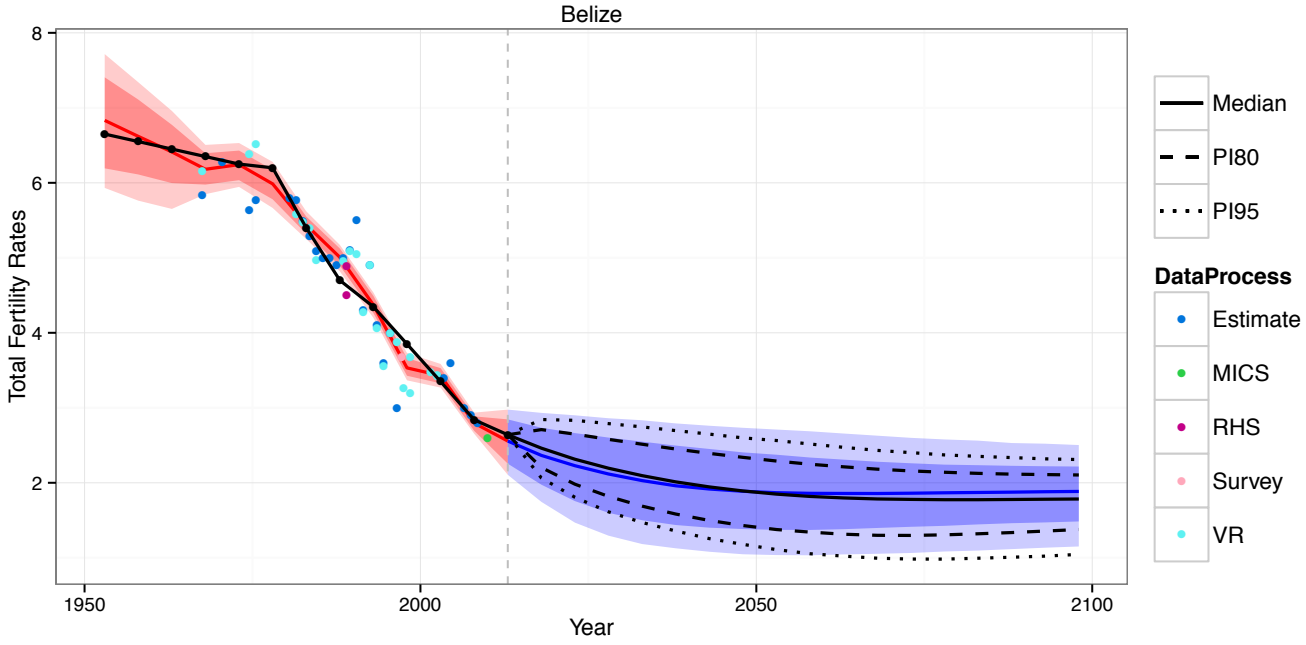

Bahrain
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
GFHS
VR

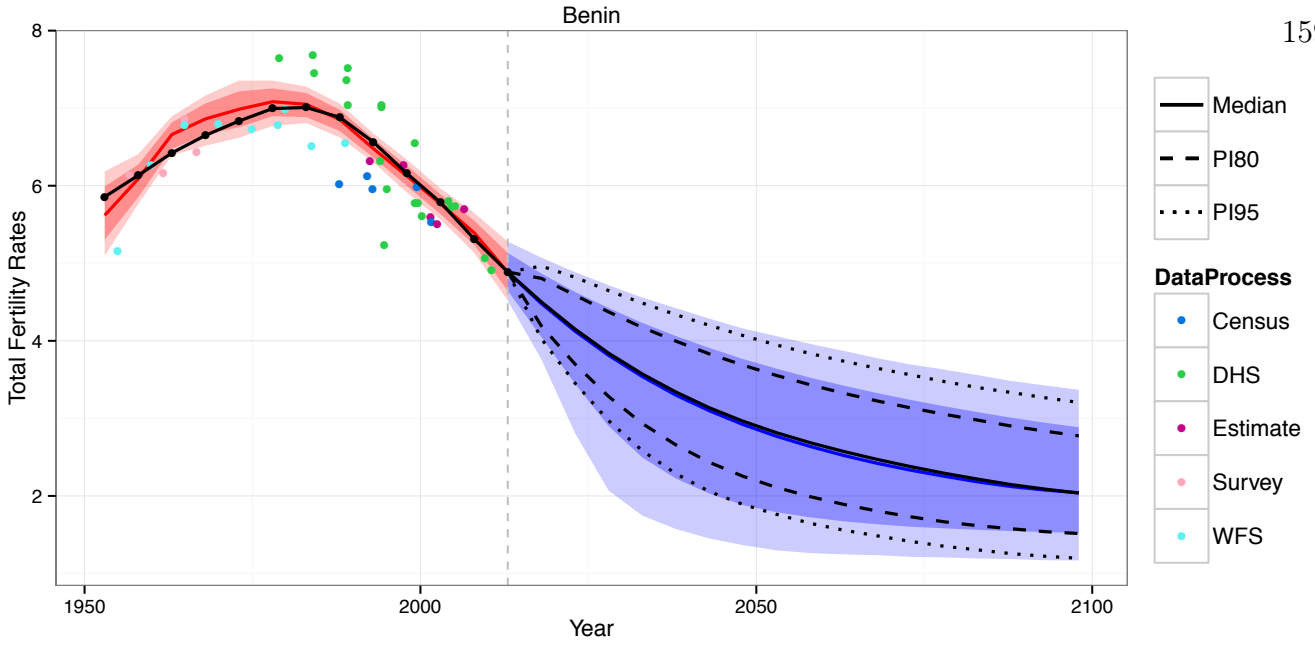

Bangladesh
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
CPS
DHS
DHS−NS
Estimate
MICS
PGE
Survey
SVRS
WFS

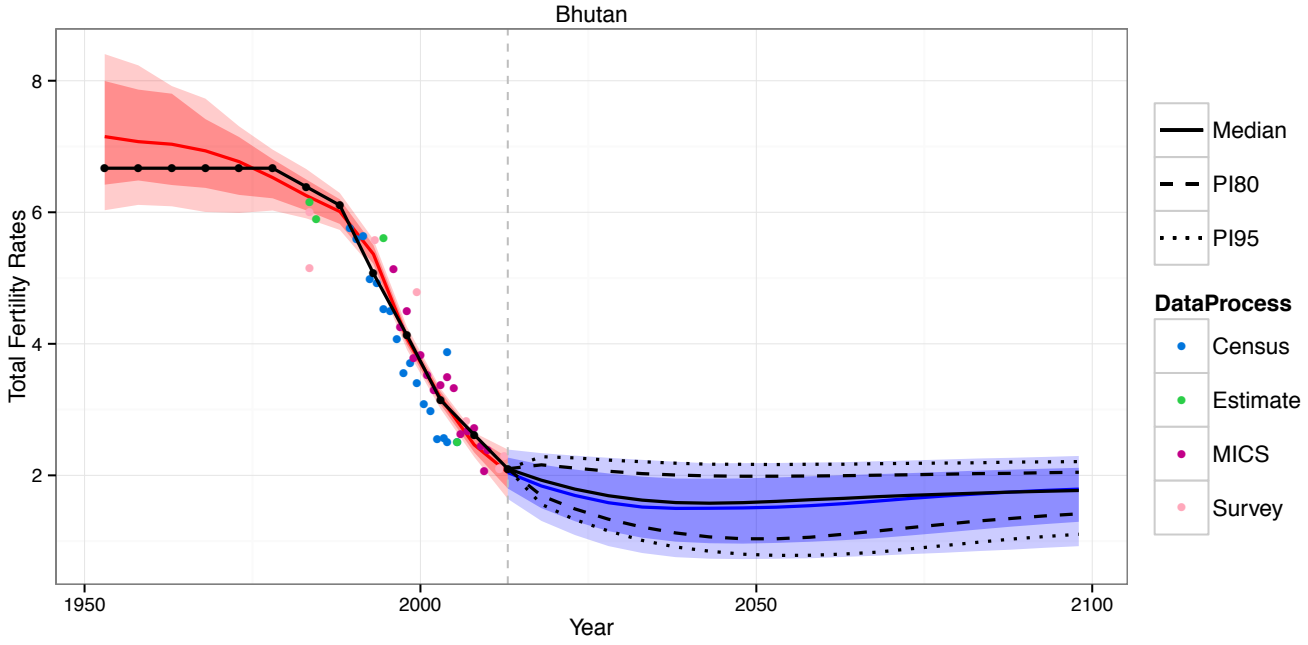

Barbados
Total Fertility Rates
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
CPS
Estimate
MICS
VR

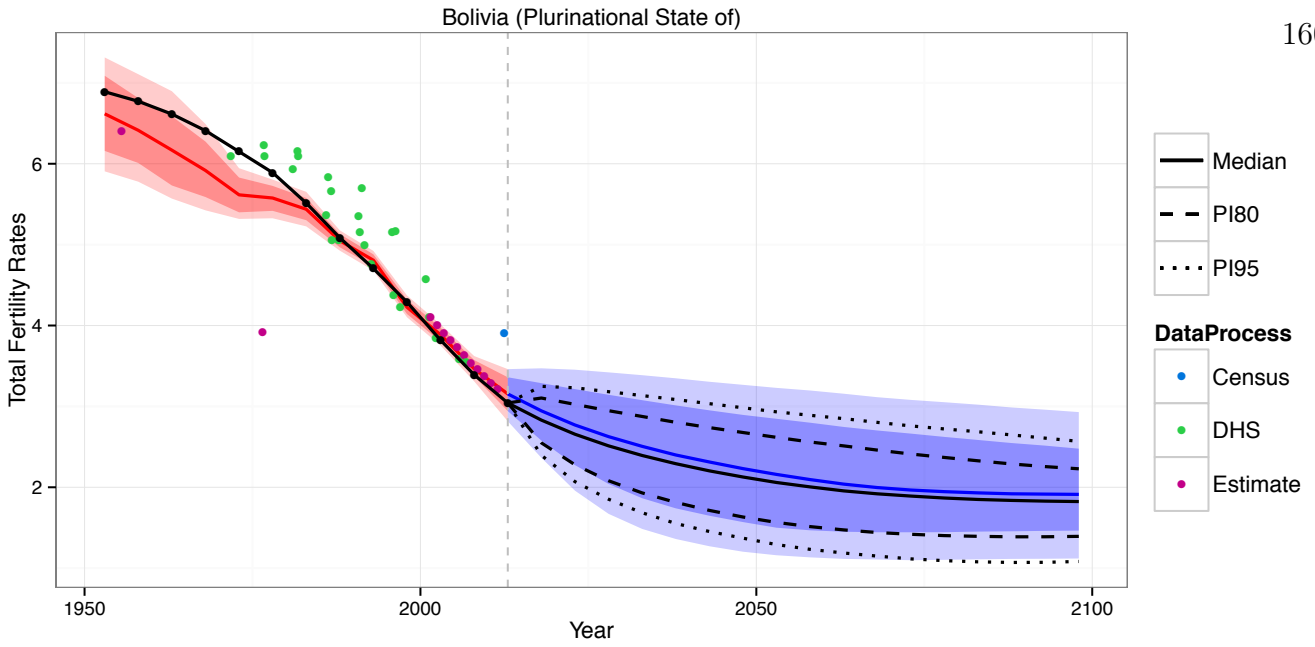

Belarus
Total Fertility Rates
4
3
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

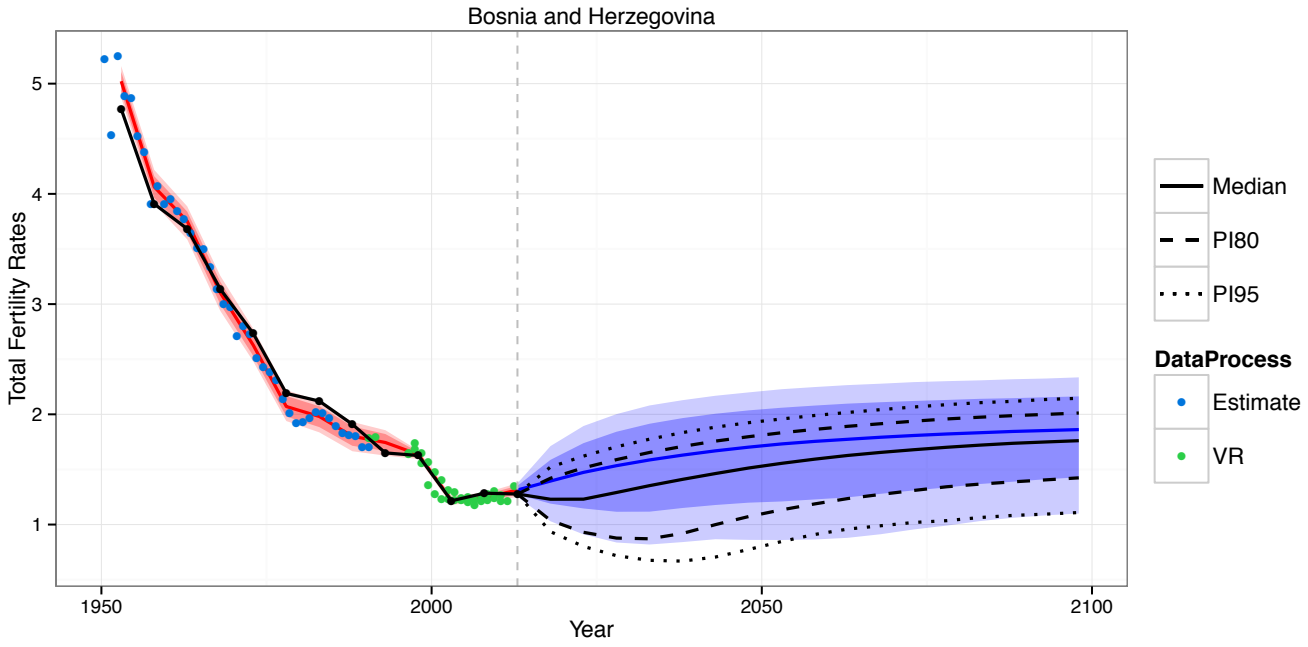

Belgium
Total Fertility Rates
2.75
2.50
2.25
2.00
1.75
1.50
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

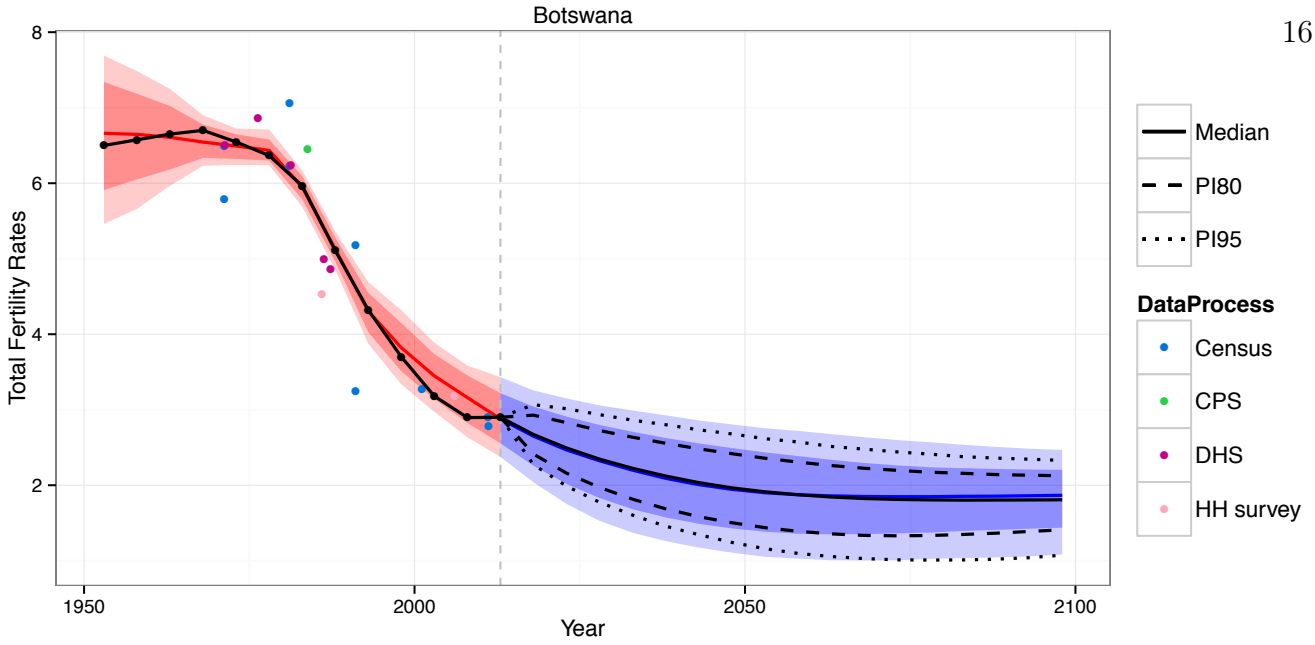

Belize
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
MICS
RHS
Survey
VR

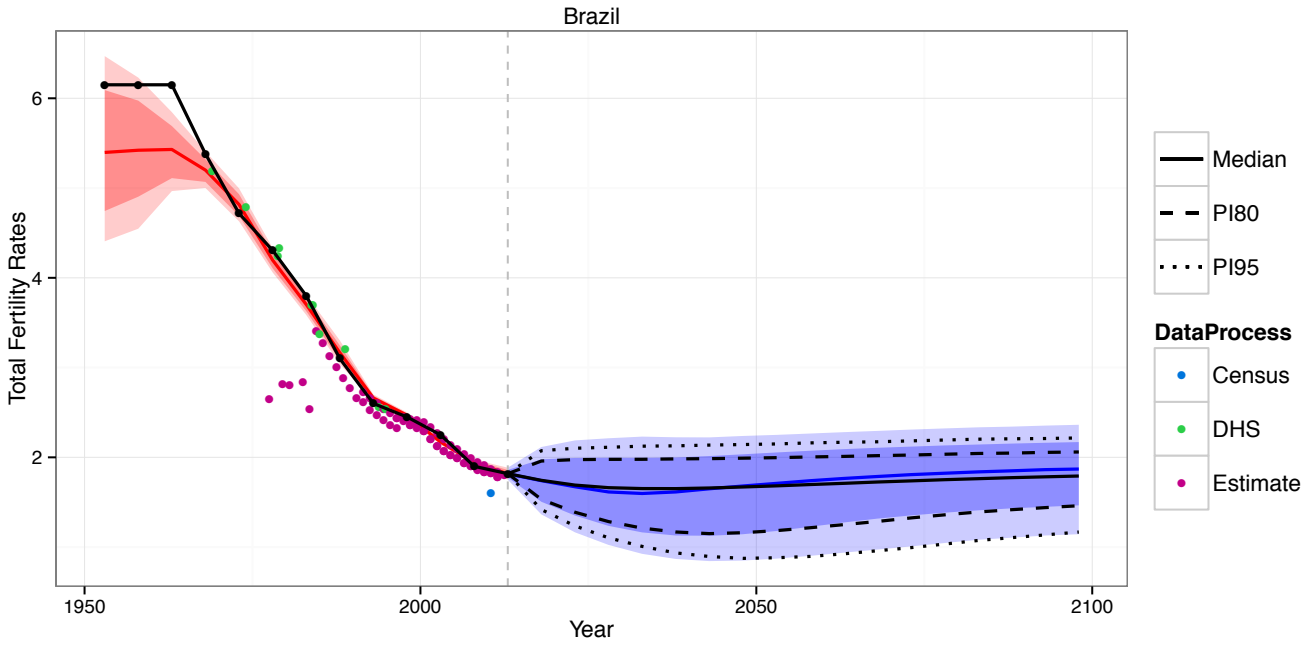

Benin
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
Survey
WFS

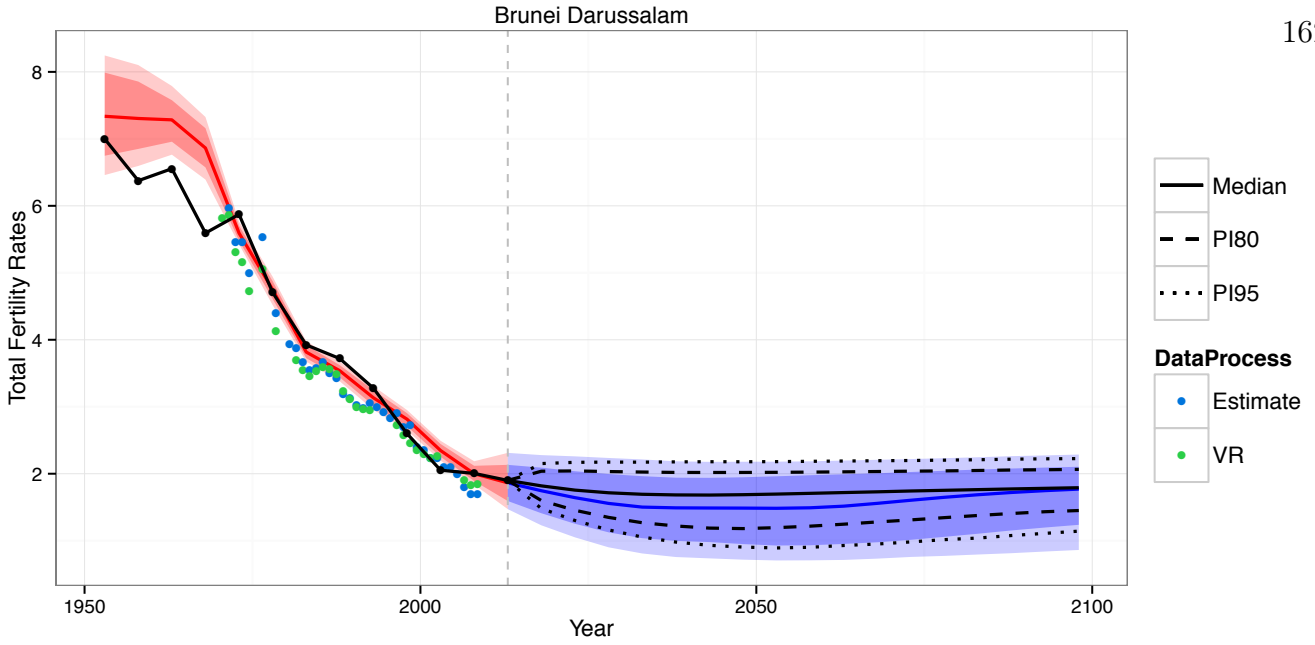

Bhutan
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
MICS
Survey

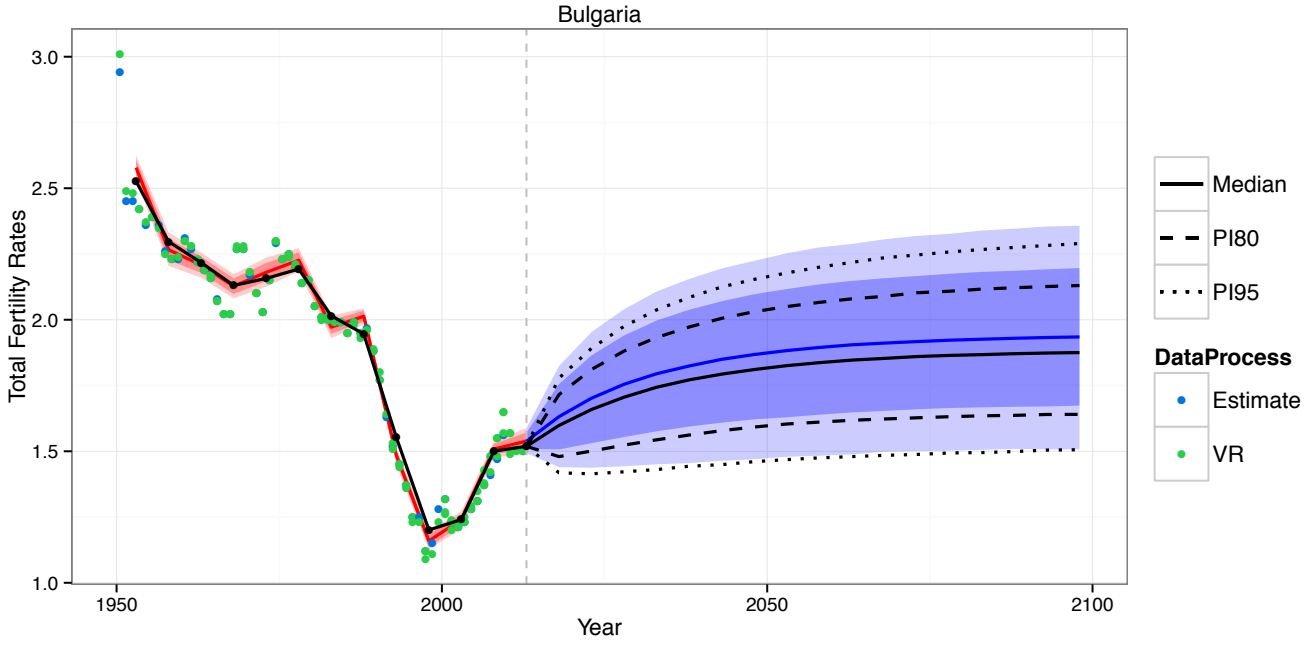

Bolivia (Plurinational State of)
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate

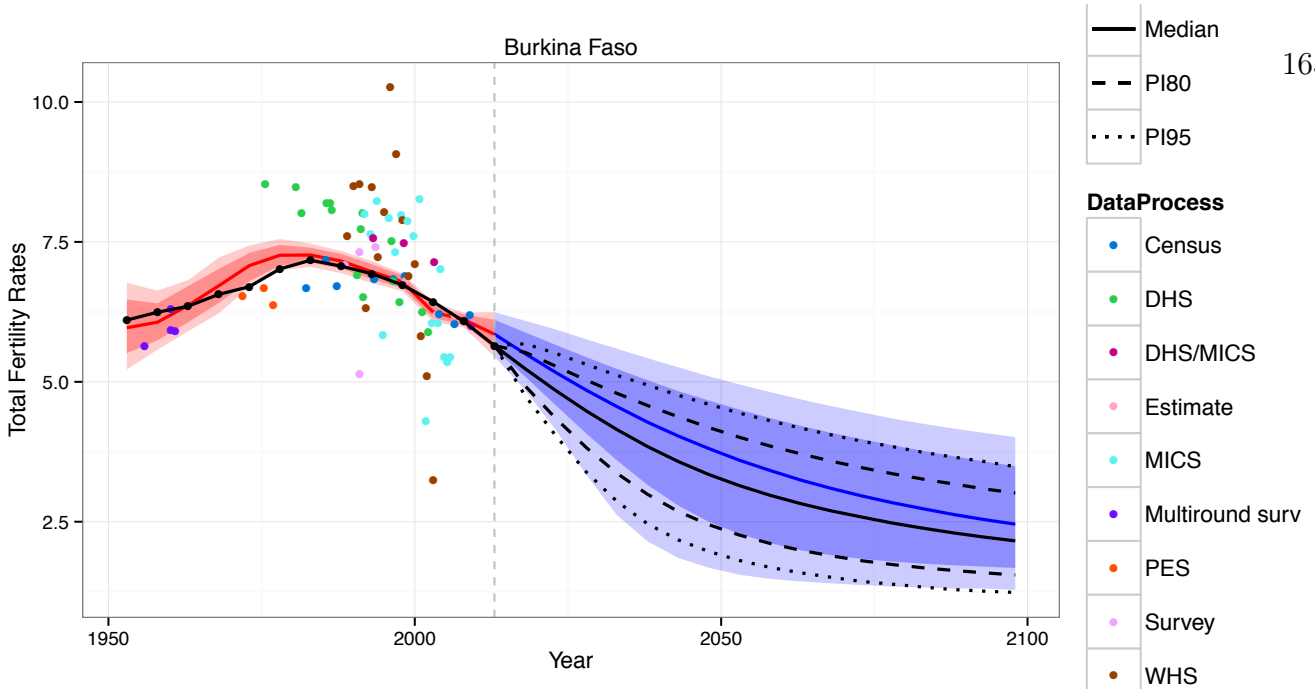

Bosnia and Herzegovina
Total Fertility Rates
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

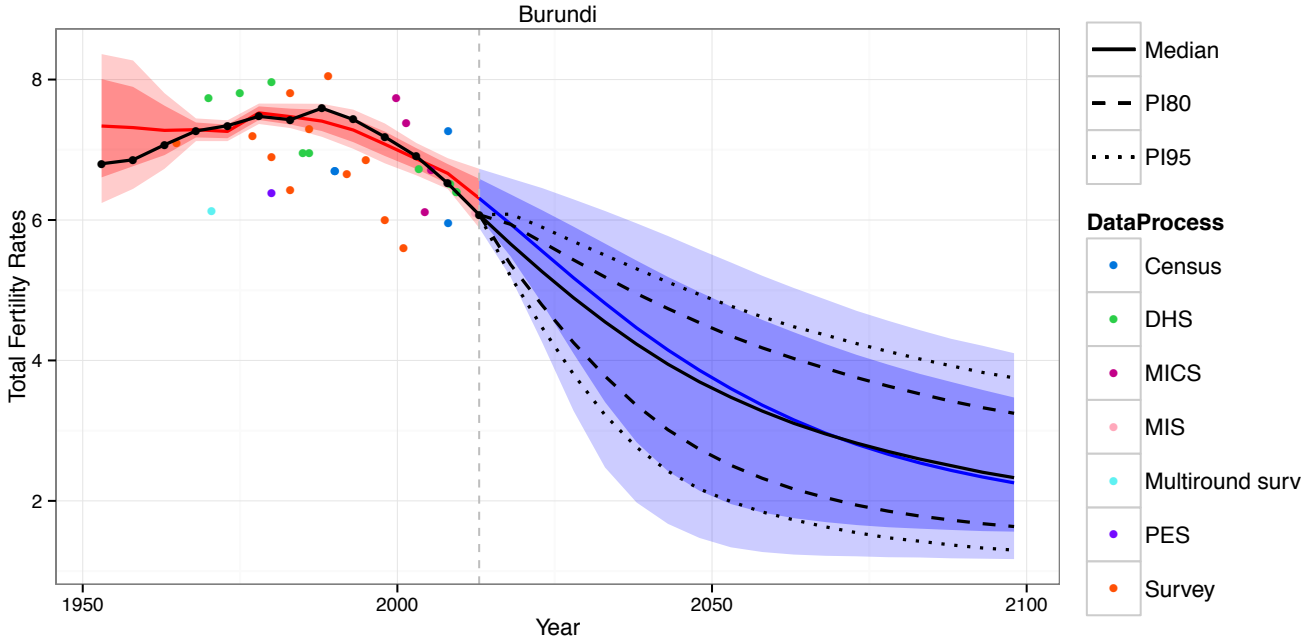

Botswana
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
DHS
HH survey

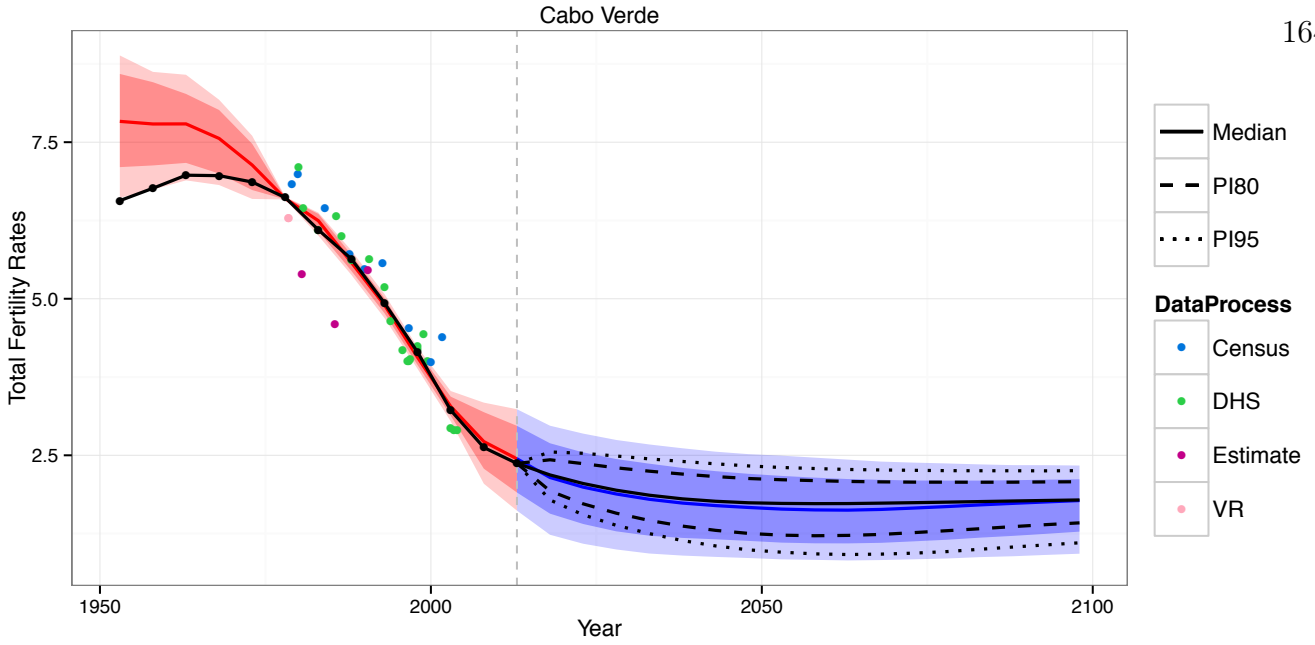

Brazil
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate

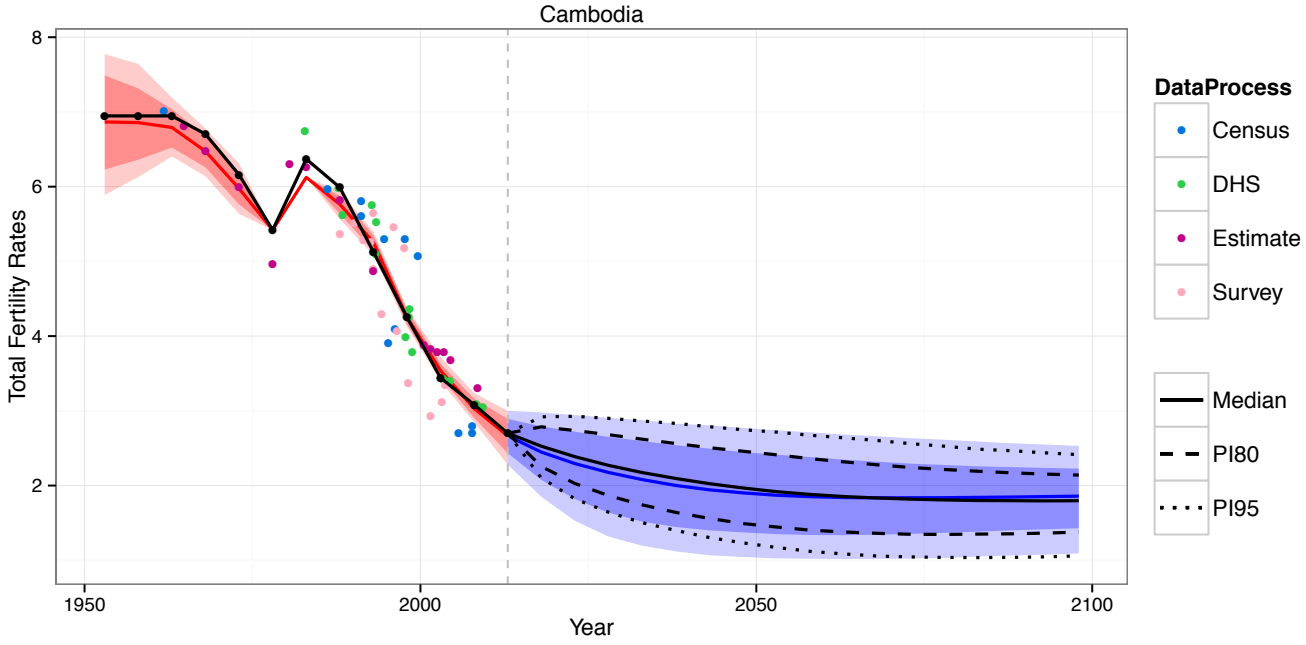

Brunei Darussalam
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

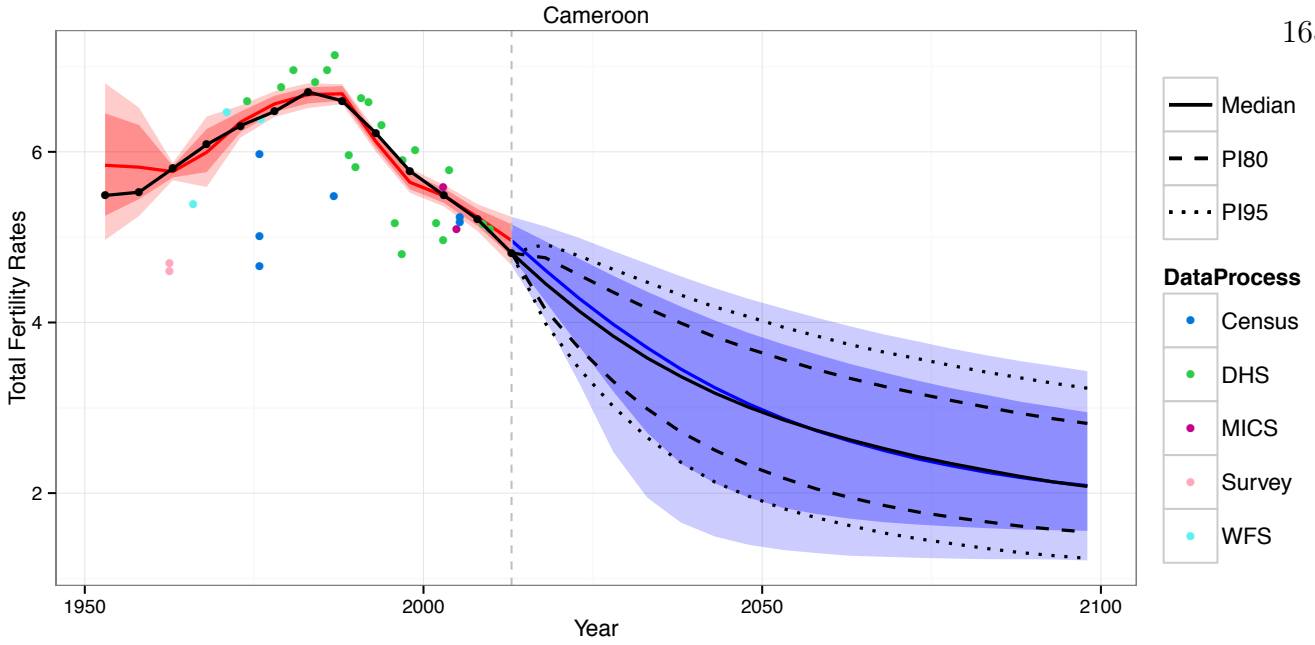

Bulgaria
Total Fertility Rates
3.0
2.5
2.0
1.5
1.0
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

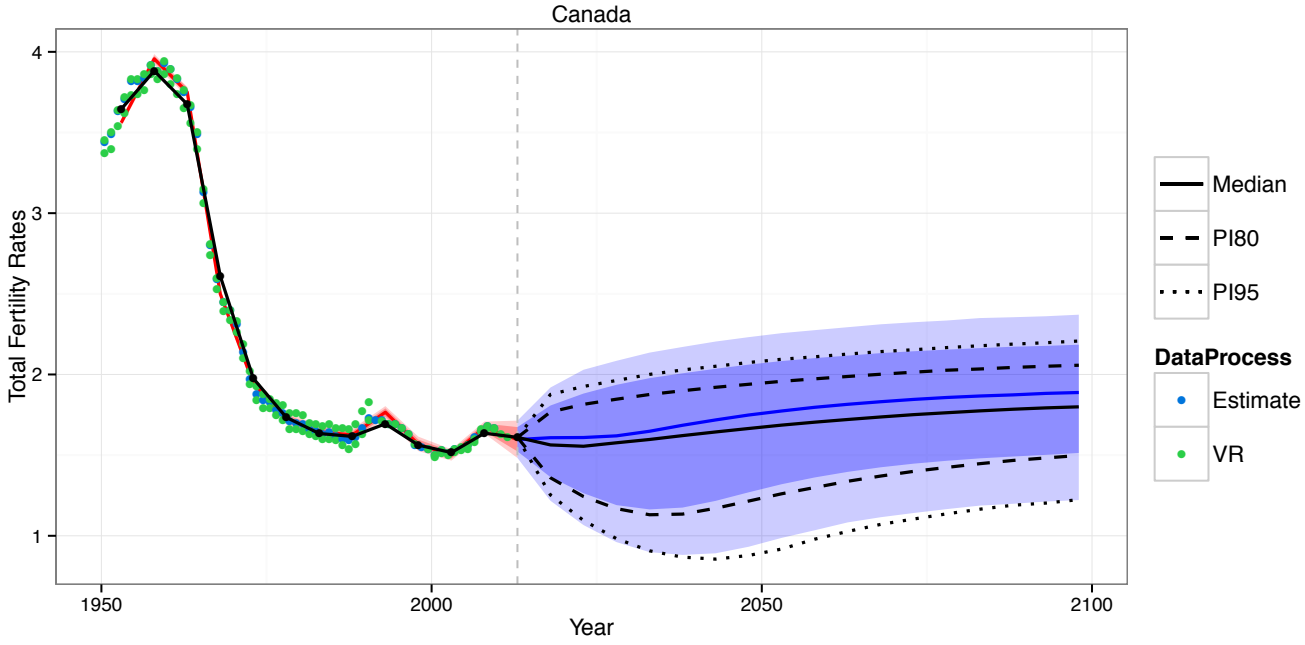

Burkina Faso
Total Fertility Rates
10.0
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
DHS/MICS
Estimate
MICS
Multiround surv
PES
Survey
WHS

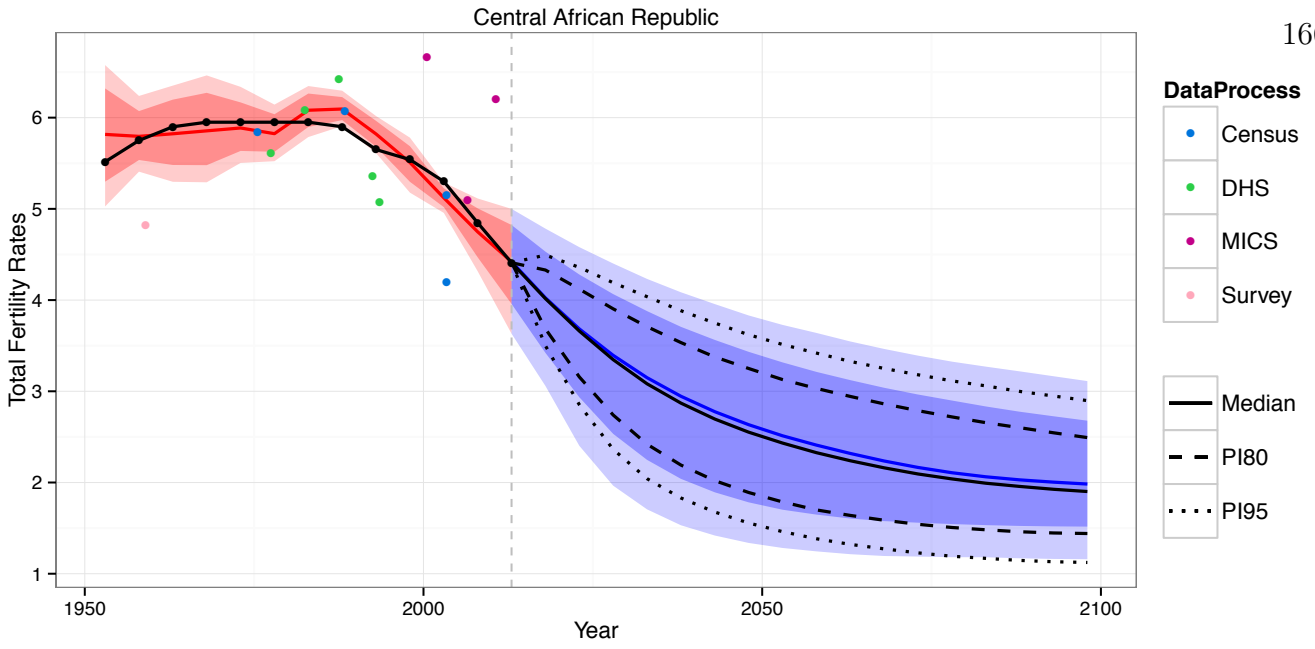

Burundi
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
MICS
MIS
Multiround surv
PES
Survey

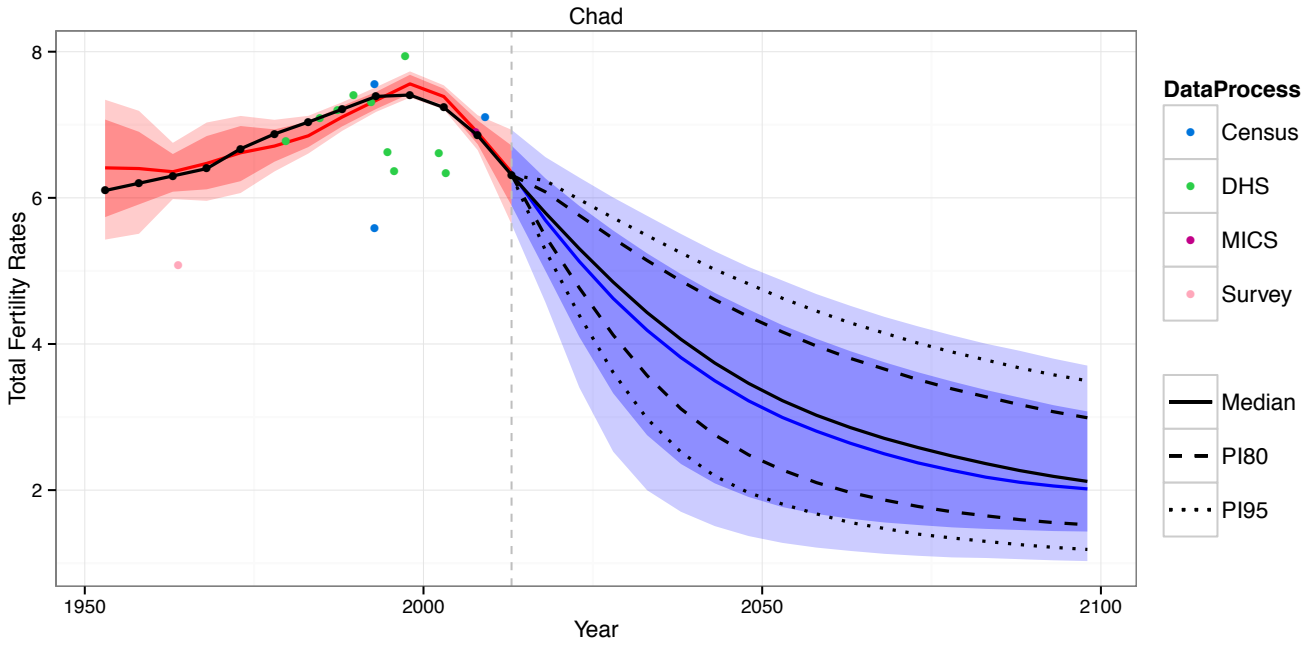

Cabo Verde
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
VR

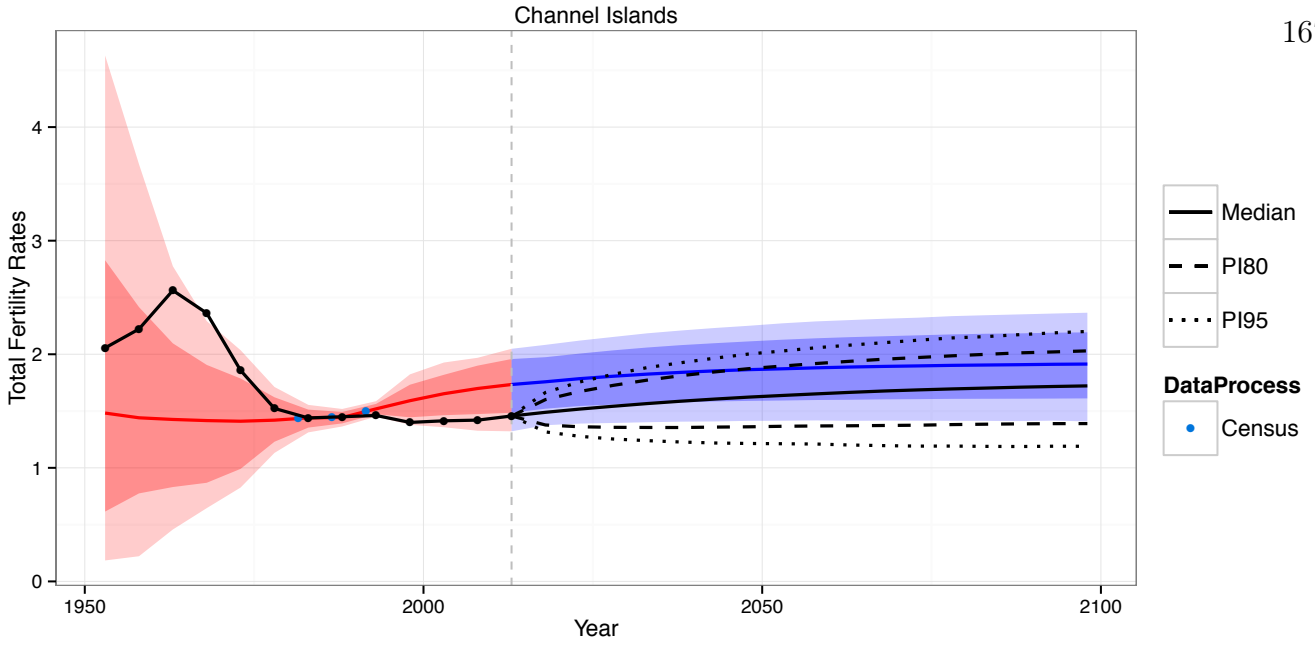

Cambodia
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Estimate
Survey
Median
PI80
PI95

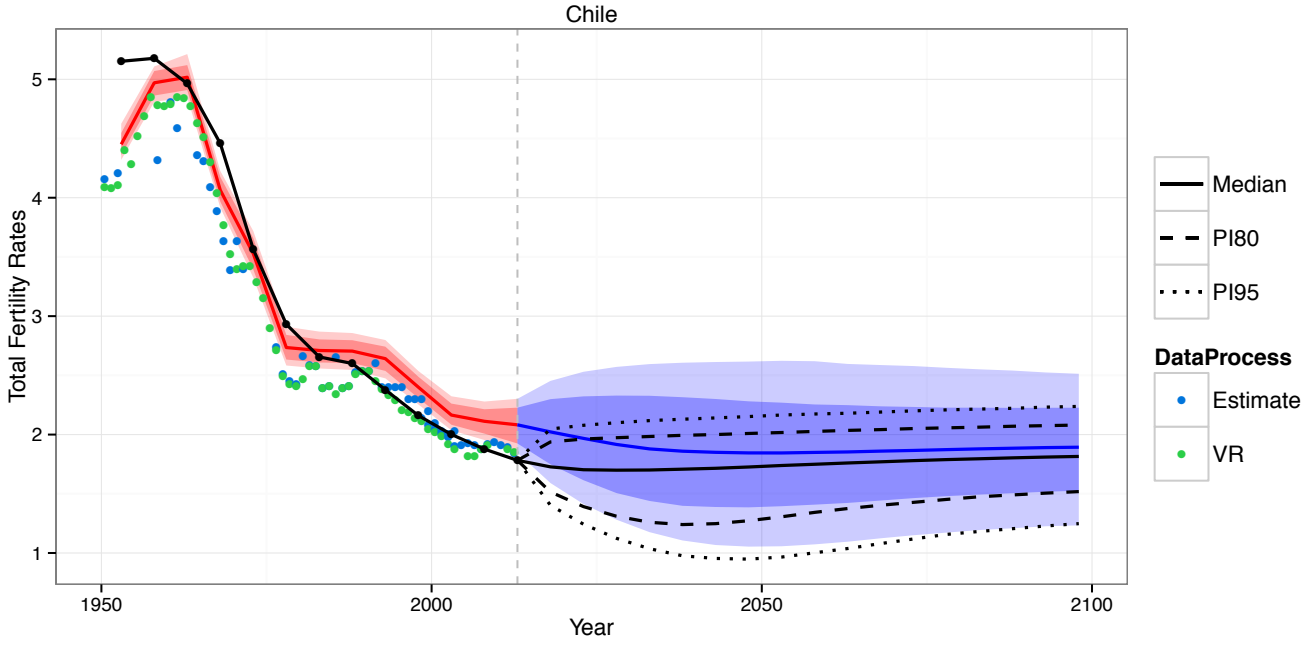

Cameroon
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
MICS
Survey
WFS

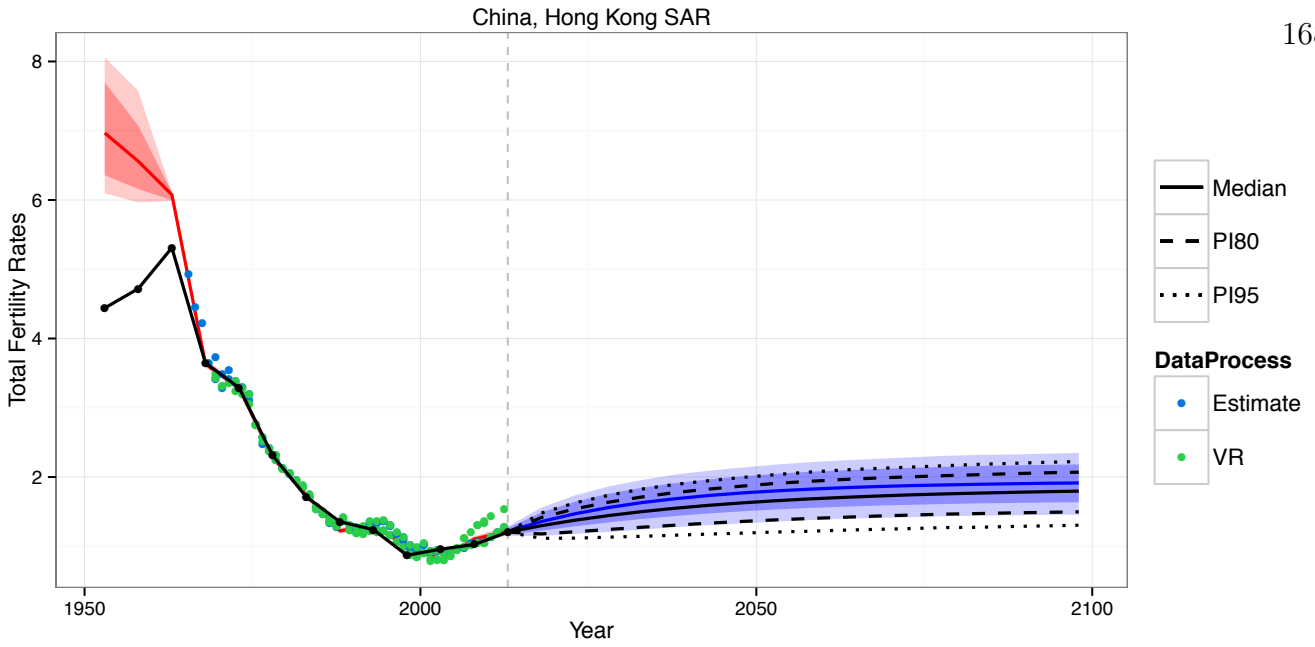

Canada
Total Fertility Rates
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

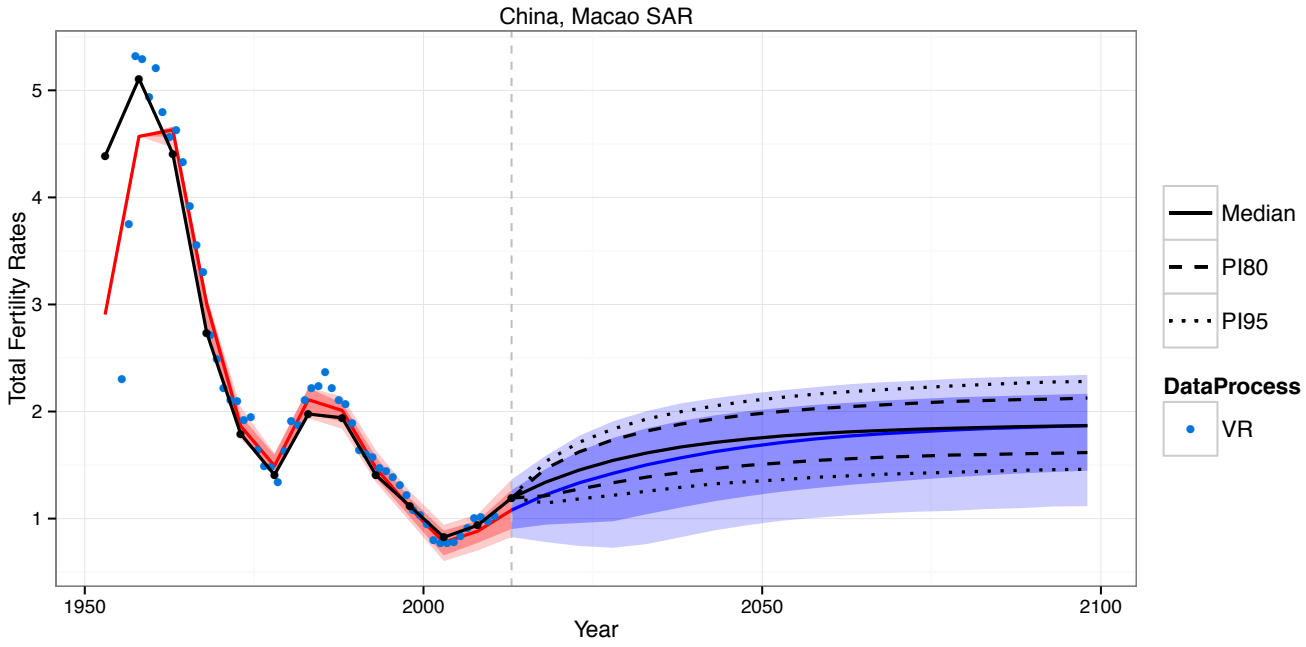

Central African Republic
Total Fertility Rates
Year
1950
2000
2050
2100
1
2
3
4
5
6
DataProcess
Census
DHS
MICS
Survey
Median
PI80
PI95

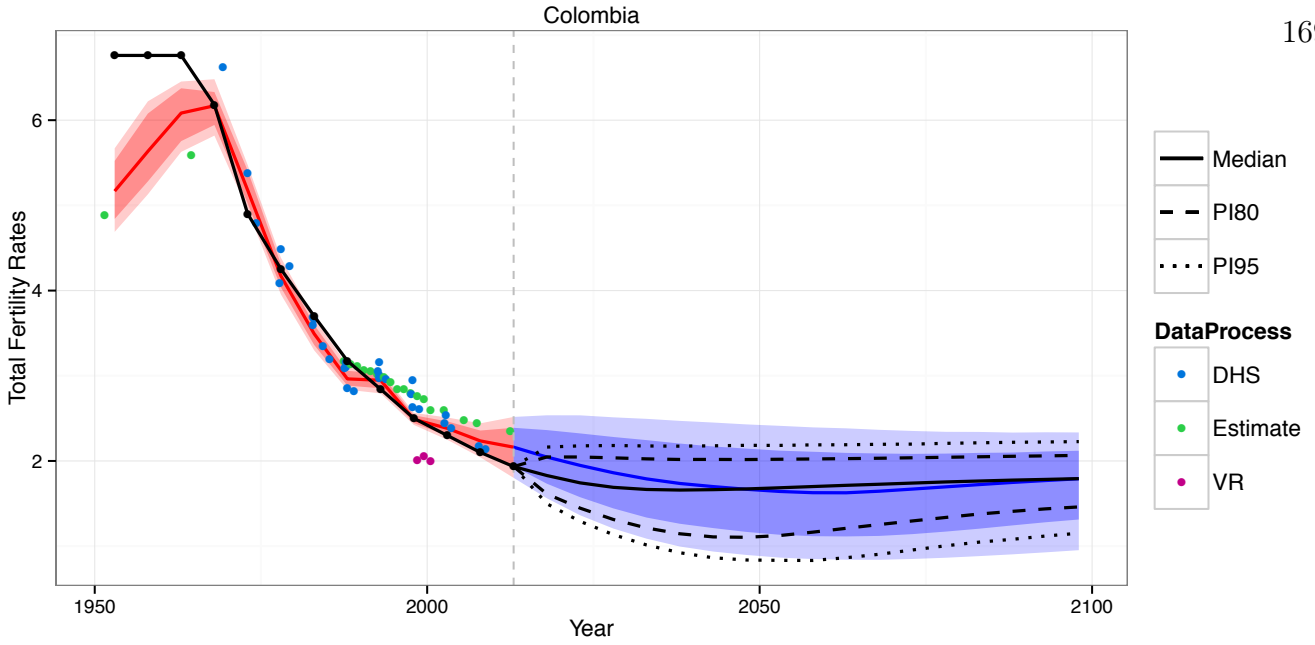

Chad
Total Fertility Rates
Year
1950
2000
2050
2100
2
4
6
8
DataProcess
Census
DHS
MICS
Survey
Median
PI80
PI95

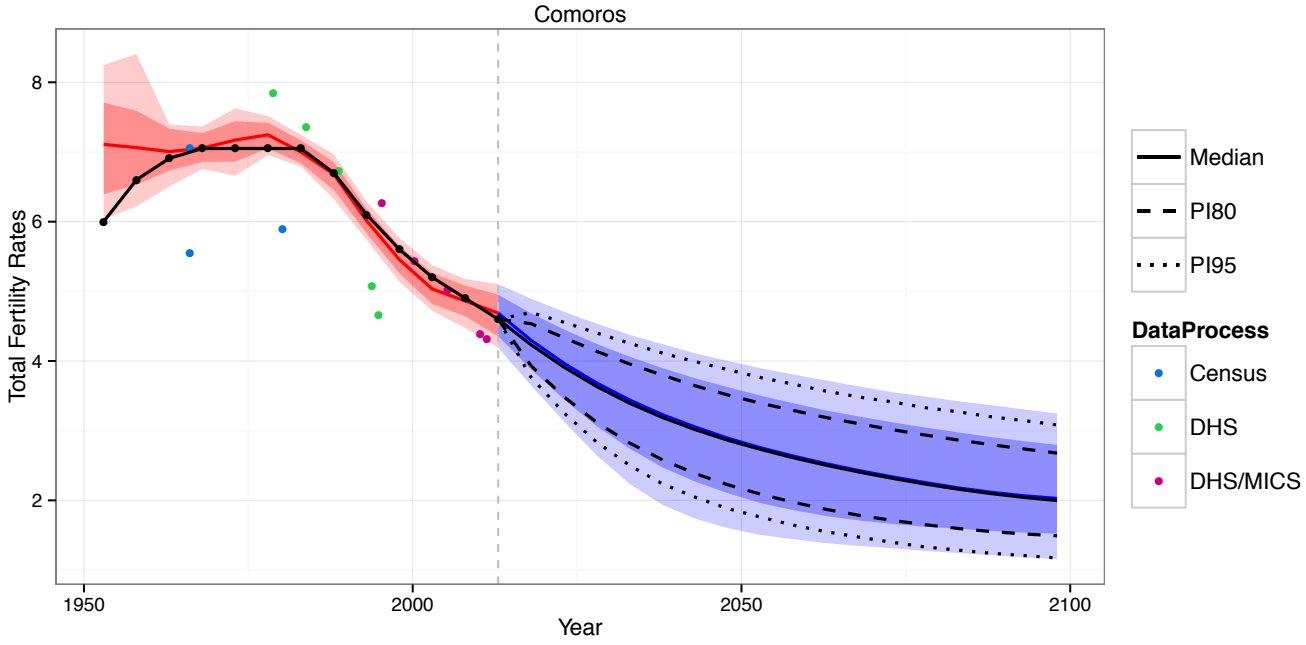

Channel Islands
Total Fertility Rates
0
1
2
3
4
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census

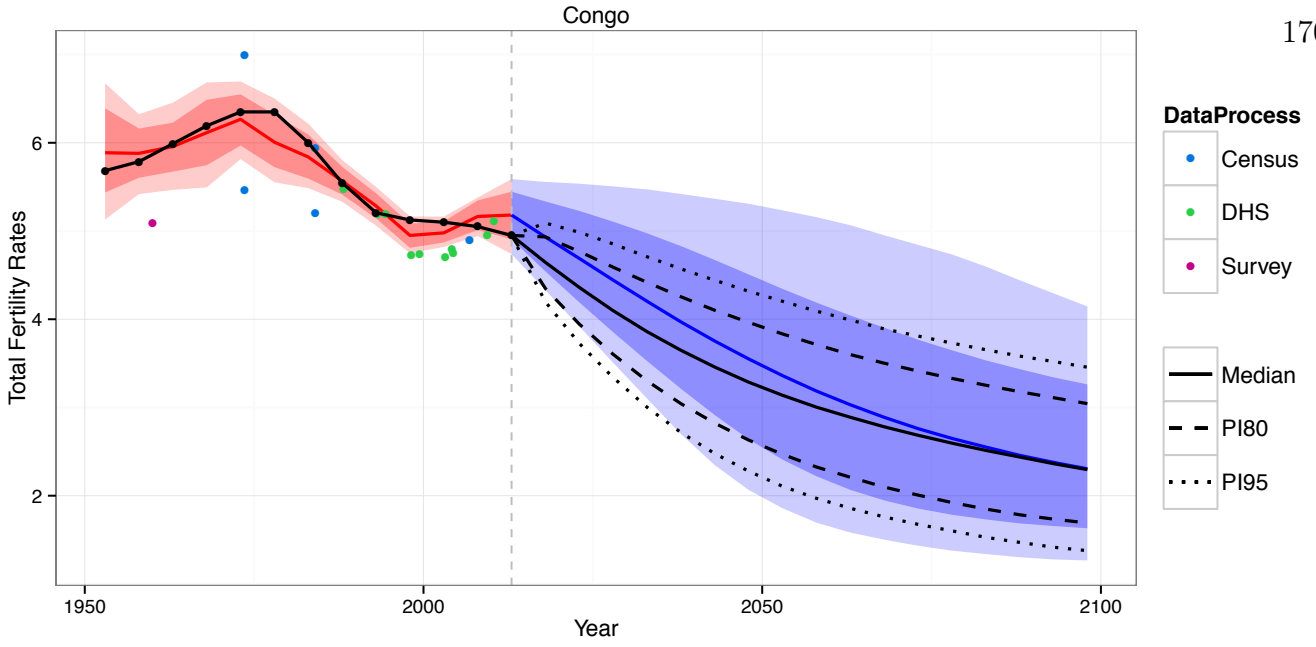

Chile
Total Fertility Rates
1
2
3
4
5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

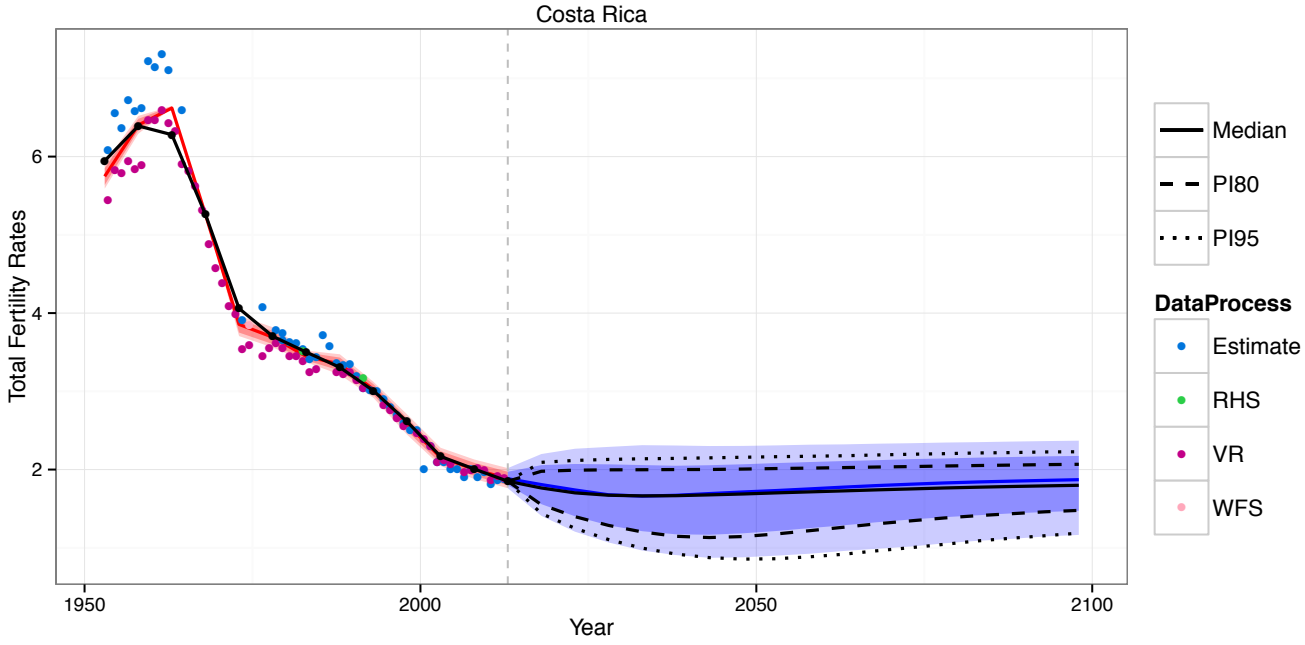

China, Hong Kong SAR
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

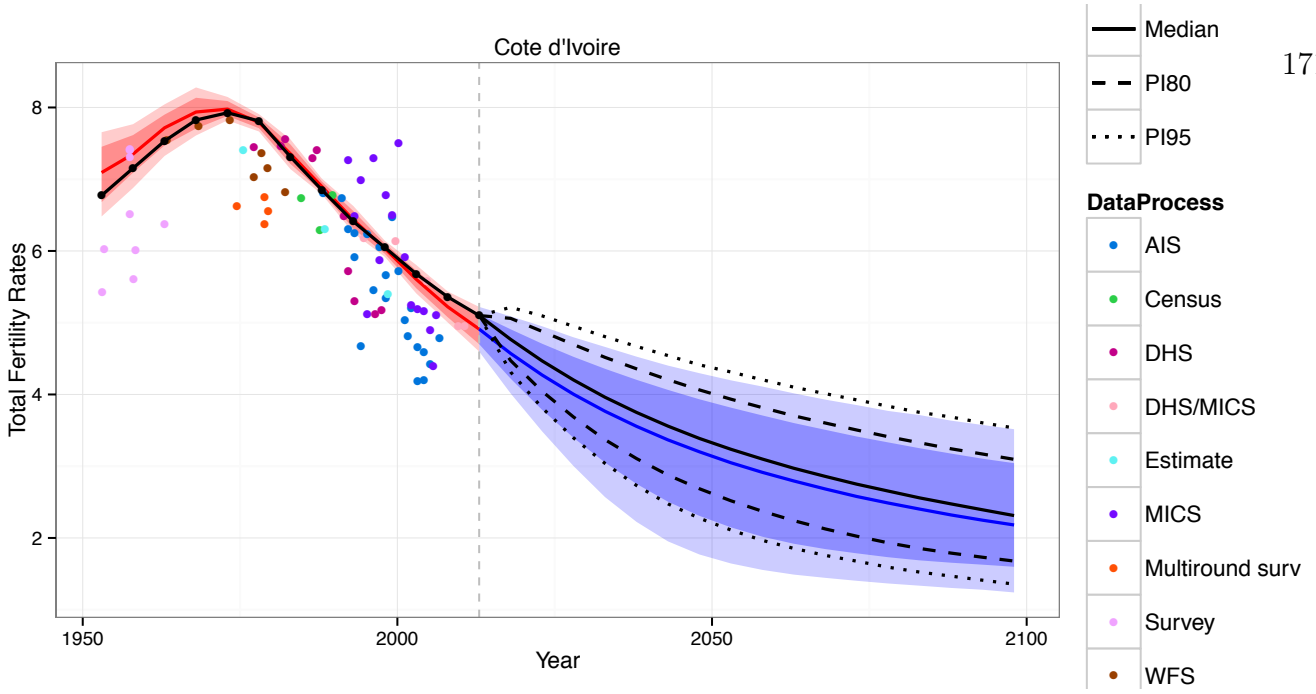

China, Macao SAR
Total Fertility Rates
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
VR

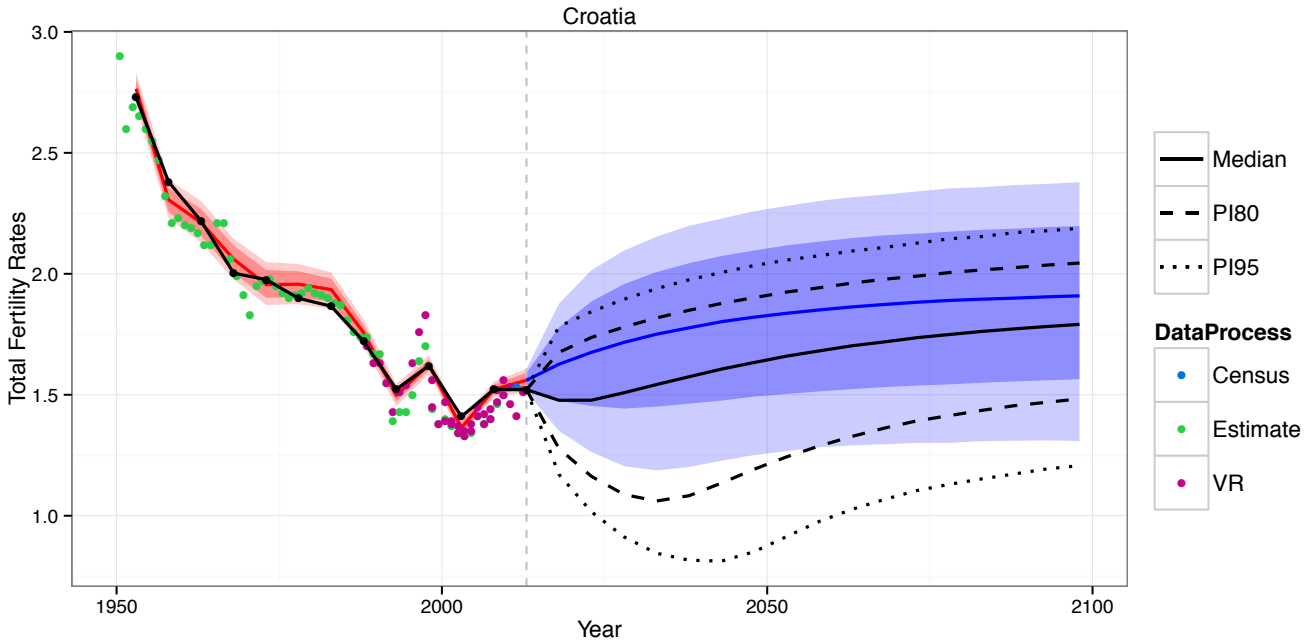

Colombia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
VR

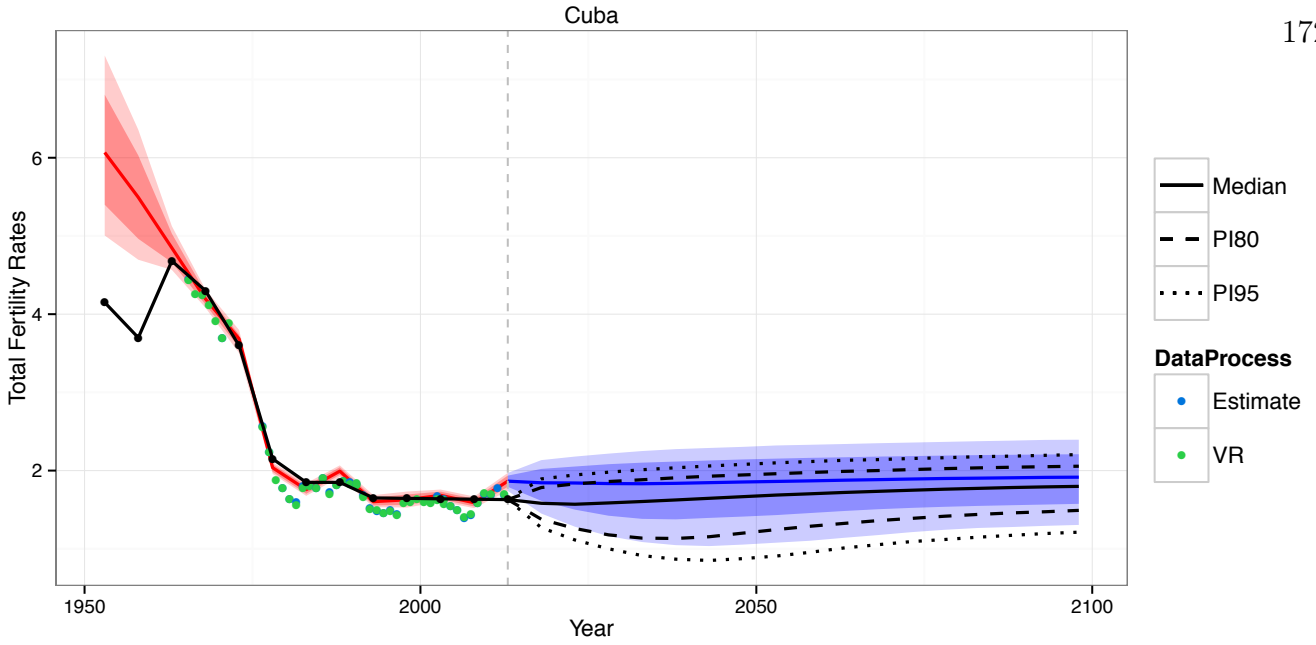

Comoros
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
DHS/MICS

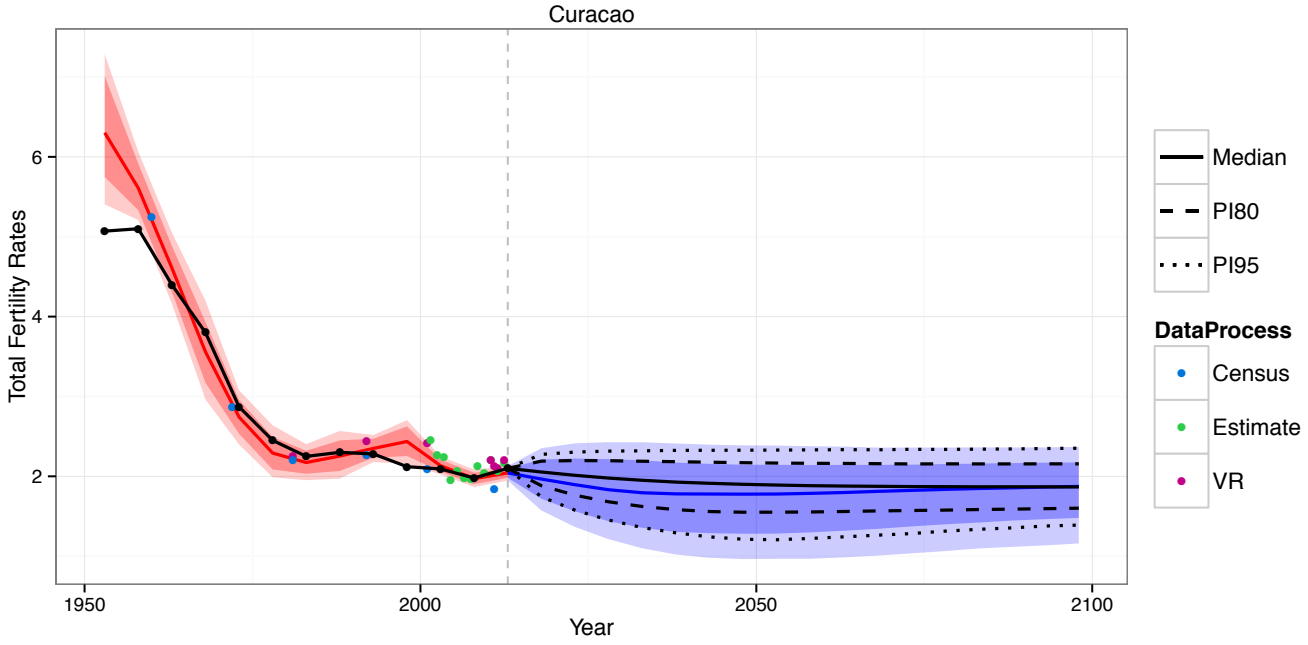

Congo
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Survey
Median
PI80
PI95

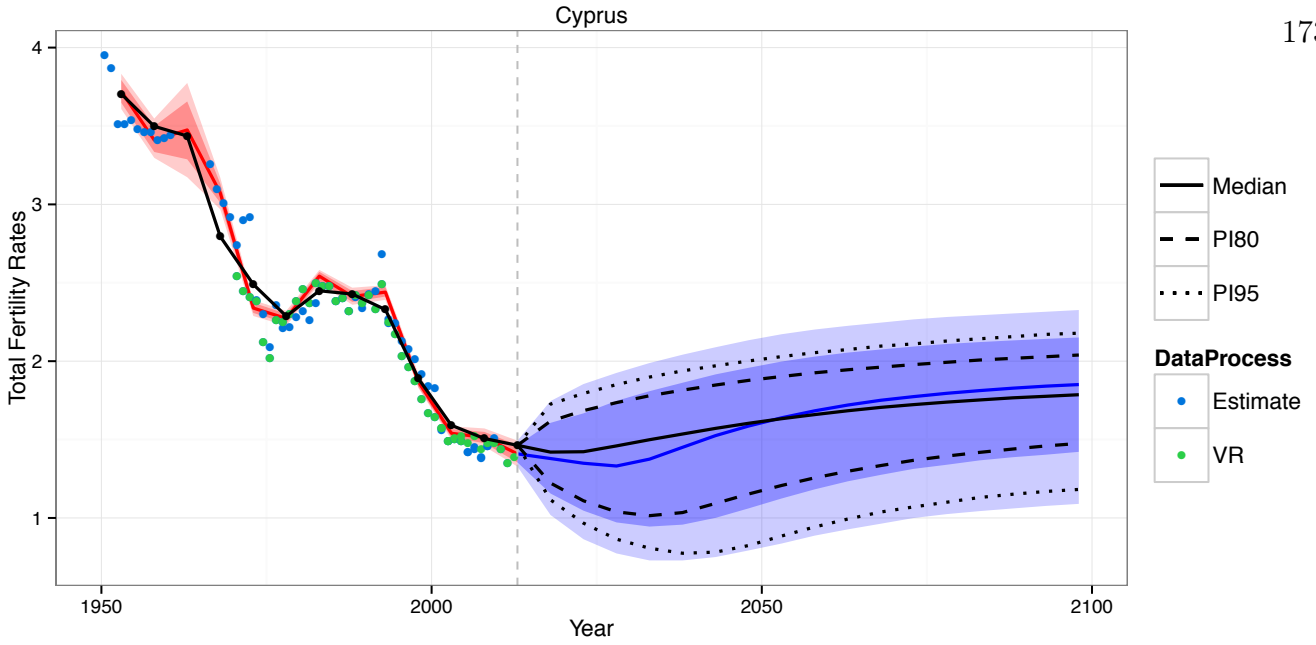

Costa Rica
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
RHS
VR
WFS

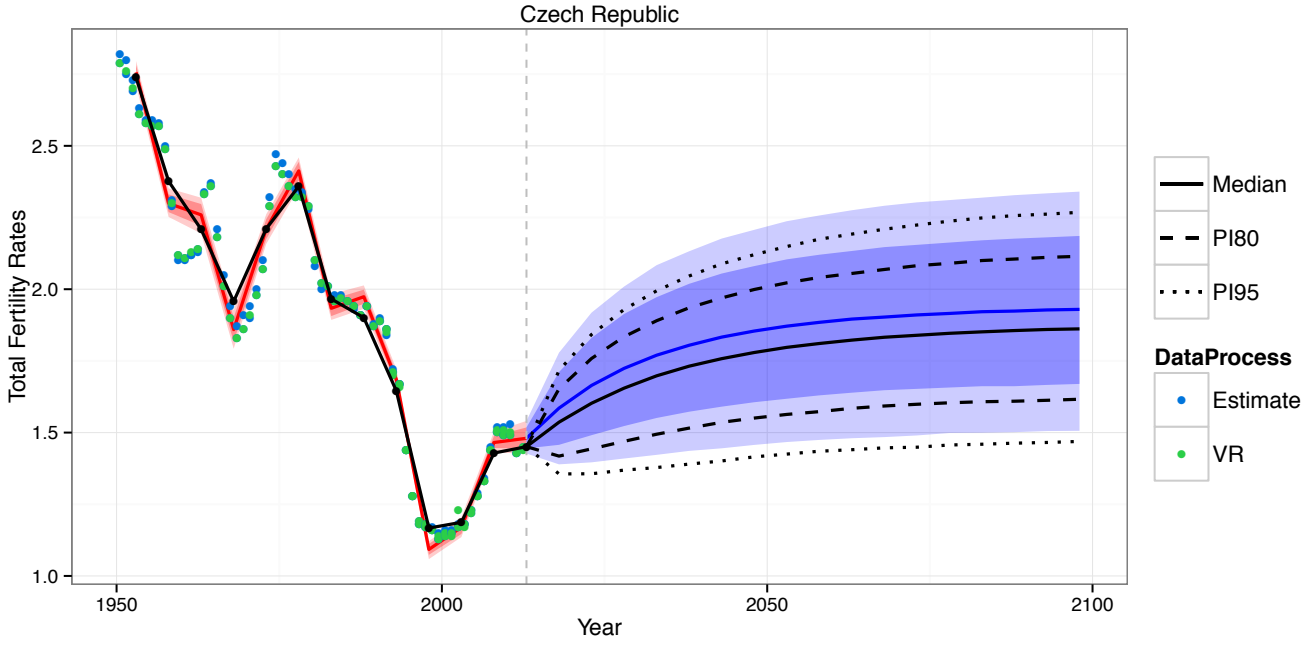

Cote d'Ivoire
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
AIS
Census
DHS
DHS/MICS
Estimate
MICS
Multiround surv
Survey
WFS

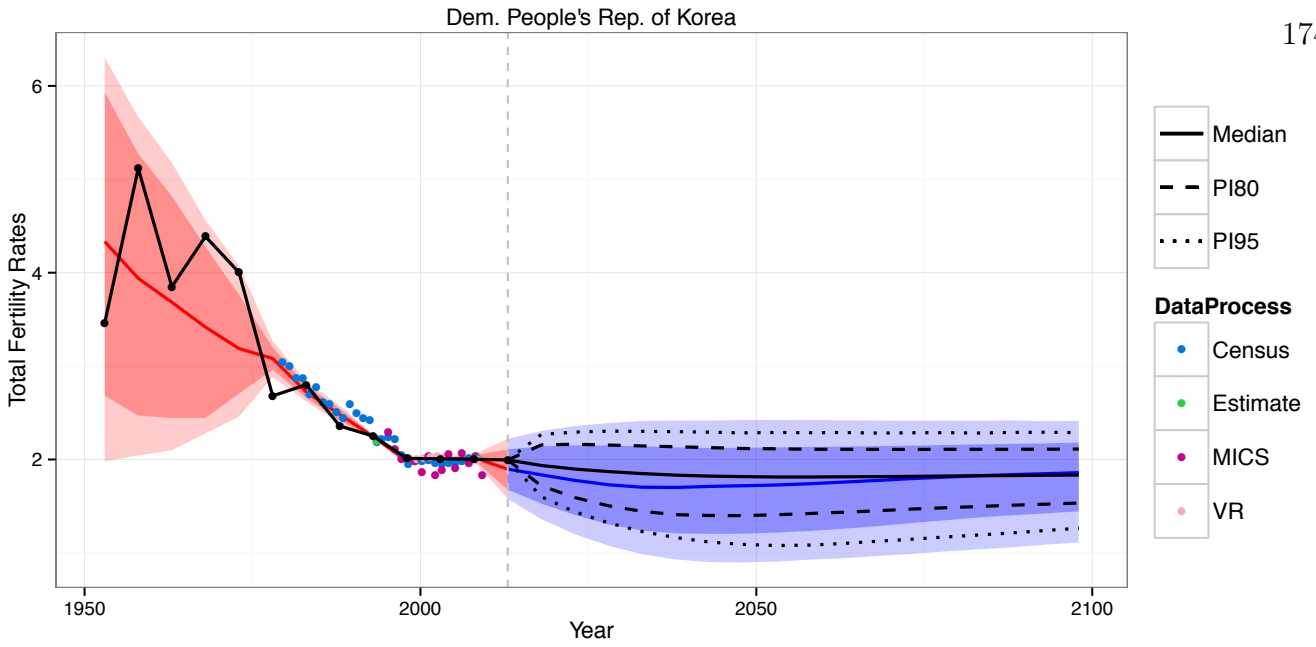

Croatia
Total Fertility Rates
3.0
2.5
2.0
1.5
1.0
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

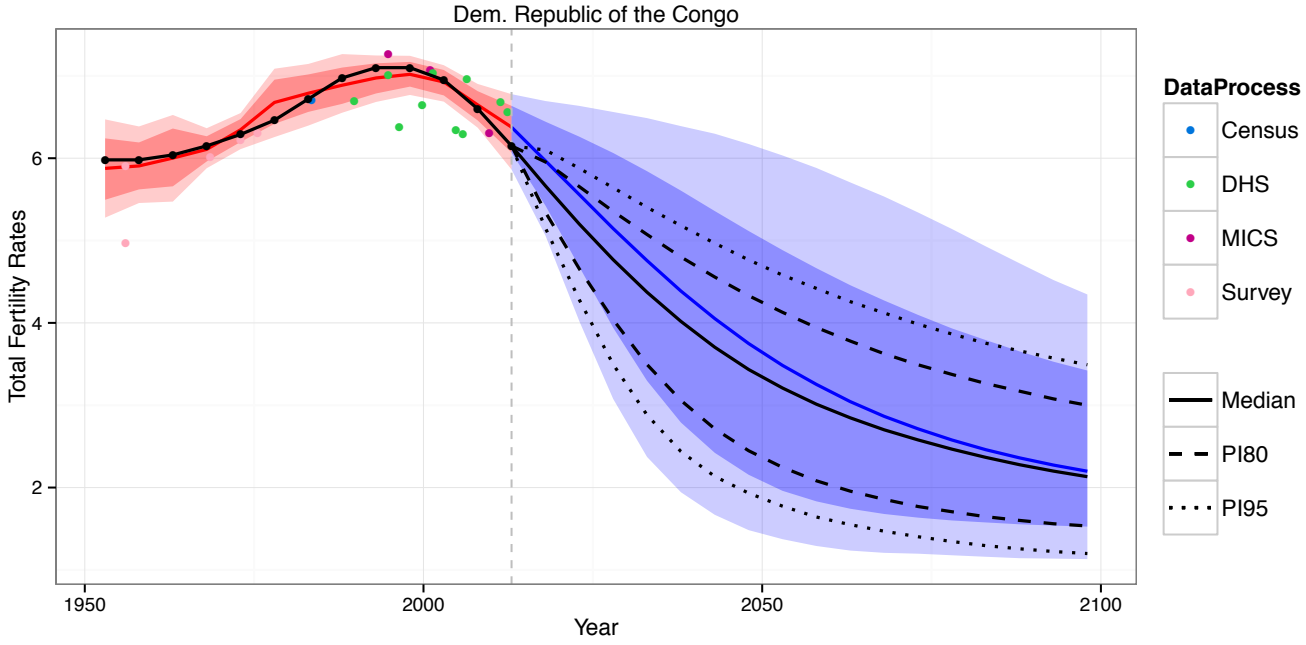

Cuba
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

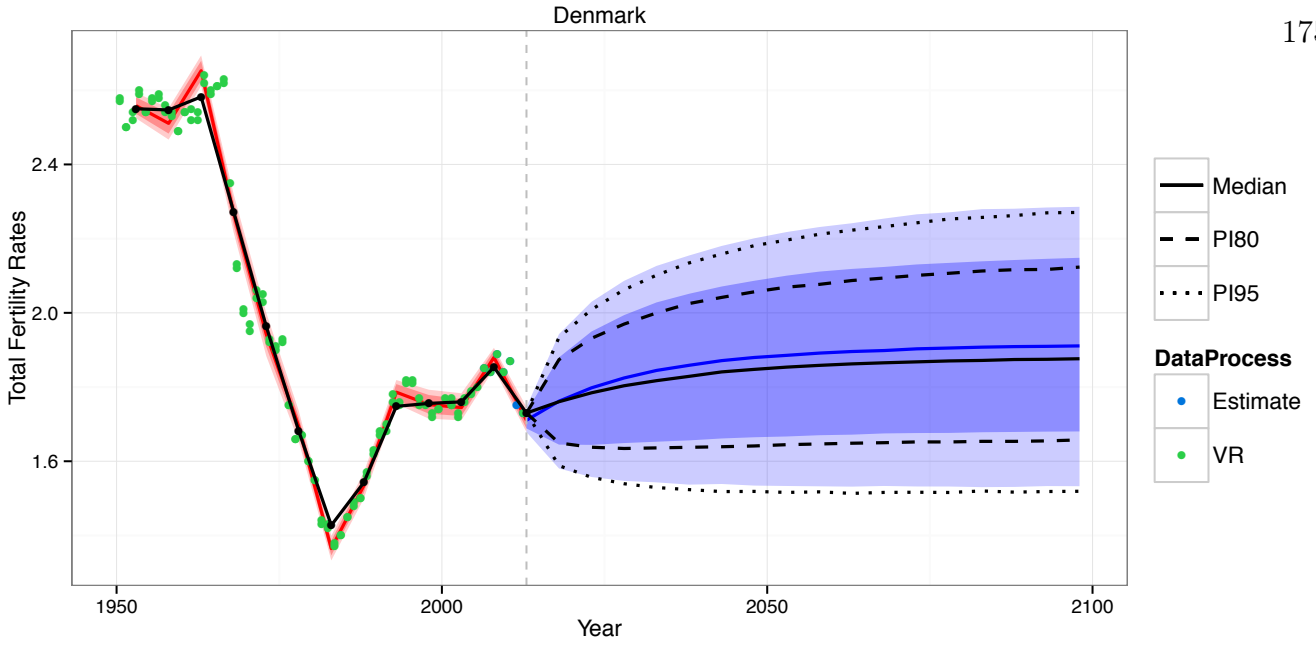

Curacao
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

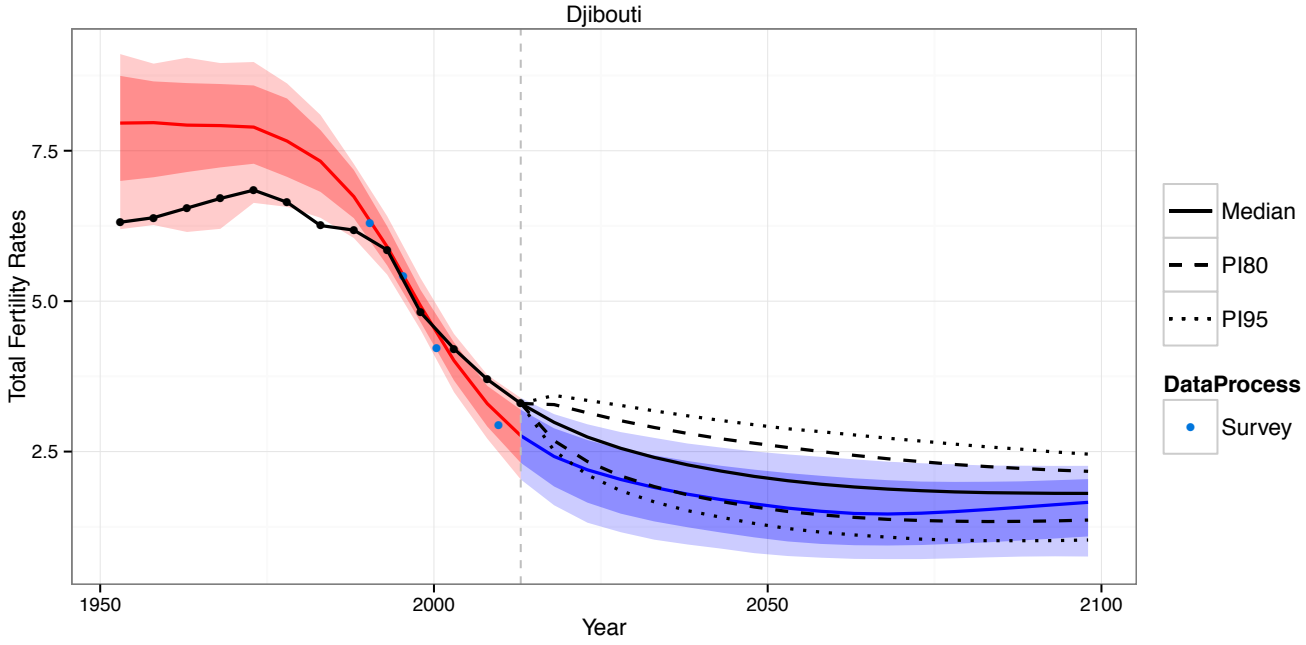

Cyprus
Total Fertility Rates
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

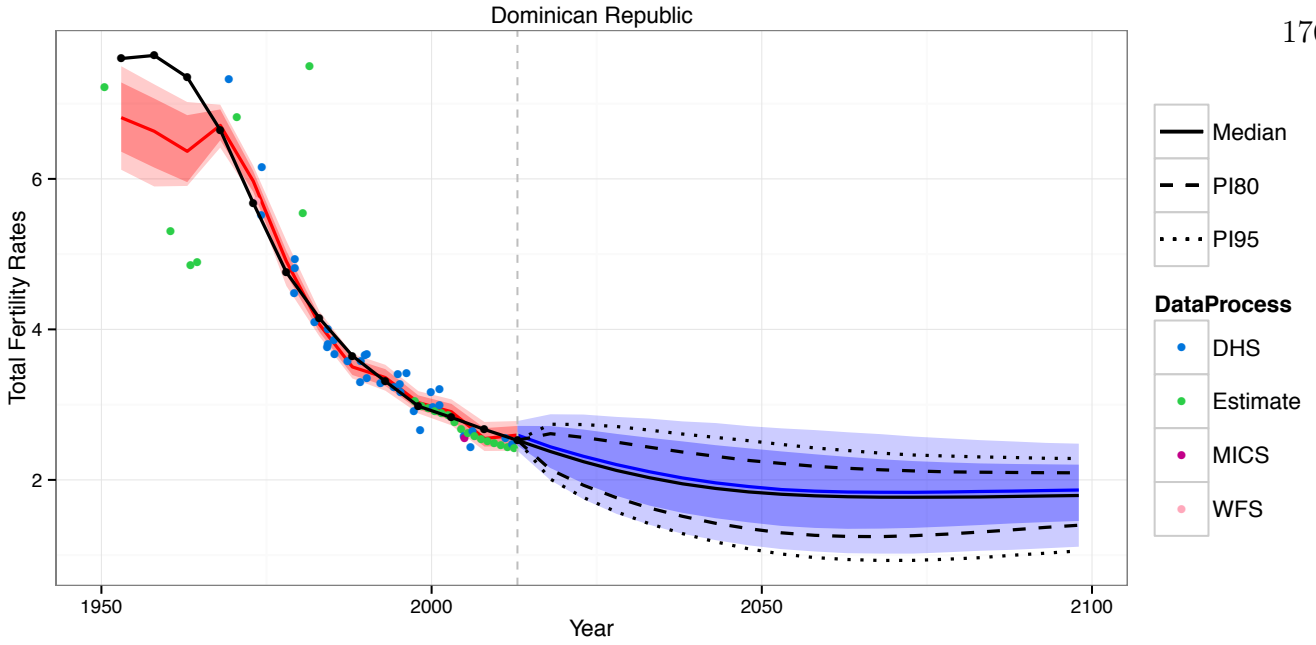

Czech Republic
Total Fertility Rates
2.5
2.0
1.5
1.0
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

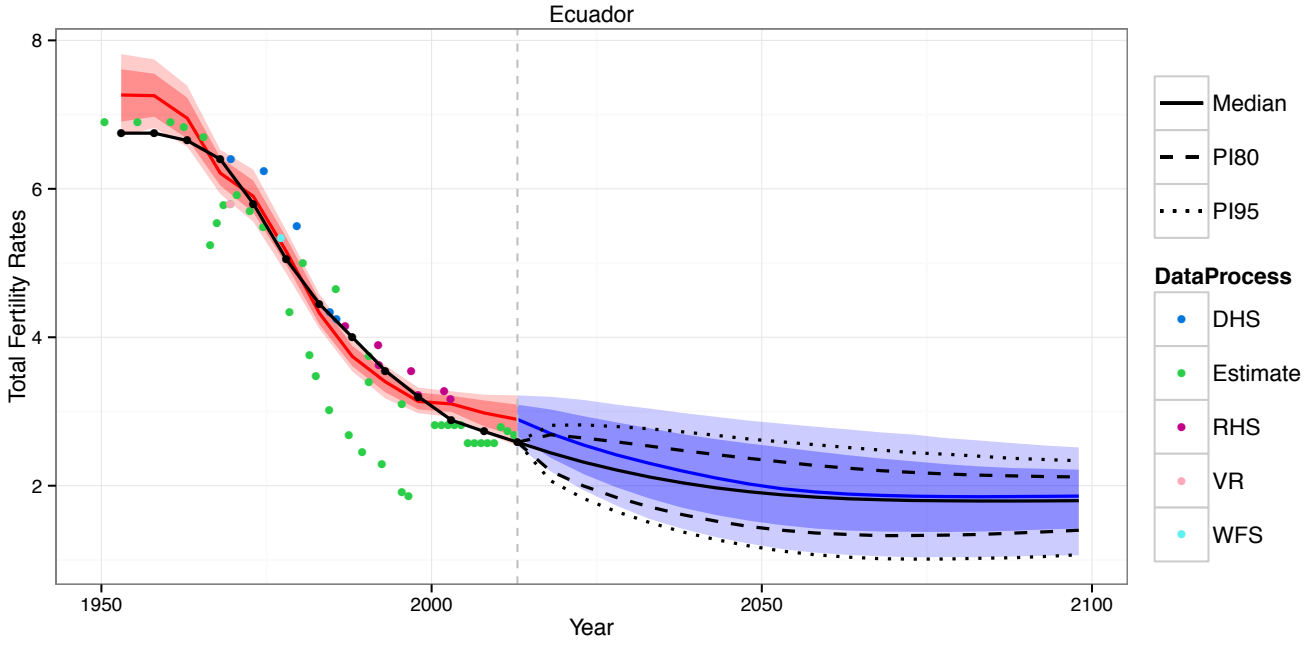

Dem. People's Rep. of Korea
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
MICS
VR

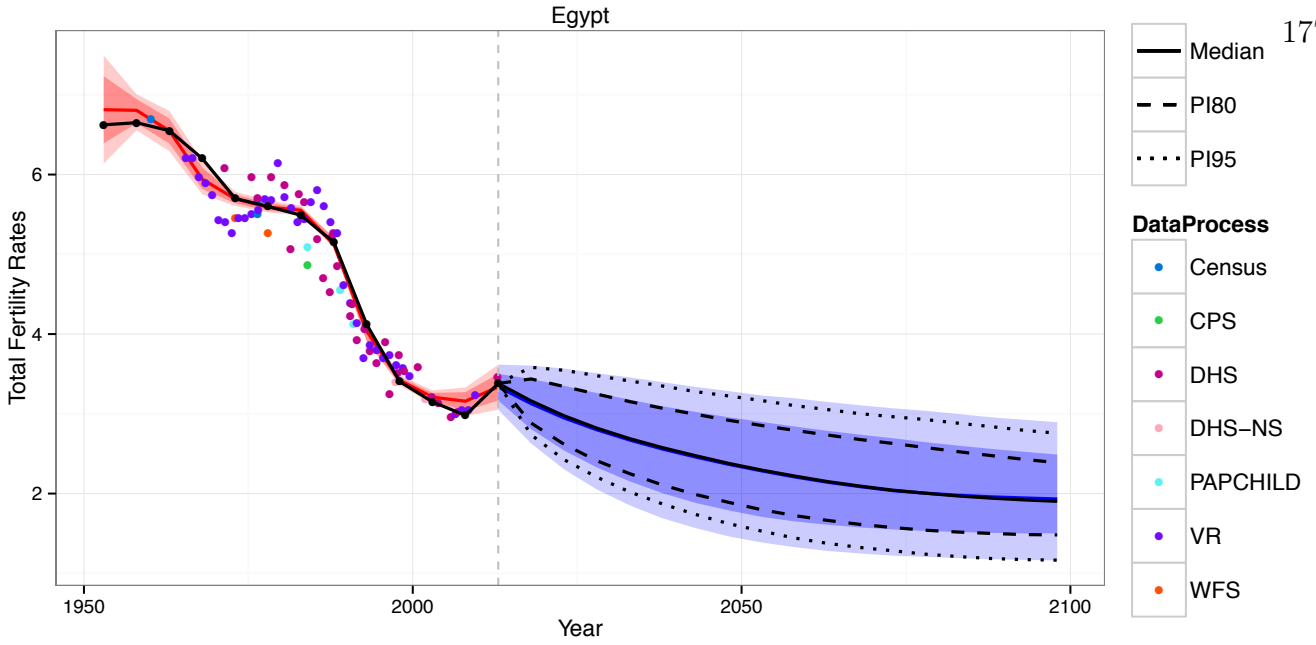

Dem. Republic of the Congo
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
DHS
MICS
Survey
Median
PI80
PI95

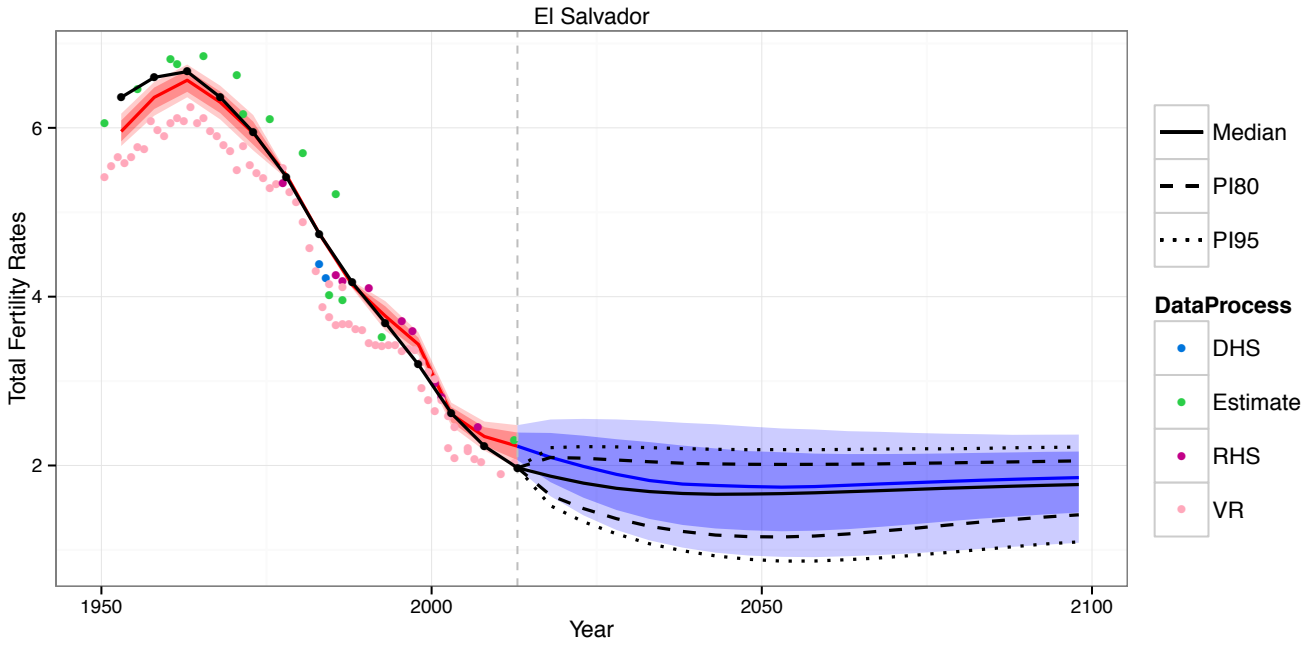

Denmark
Total Fertility Rates
2.4
2.0
1.6
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

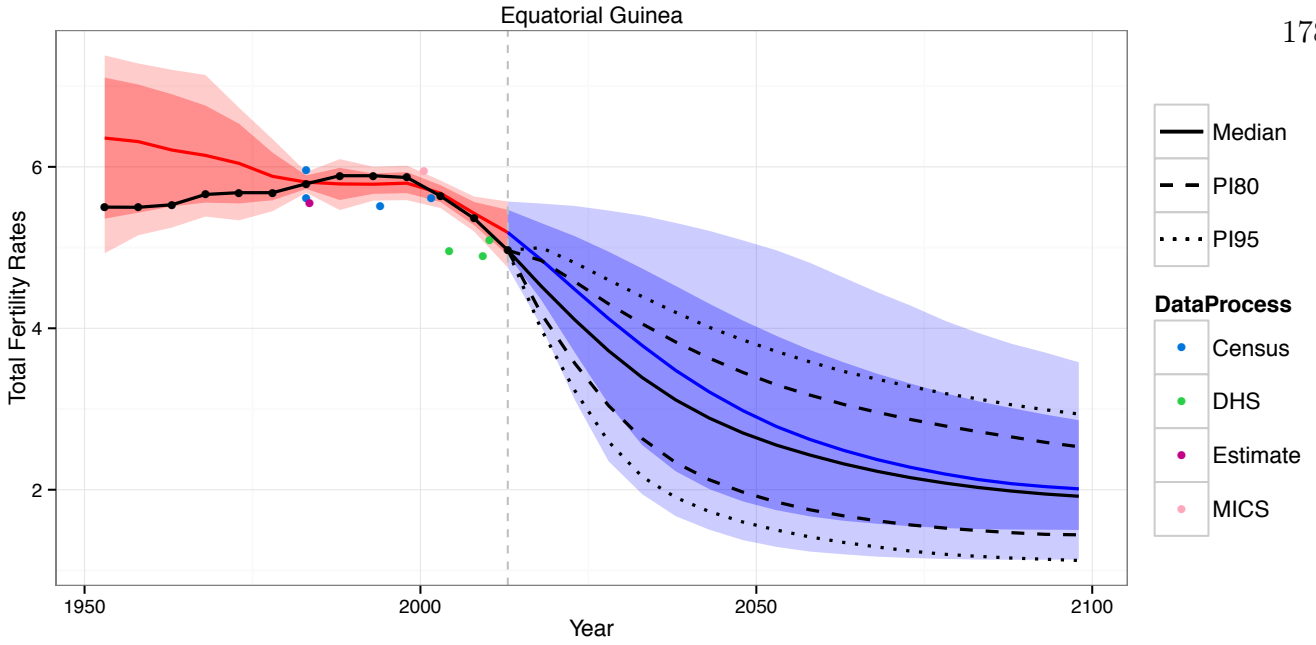

Djibouti
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Survey

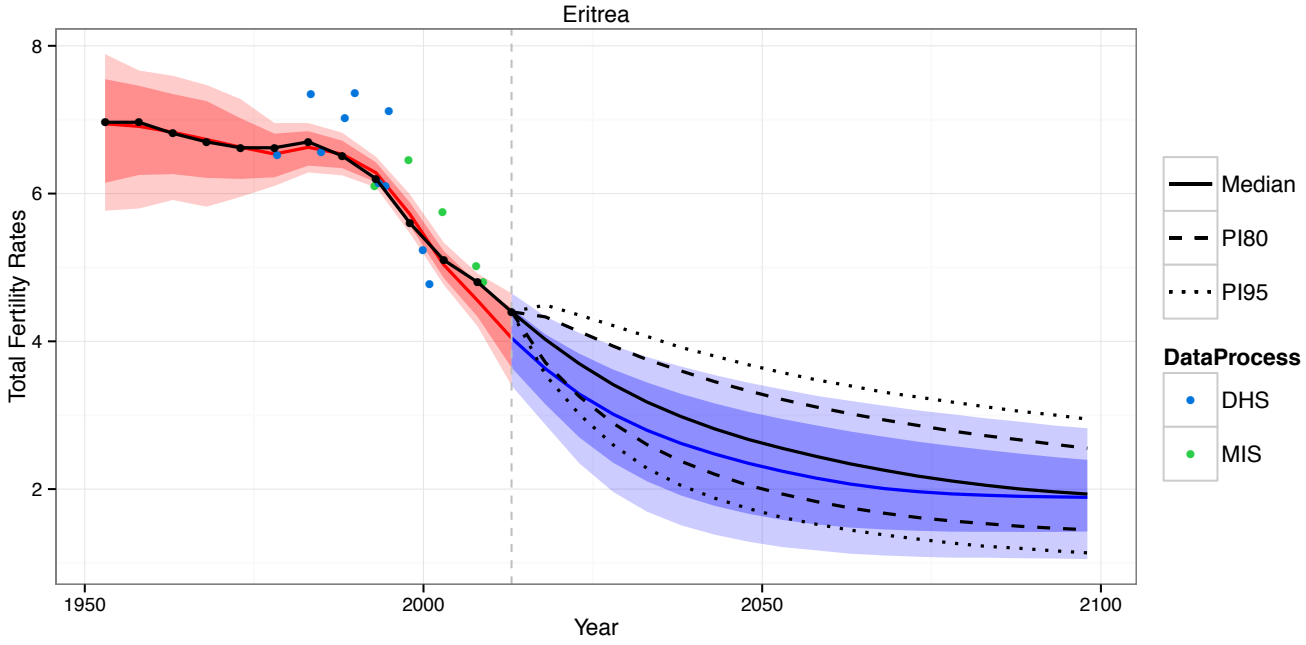

Dominican Republic
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
MICS
WFS

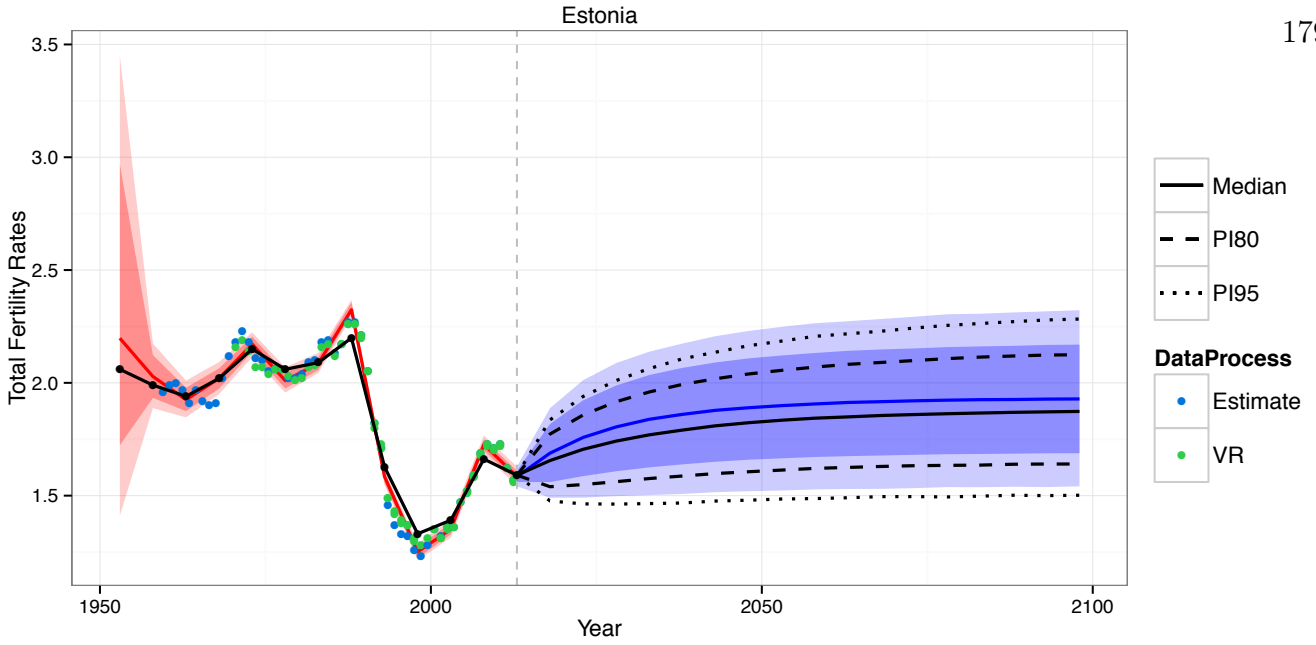

Ecuador
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
RHS
VR
WFS

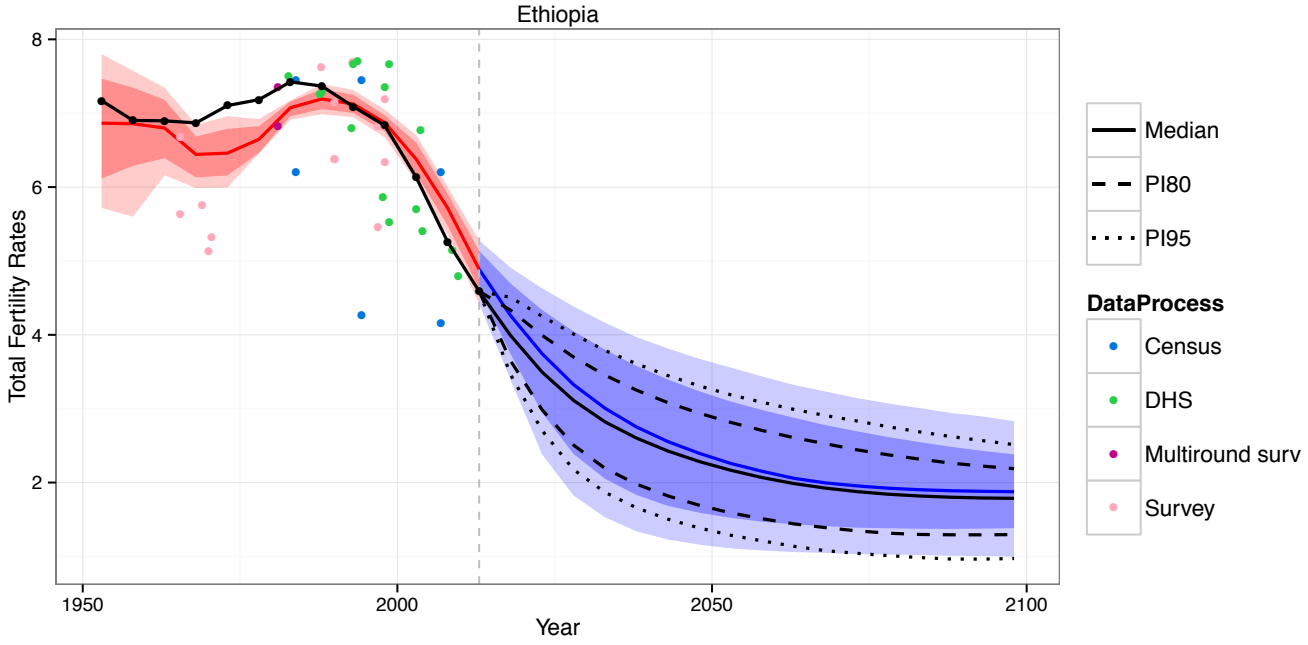

Egypt
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
DHS
DHS−NS
PAPCHILD
VR
WFS

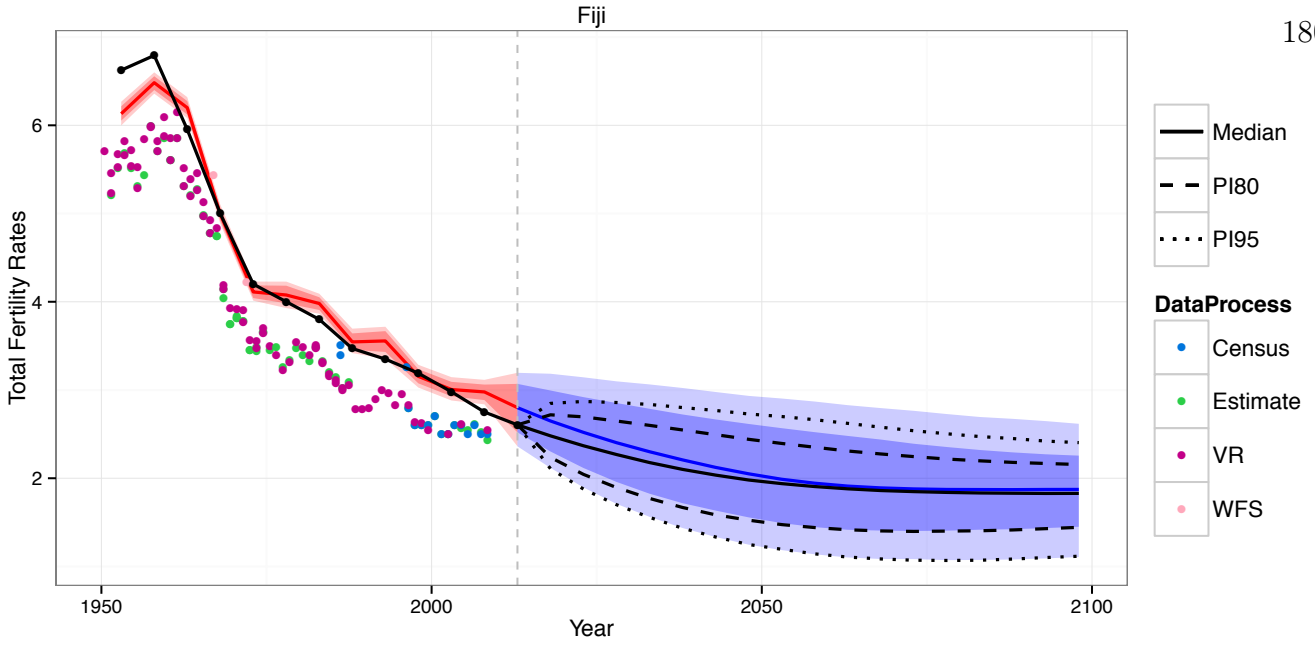

El Salvador
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
RHS
VR

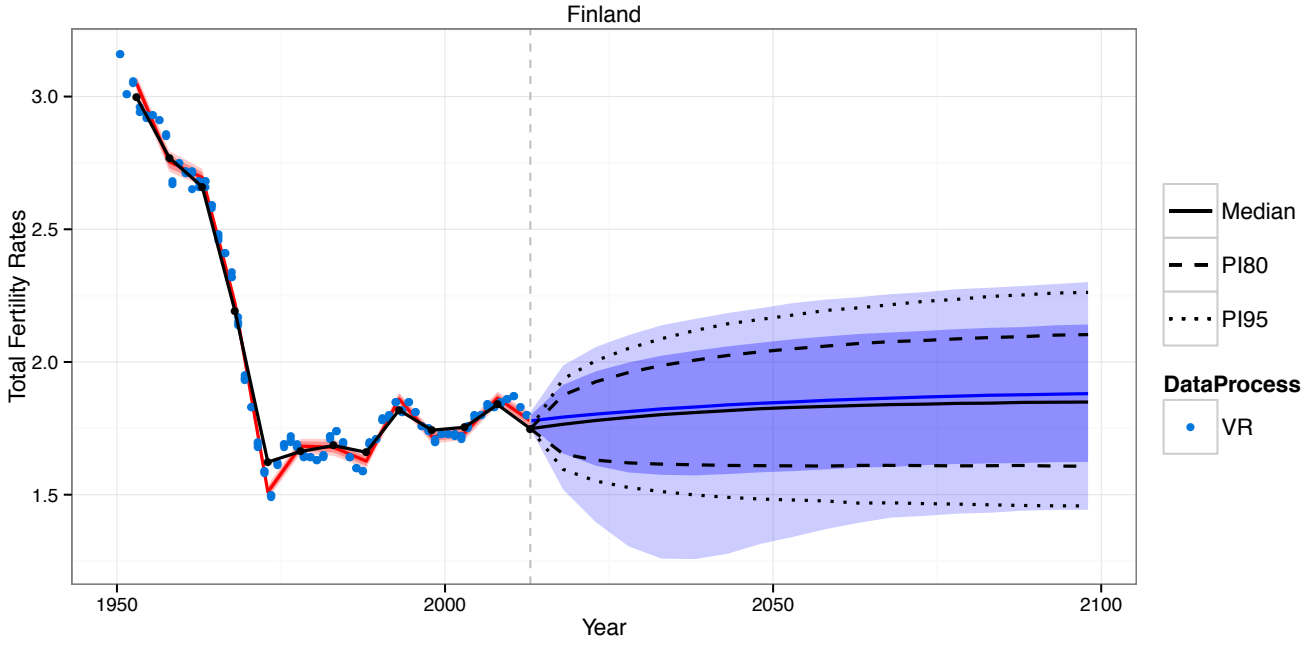

Equatorial Guinea
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS

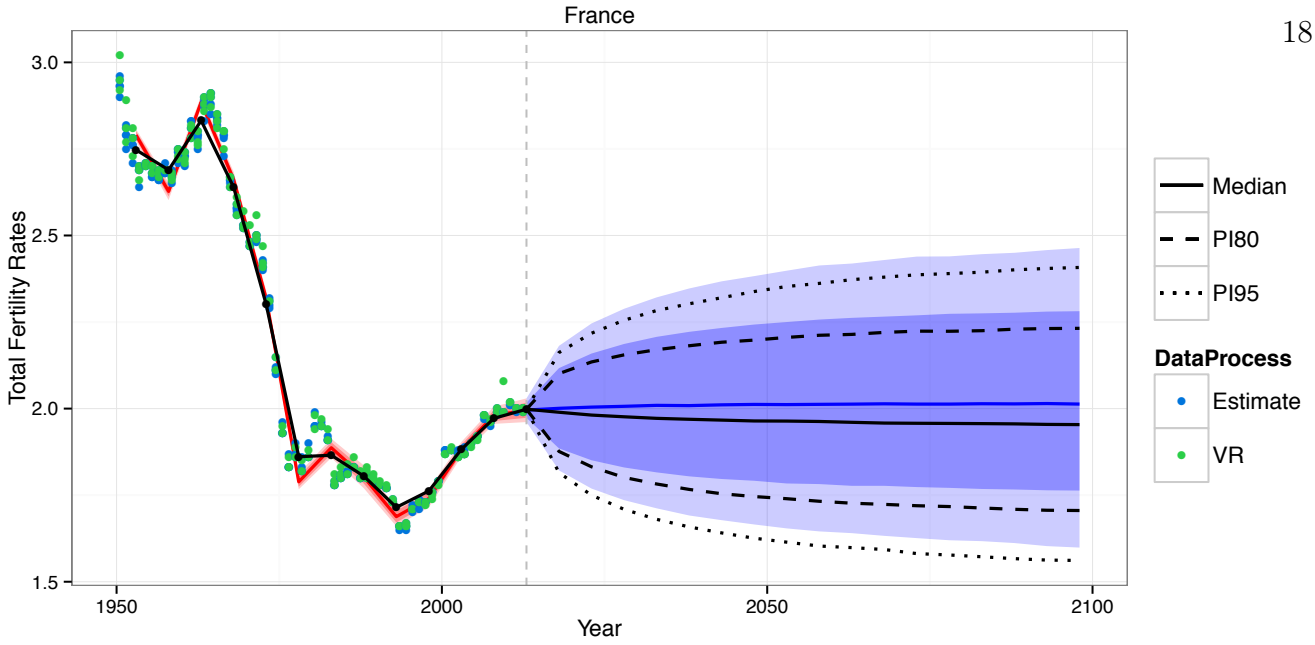

Eritrea
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
MIS

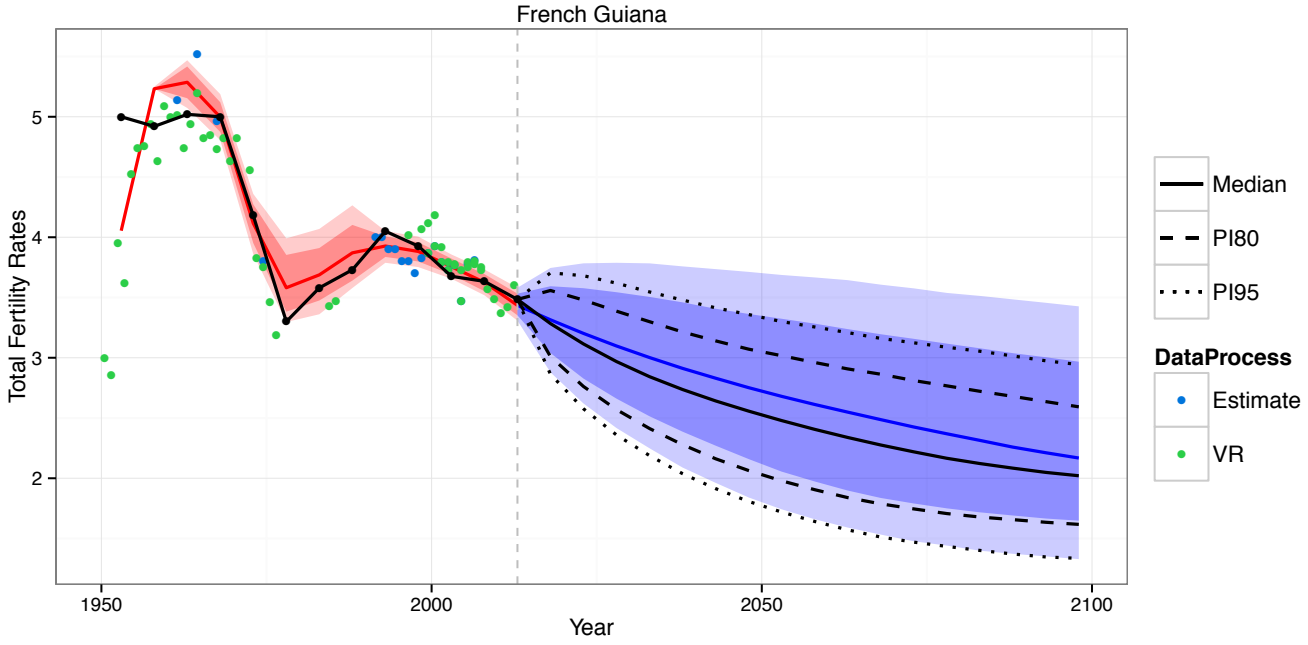

Estonia
Total Fertility Rates
3.5
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

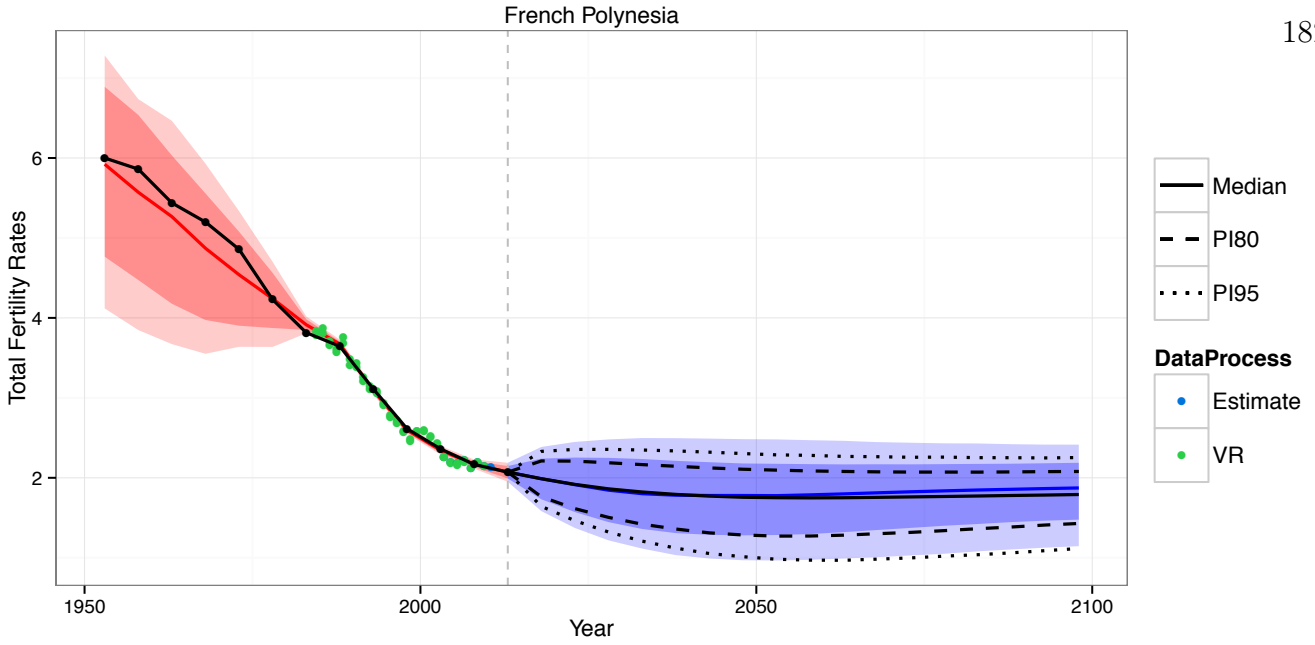

Ethiopia
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Multiround surv
Survey

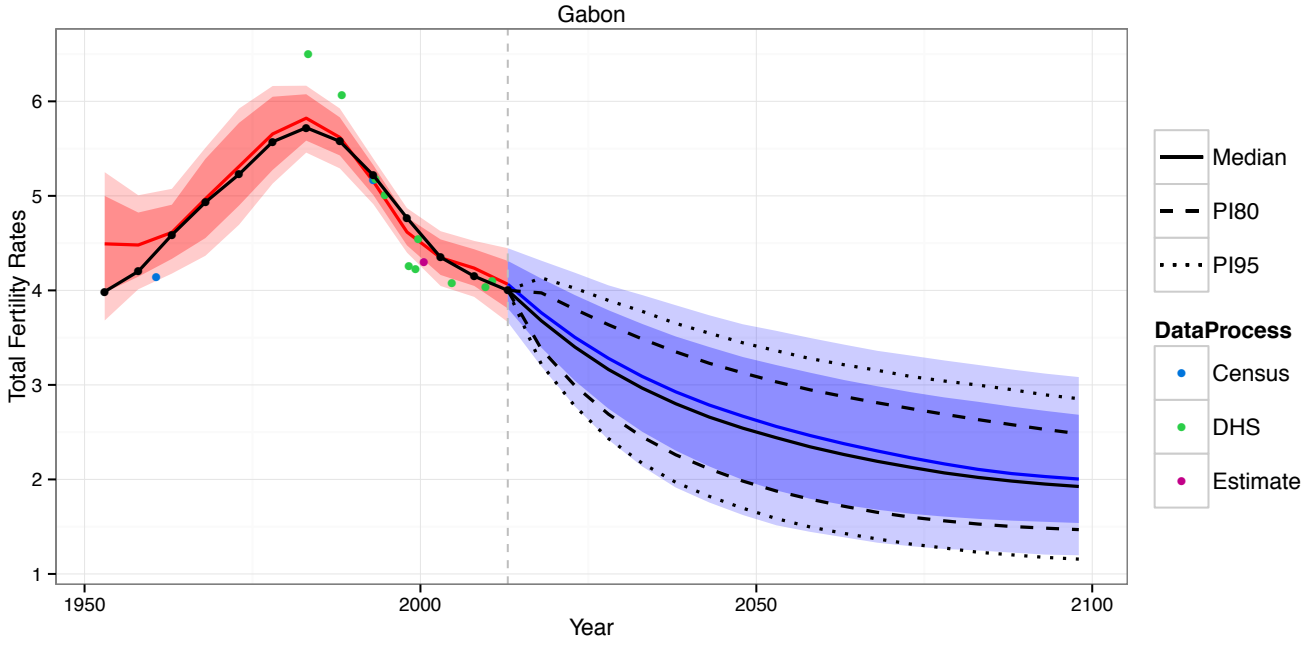

Fiji
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR
WFS

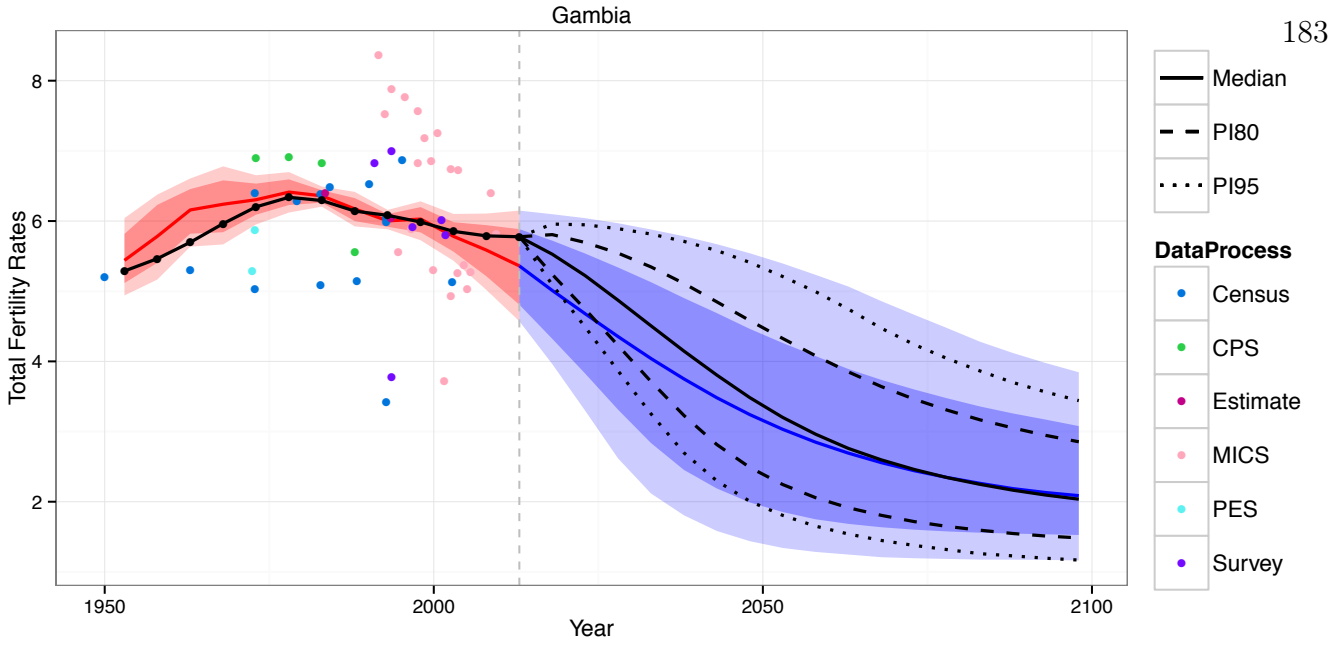

Finland
Total Fertility Rates
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
VR

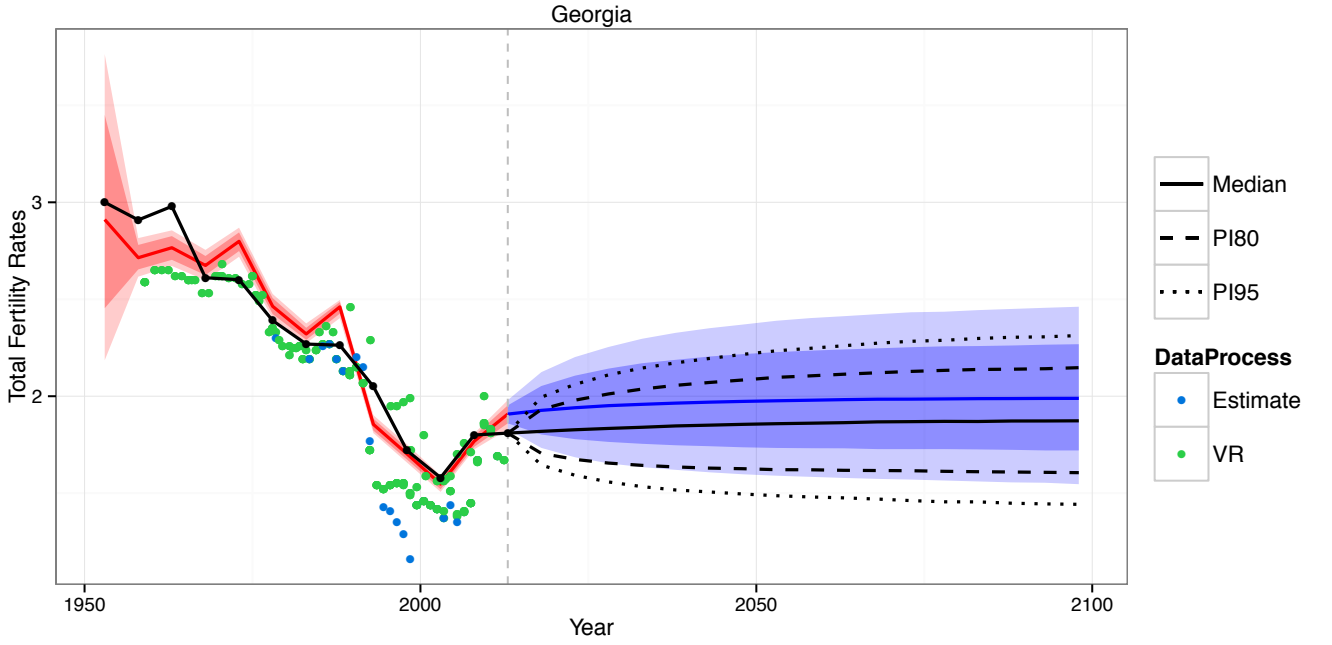

France
Total Fertility Rates
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

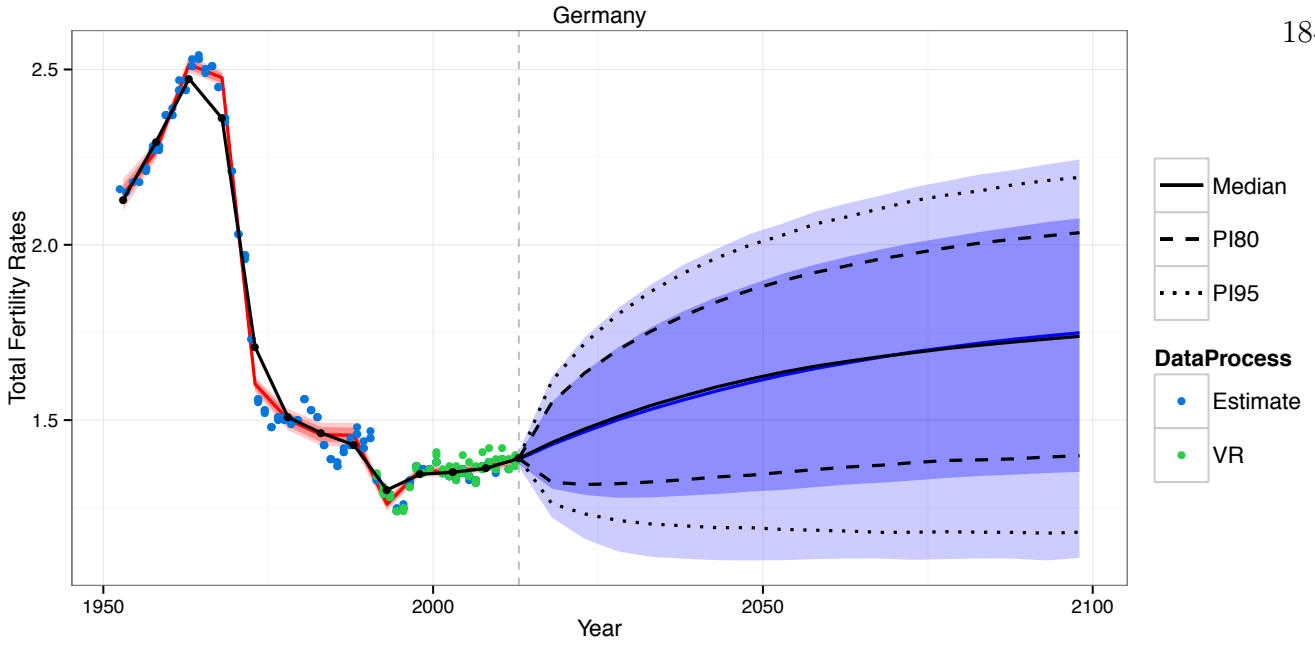

French Guiana
Total Fertility Rates
5
4
3
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

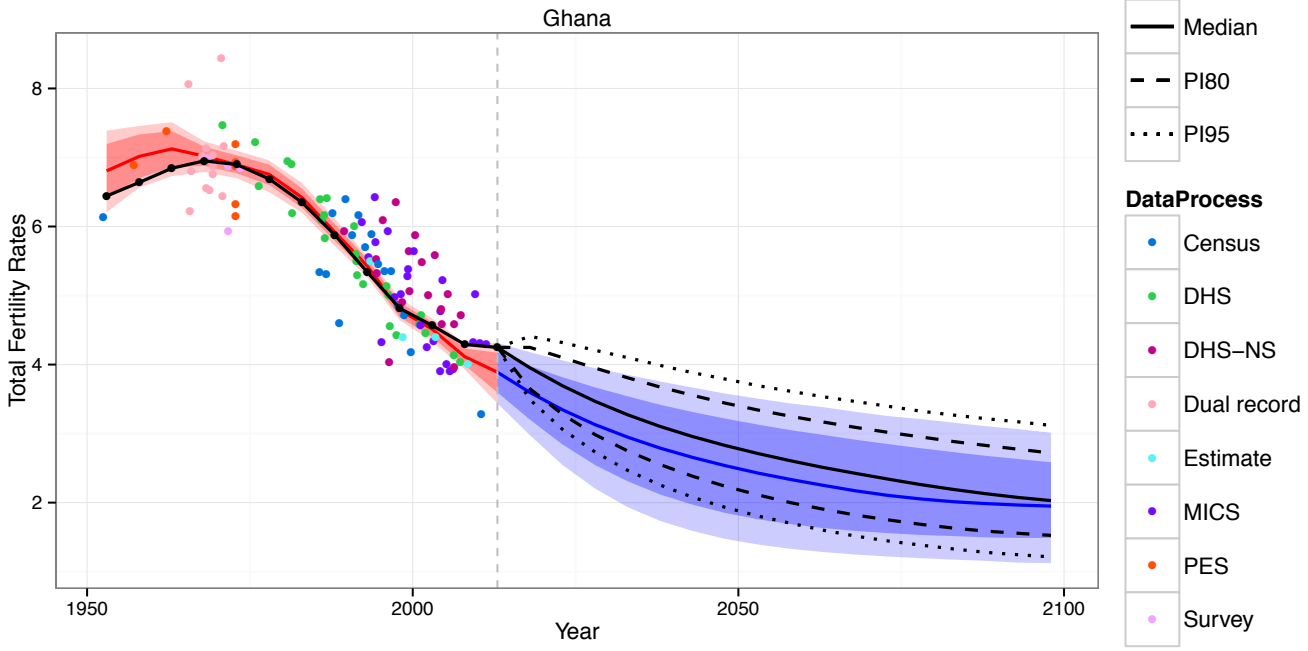

French Polynesia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

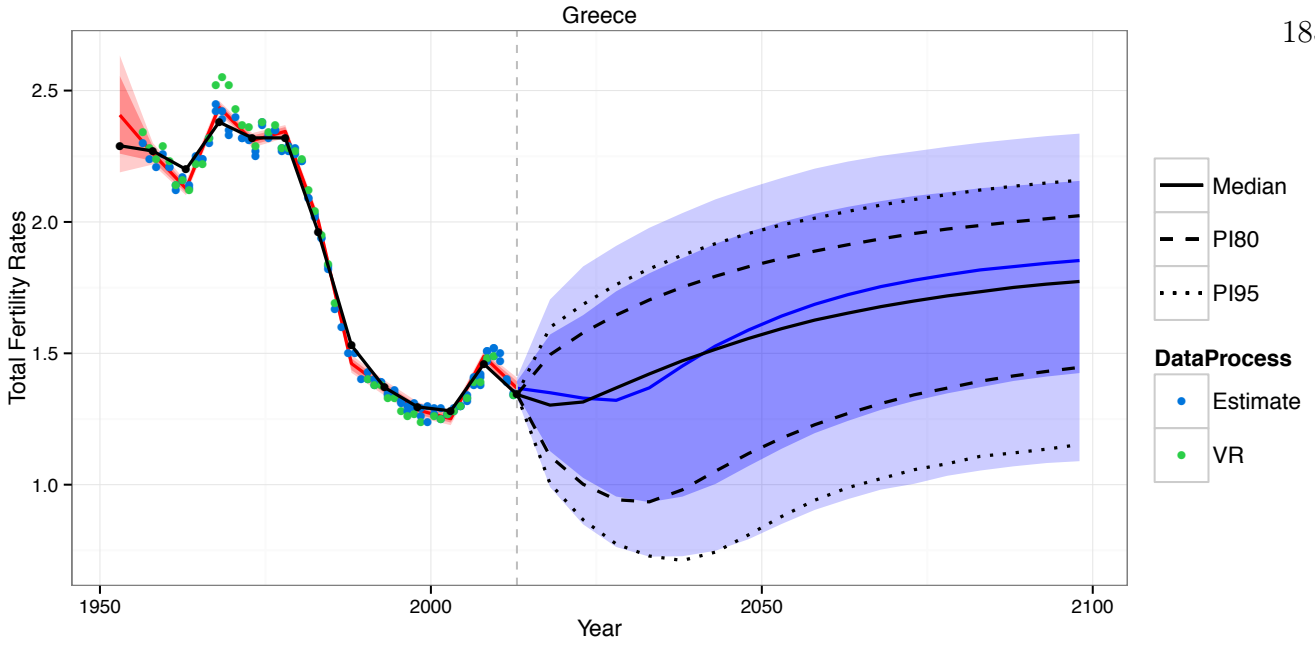

Gabon
Total Fertility Rates
6
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate

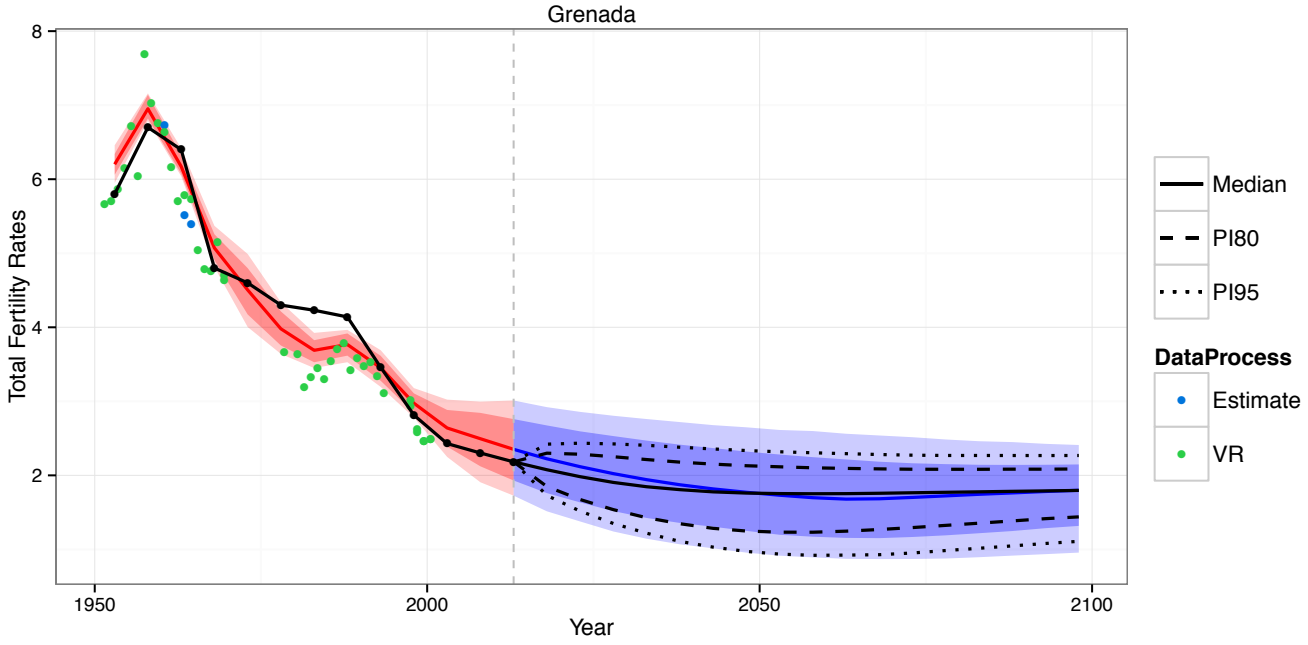

Gambia
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
Estimate
MICS
PES
Survey

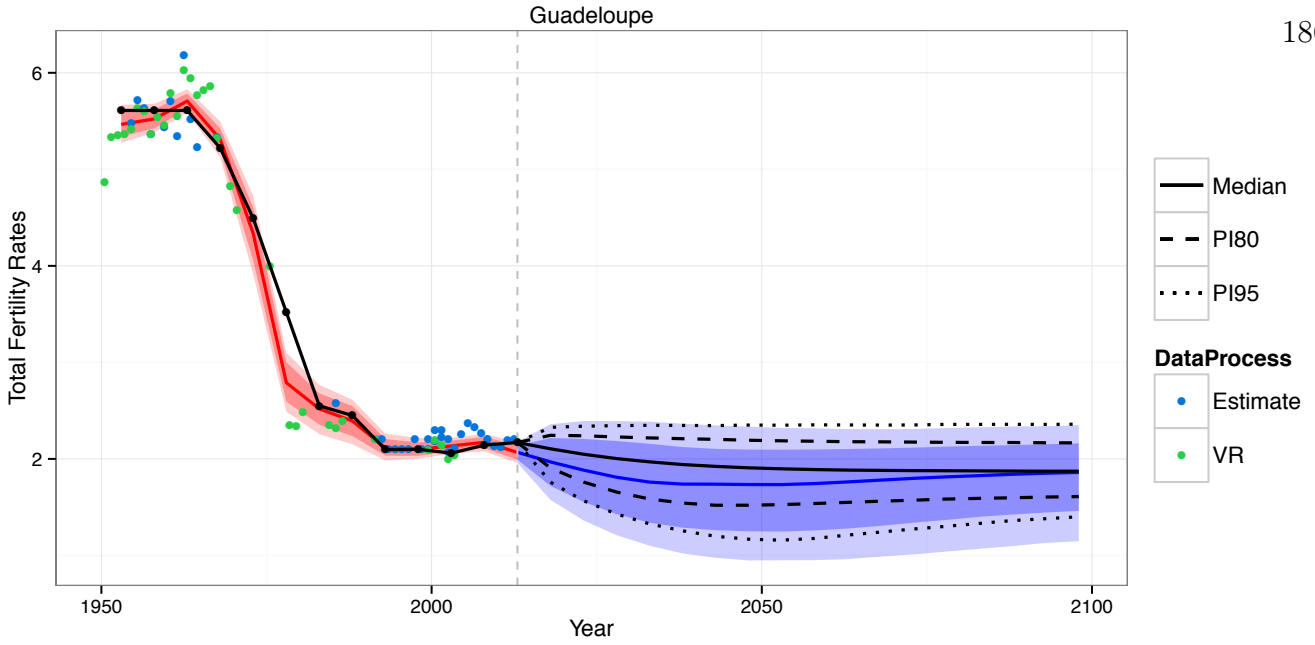

Georgia
Total Fertility Rates
3
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

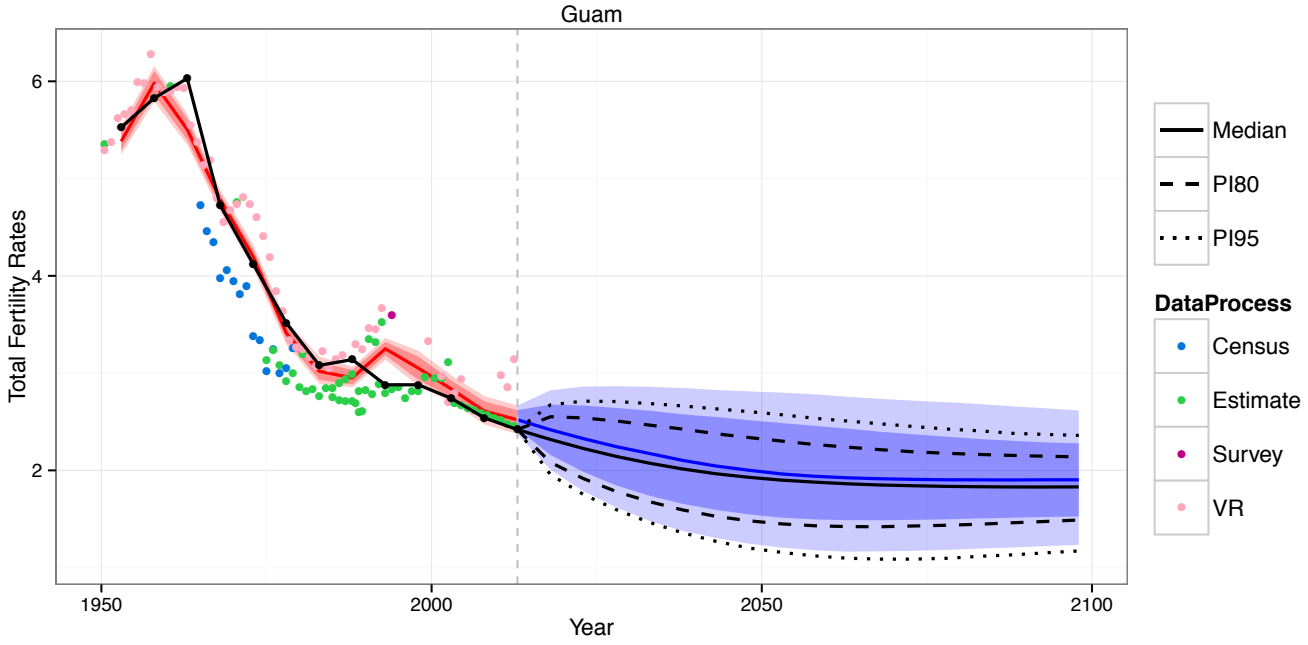

Germany
Total Fertility Rates
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

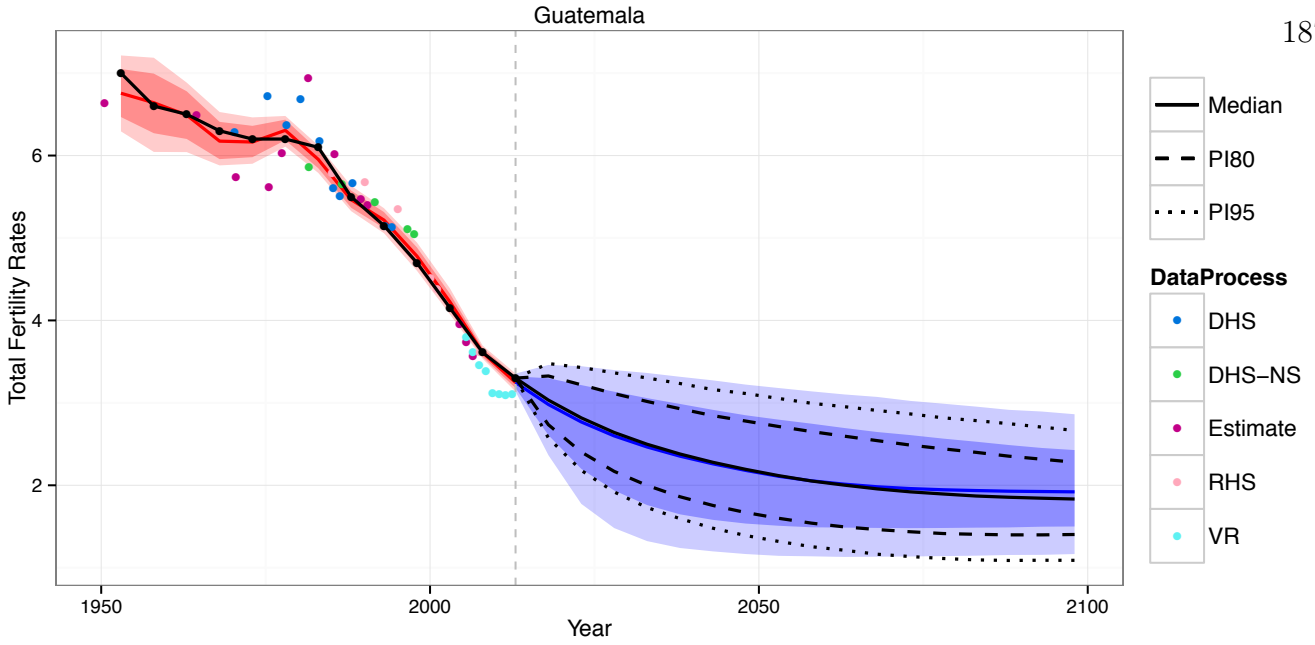

Ghana
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
DHS–NS
Dual record
Estimate
MICS
PES
Survey

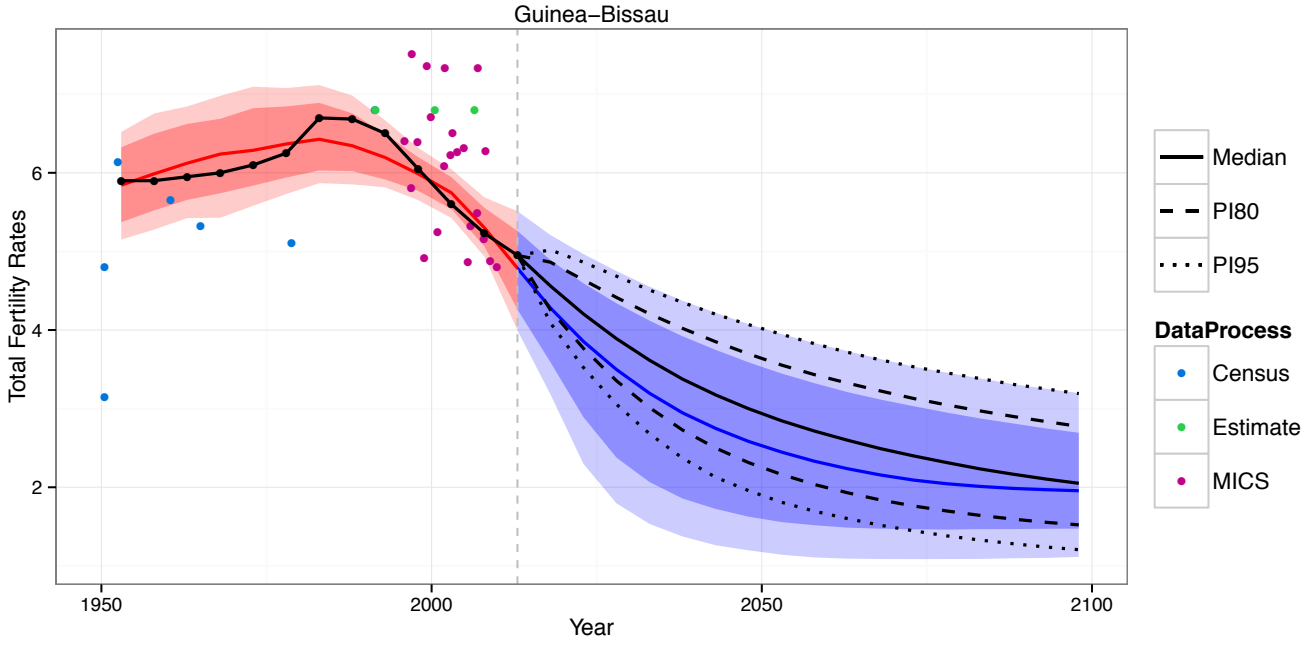

Greece
Total Fertility Rates
2.5
2.0
1.5
1.0
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

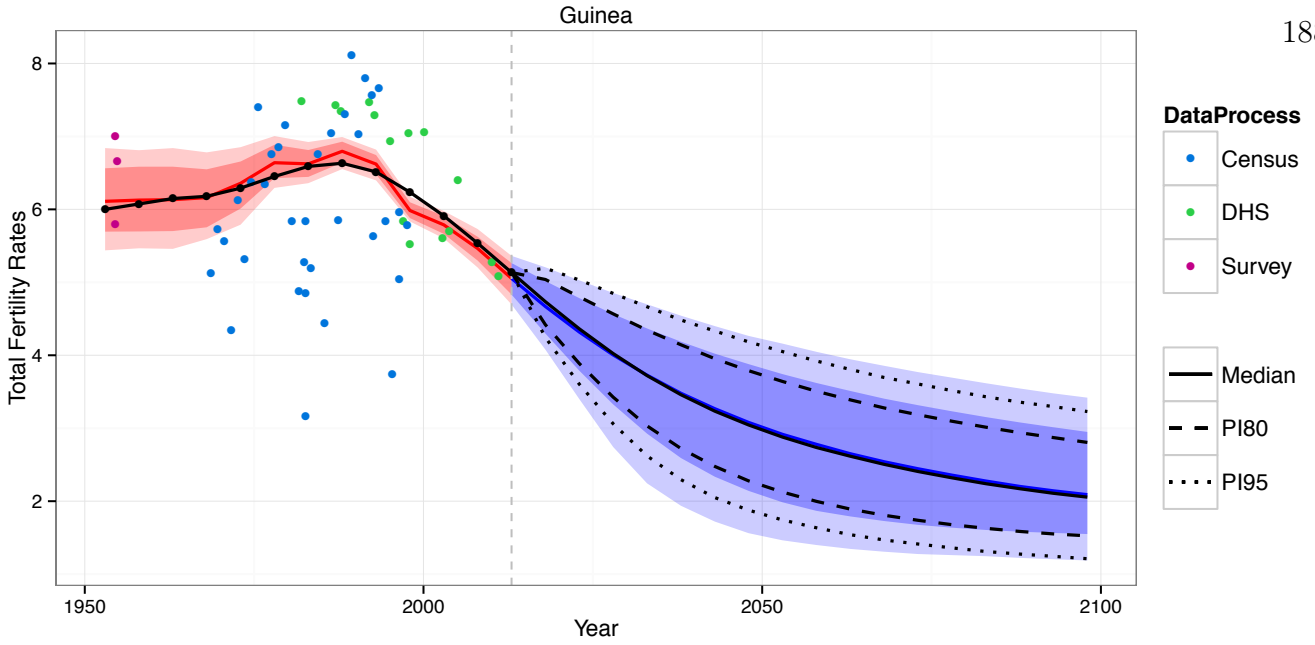

Grenada
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

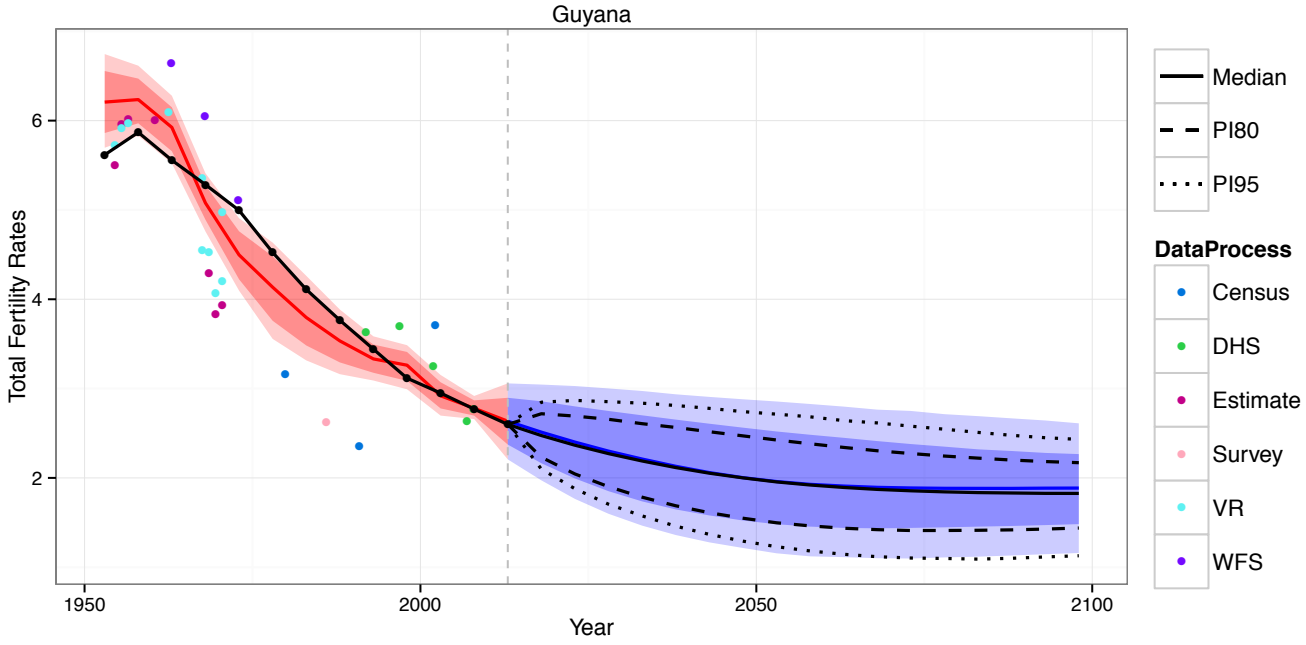

Guadeloupe
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

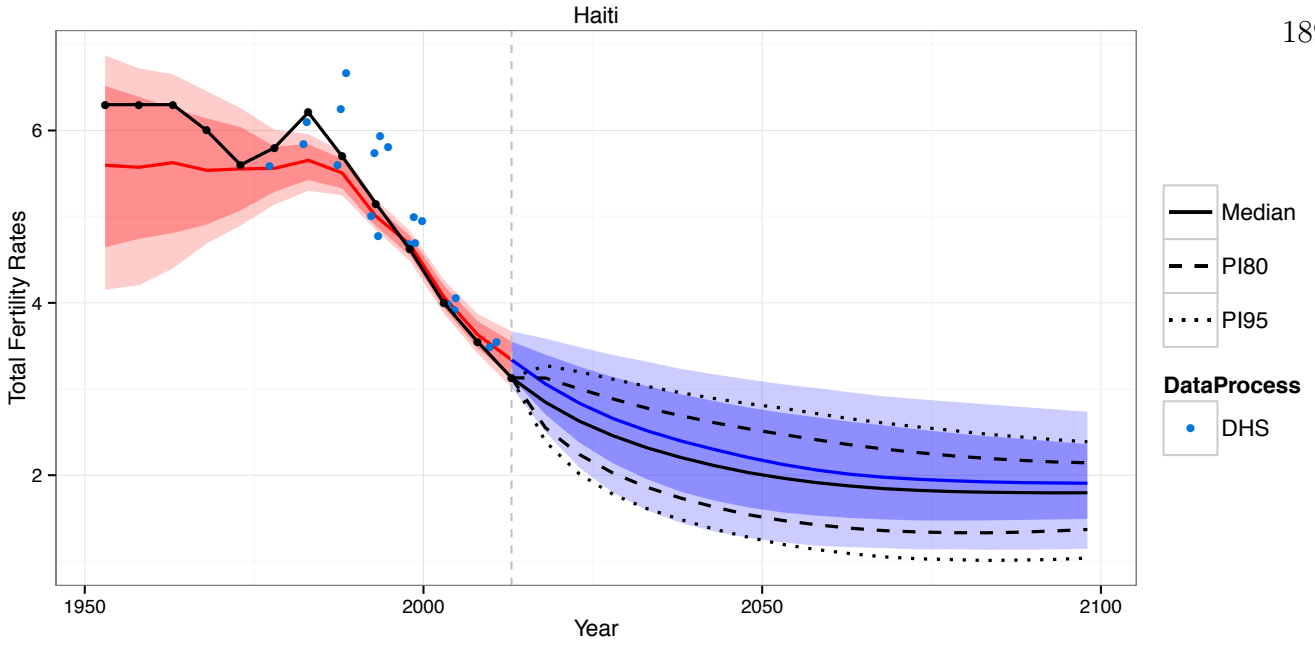

Guam
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
Survey
VR

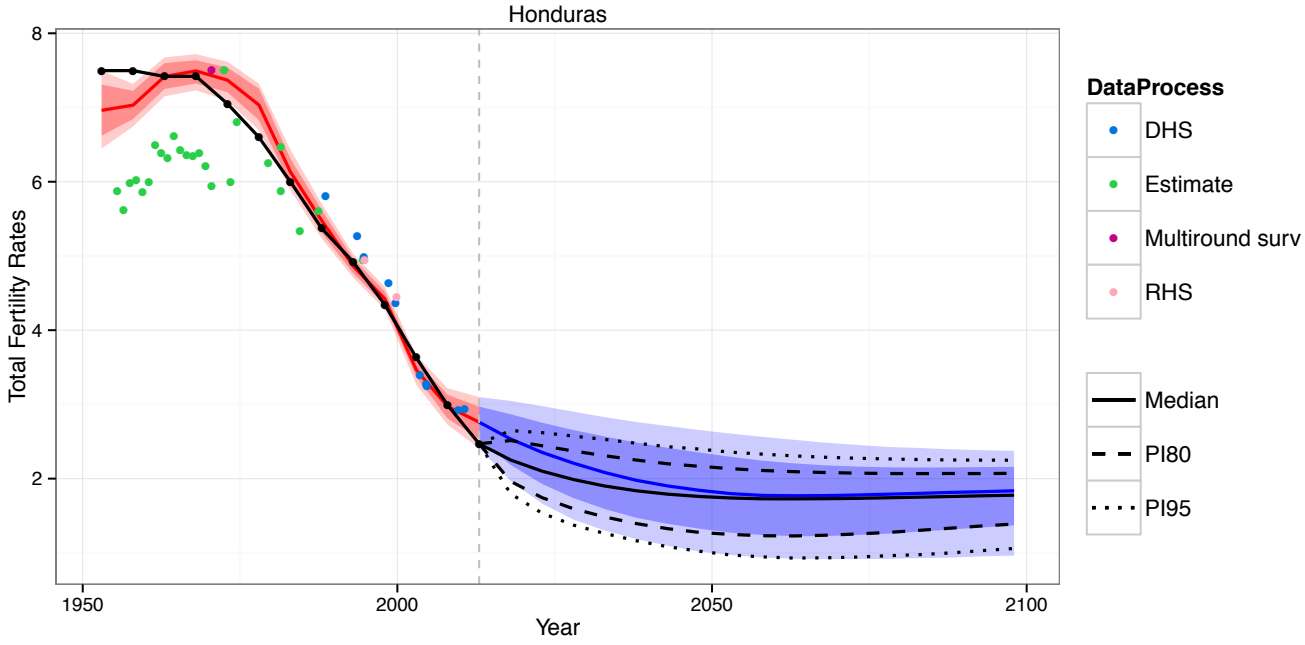

Guatemala
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
DHS−NS
Estimate
RHS
VR

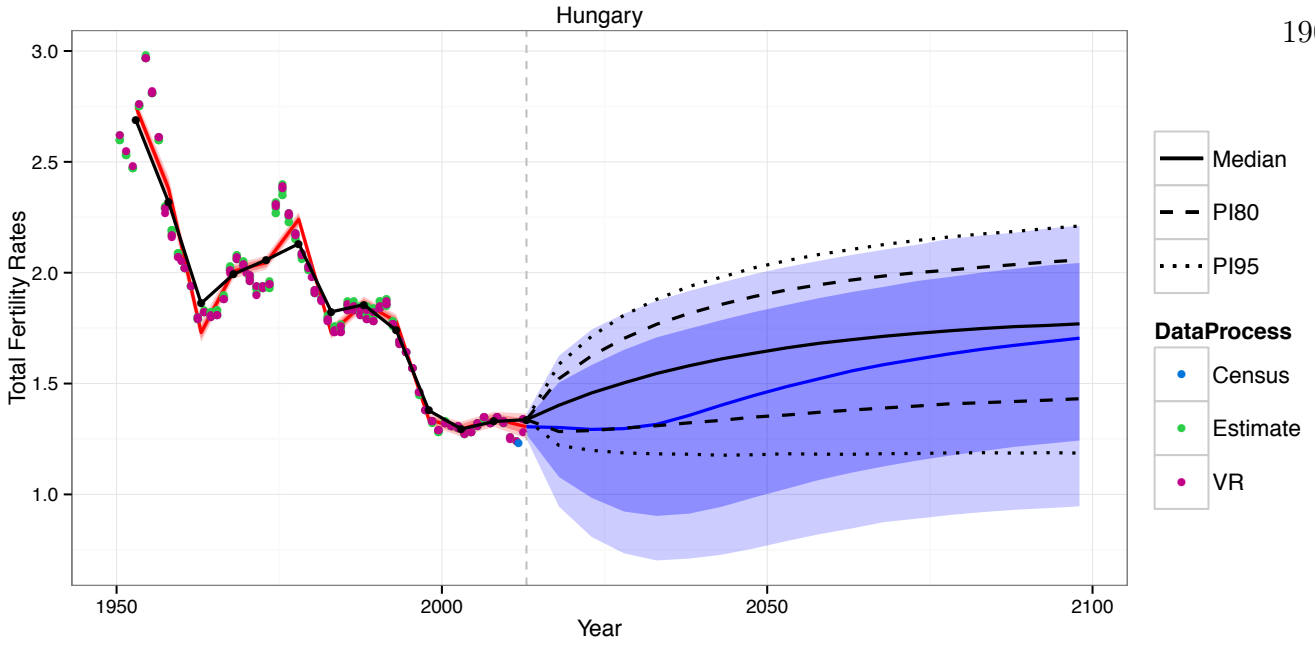

Guinea−Bissau
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
MICS

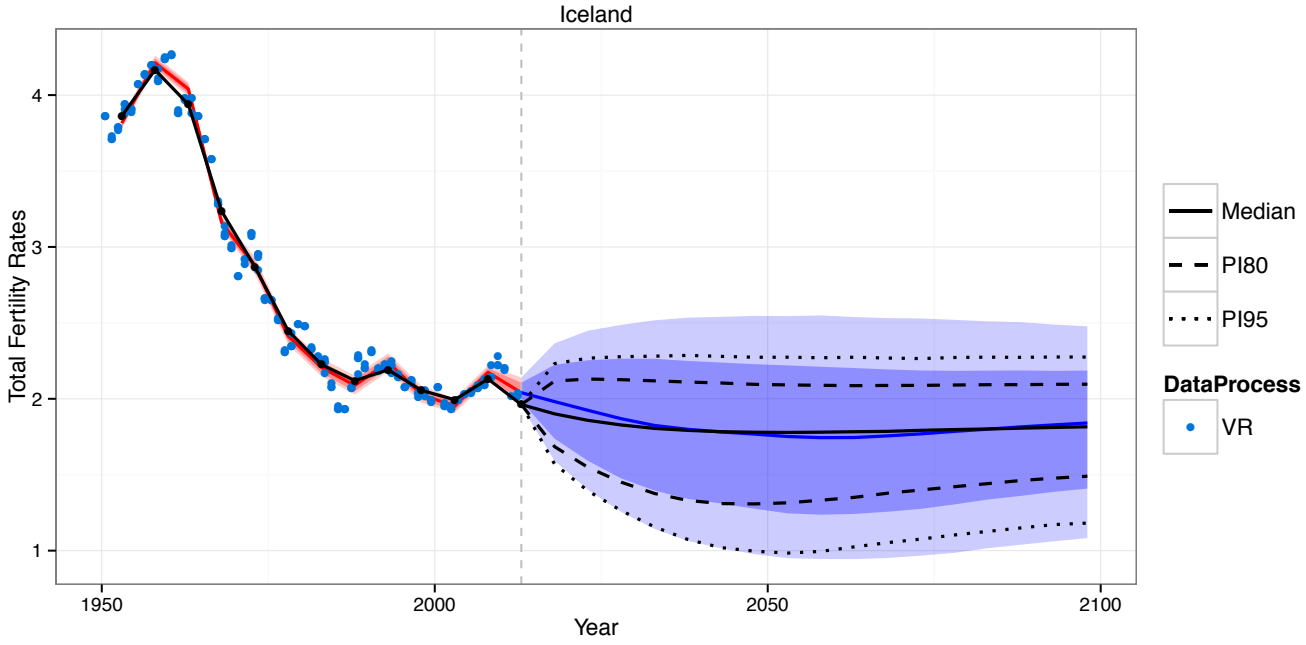

Guinea
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Survey
Median
PI80
PI95

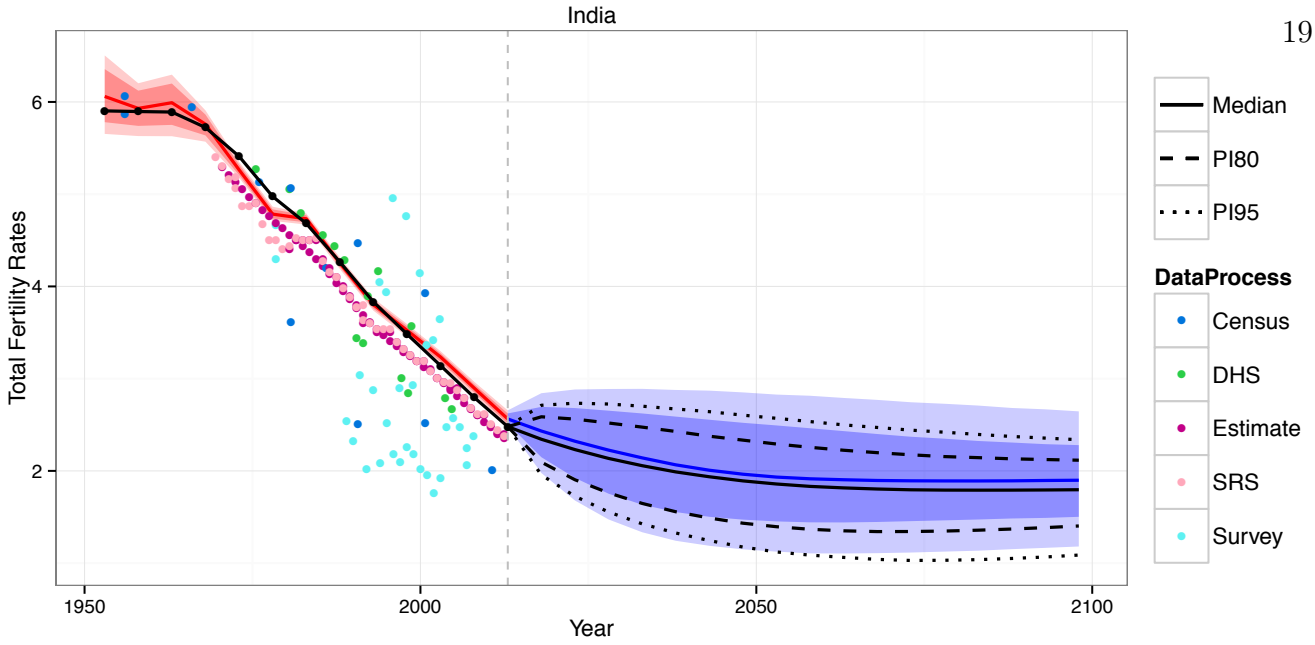

Guyana
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
Survey
VR
WFS

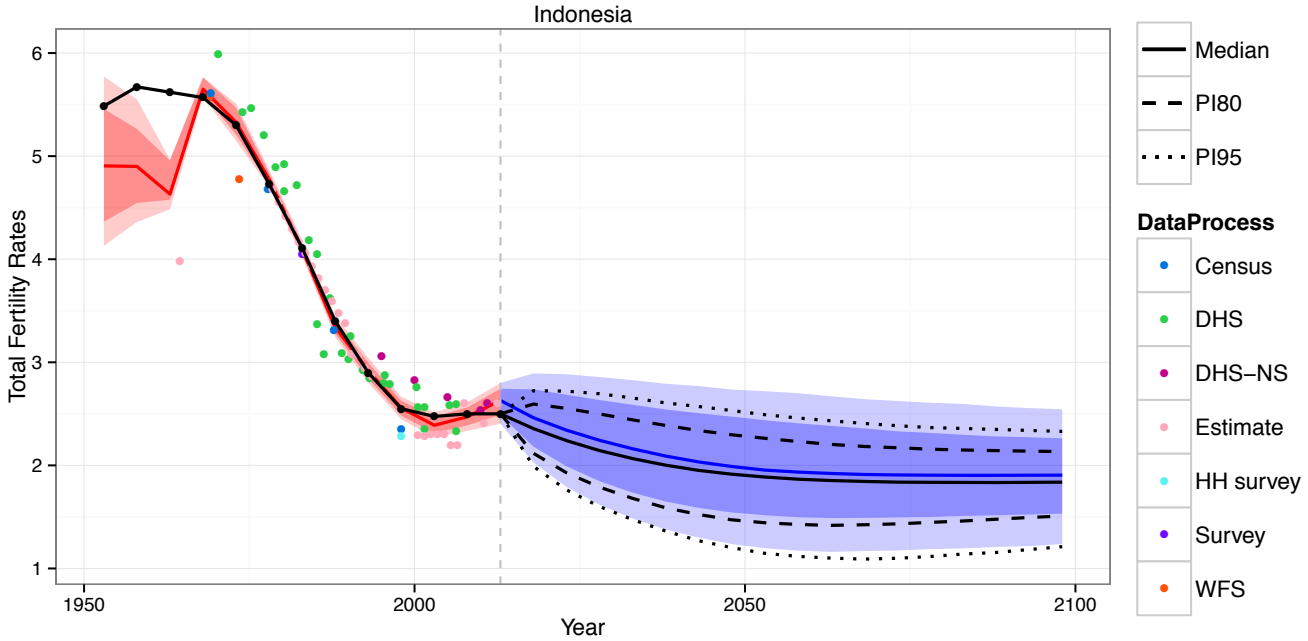

Haiti
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS

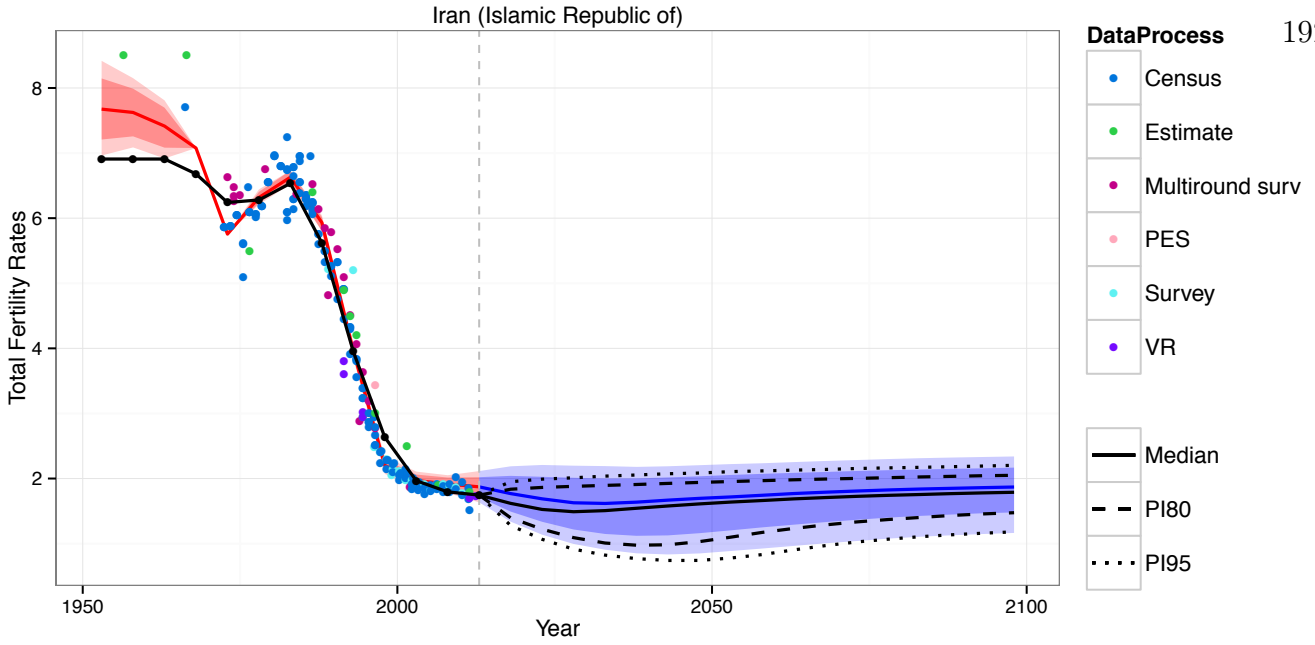

Honduras
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
DataProcess
DHS
Estimate
Multiround surv
RHS
Median
PI80
PI95

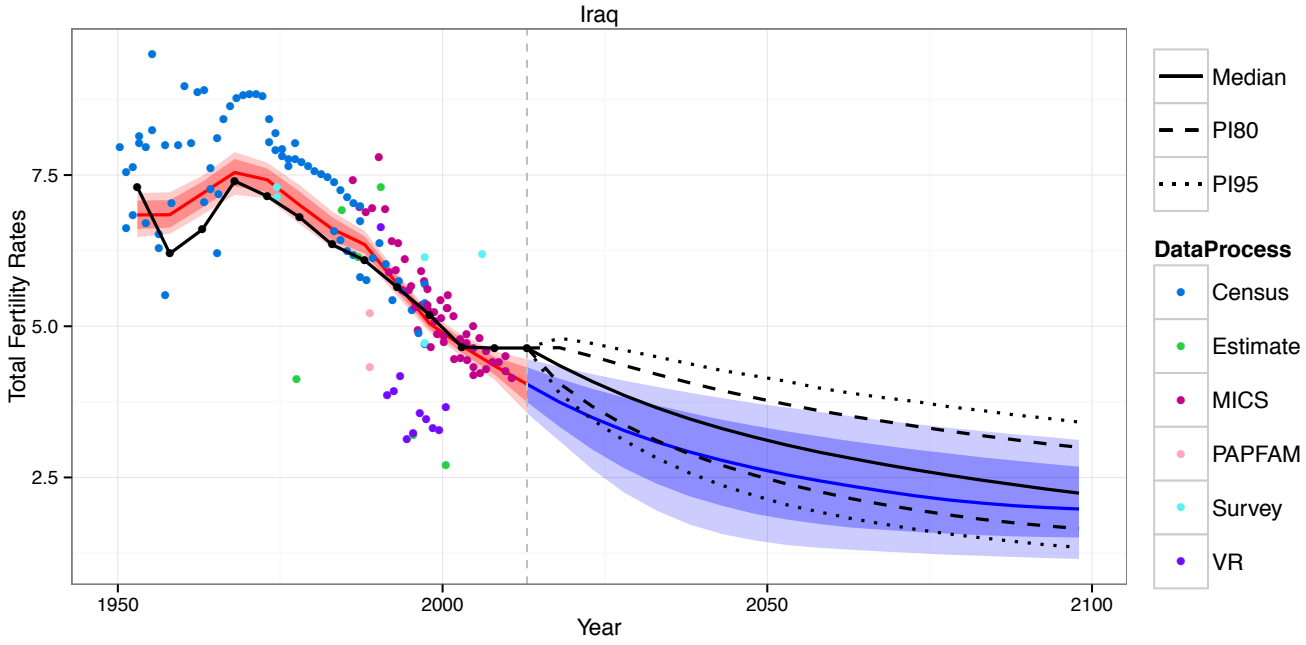

Hungary
Total Fertility Rates
3.0
2.5
2.0
1.5
1.0
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

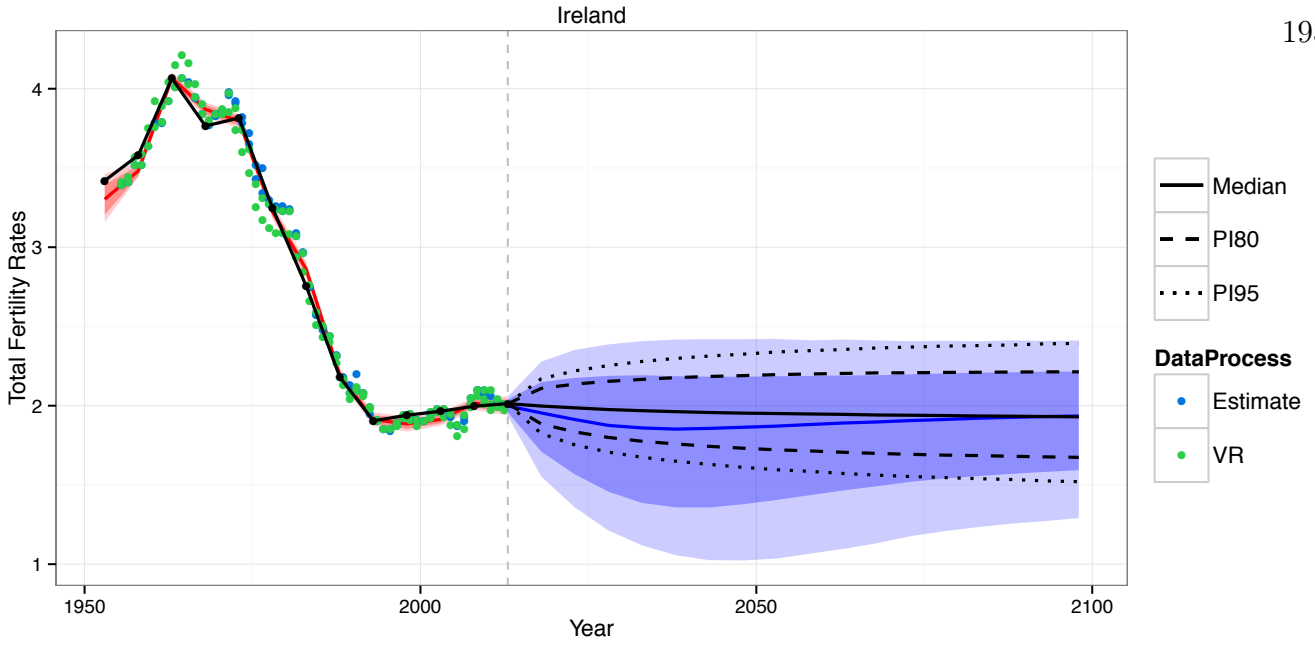

Iceland
Total Fertility Rates
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
VR

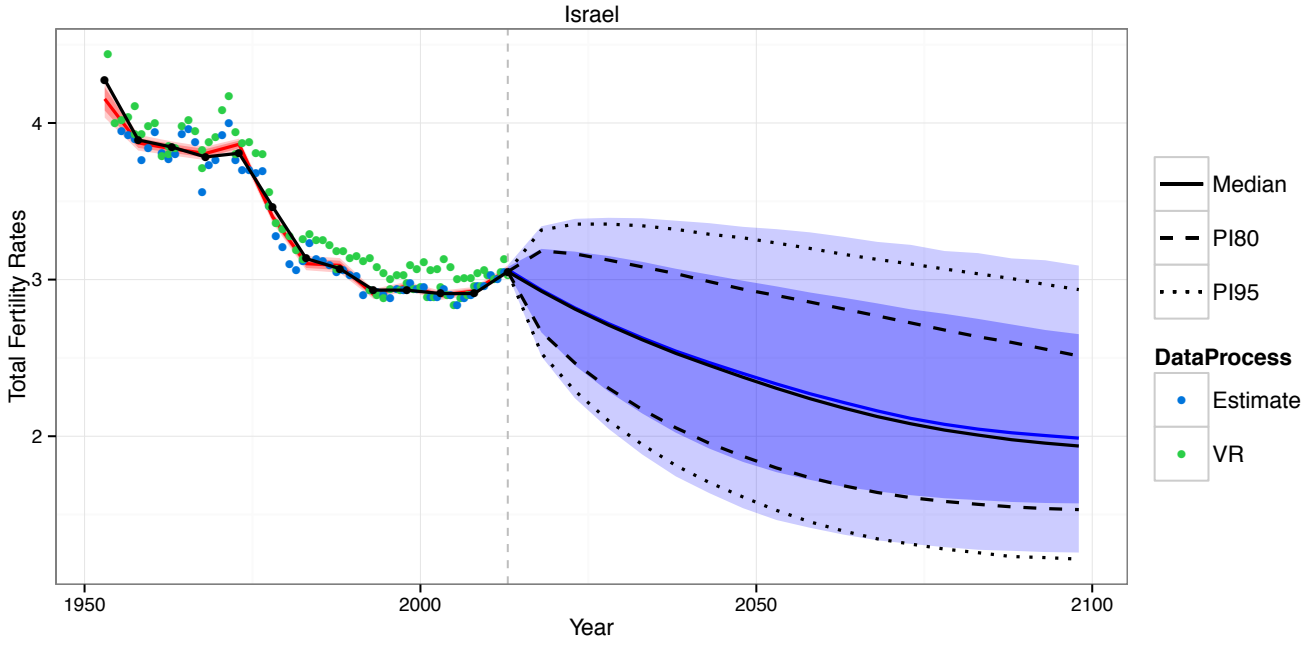

India
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
SRS
Survey

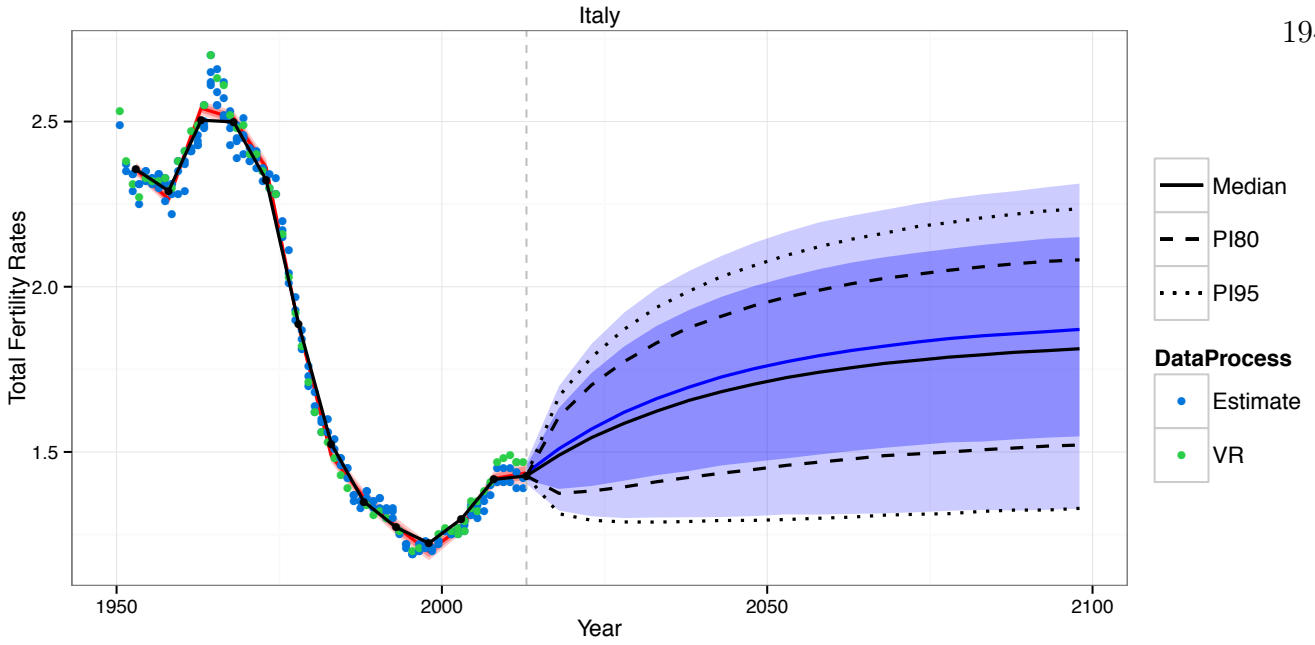

Indonesia
Total Fertility Rates
6
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
DHS−NS
Estimate
HH survey
Survey
WFS

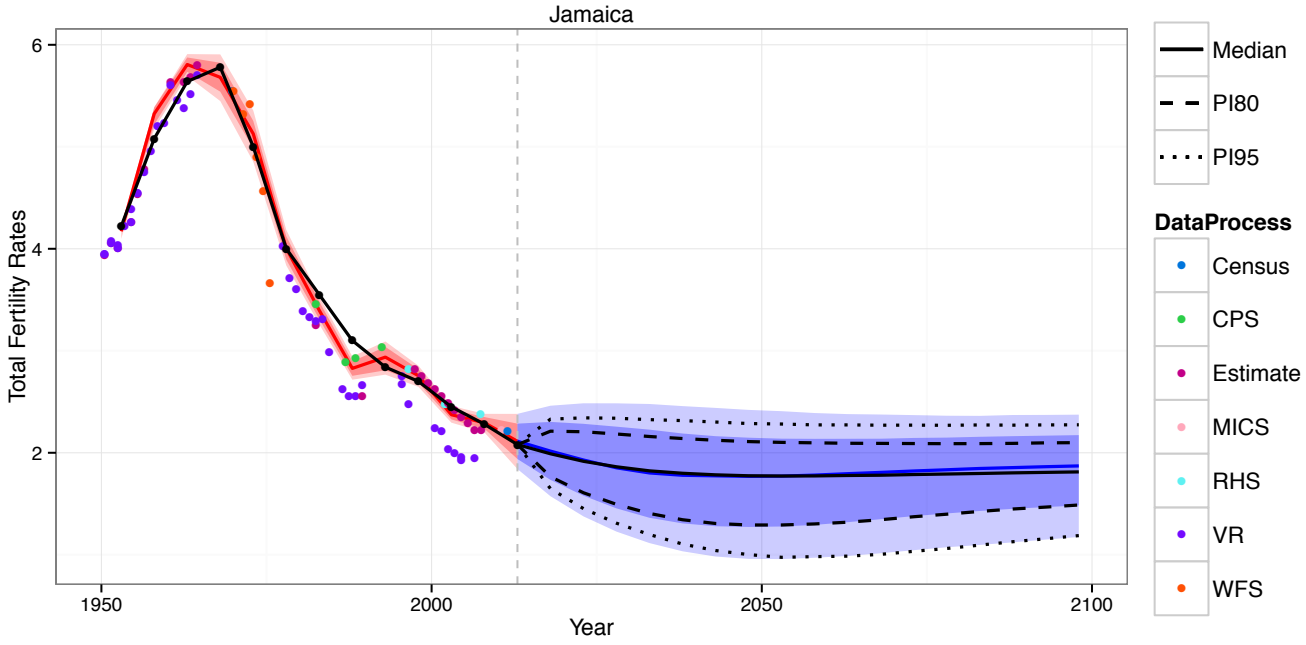

Iran (Islamic Republic of)
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
Estimate
Multiround surv
PES
Survey
VR
Median
PI80
PI95

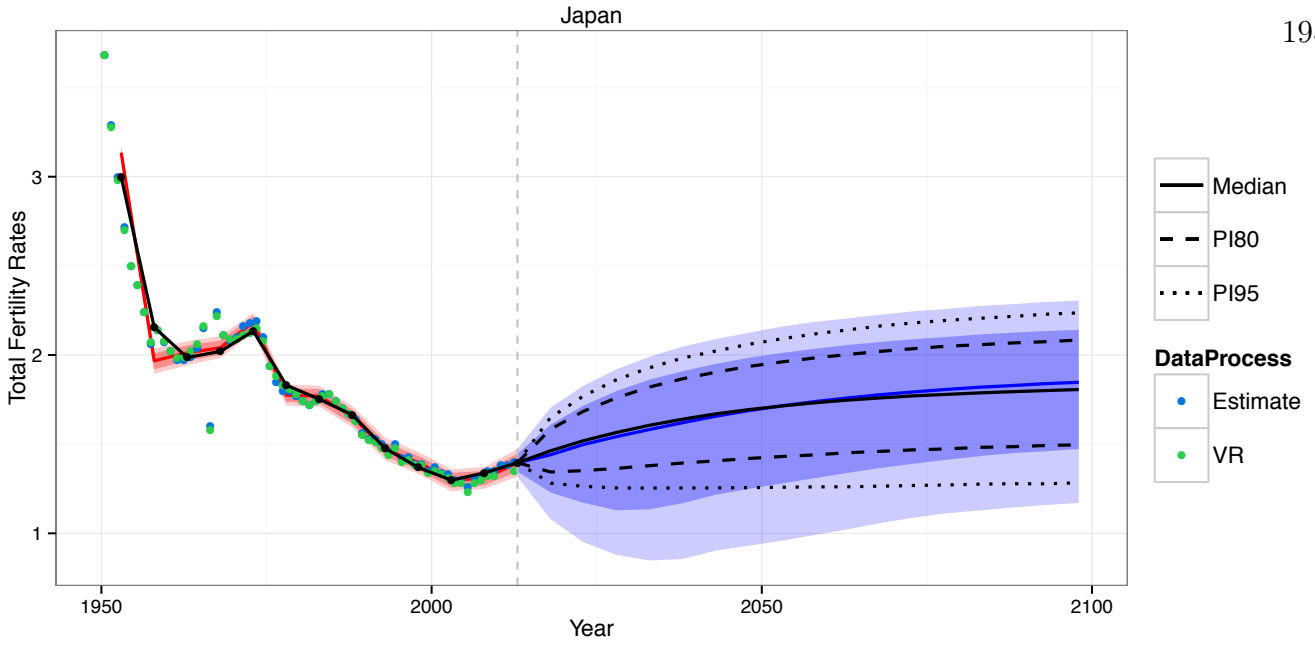

Iraq
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
MICS
PAPFAM
Survey
VR

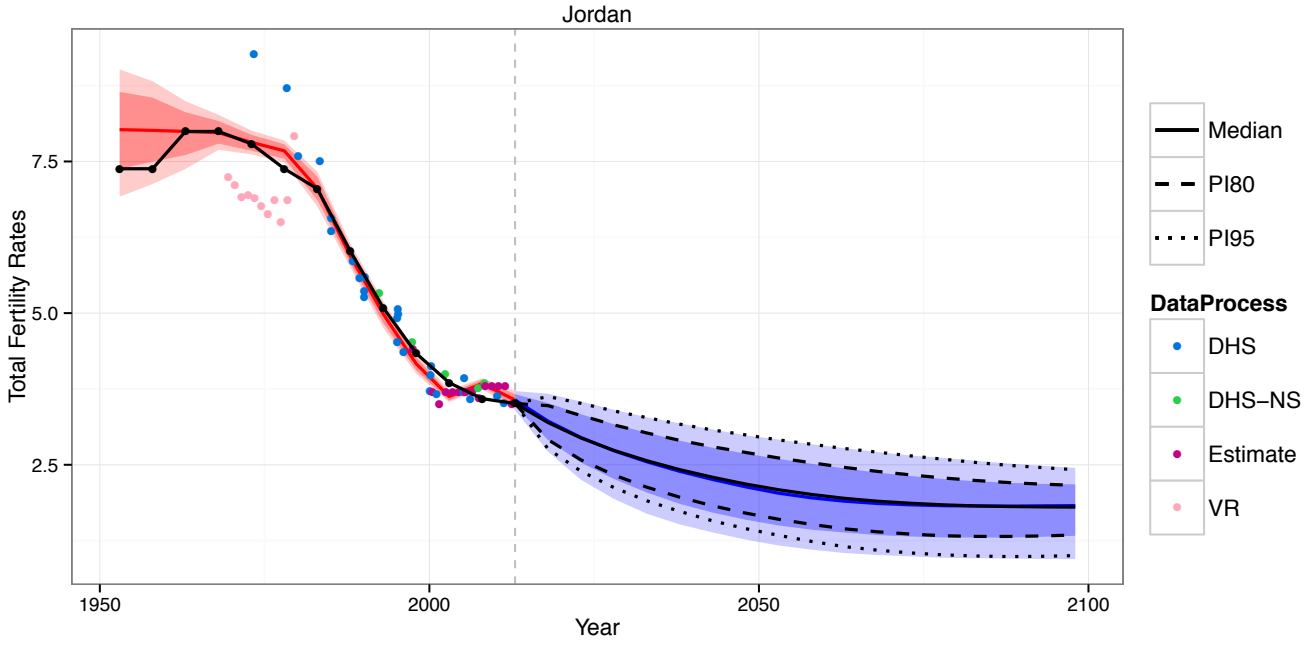

Ireland
Total Fertility Rates
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

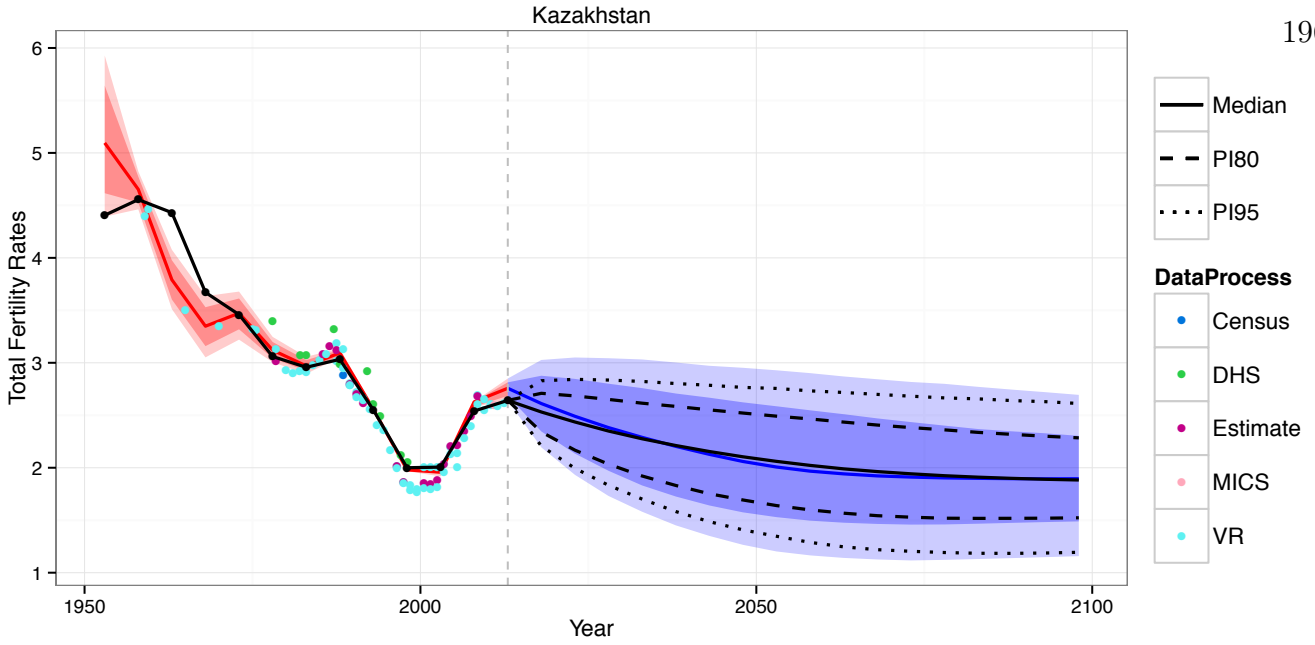

Israel
Total Fertility Rates
4
3
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

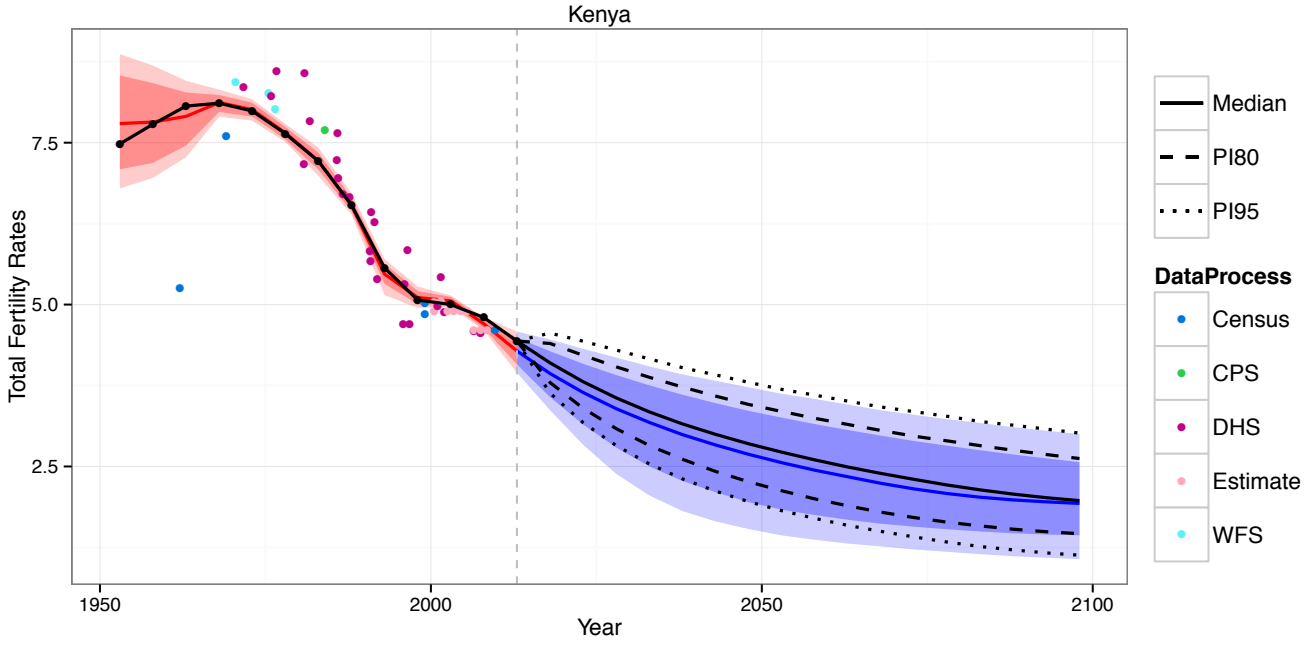

Italy
Total Fertility Rates
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

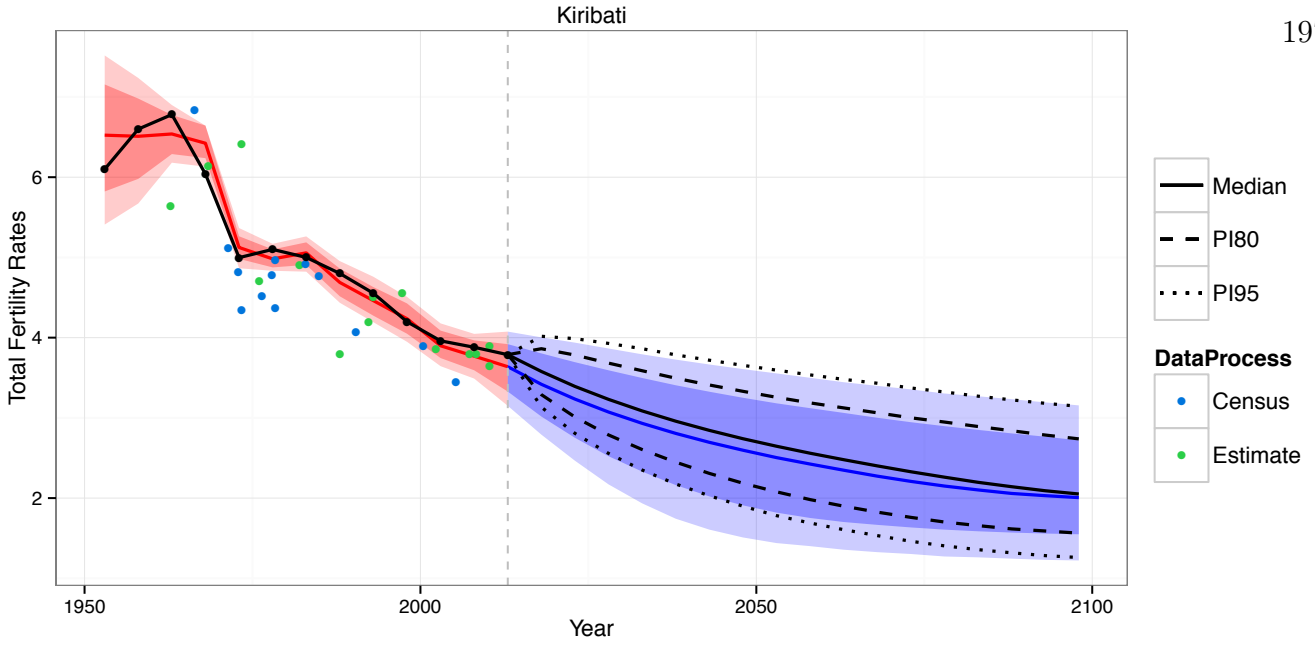

Jamaica
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
Estimate
MICS
RHS
VR
WFS

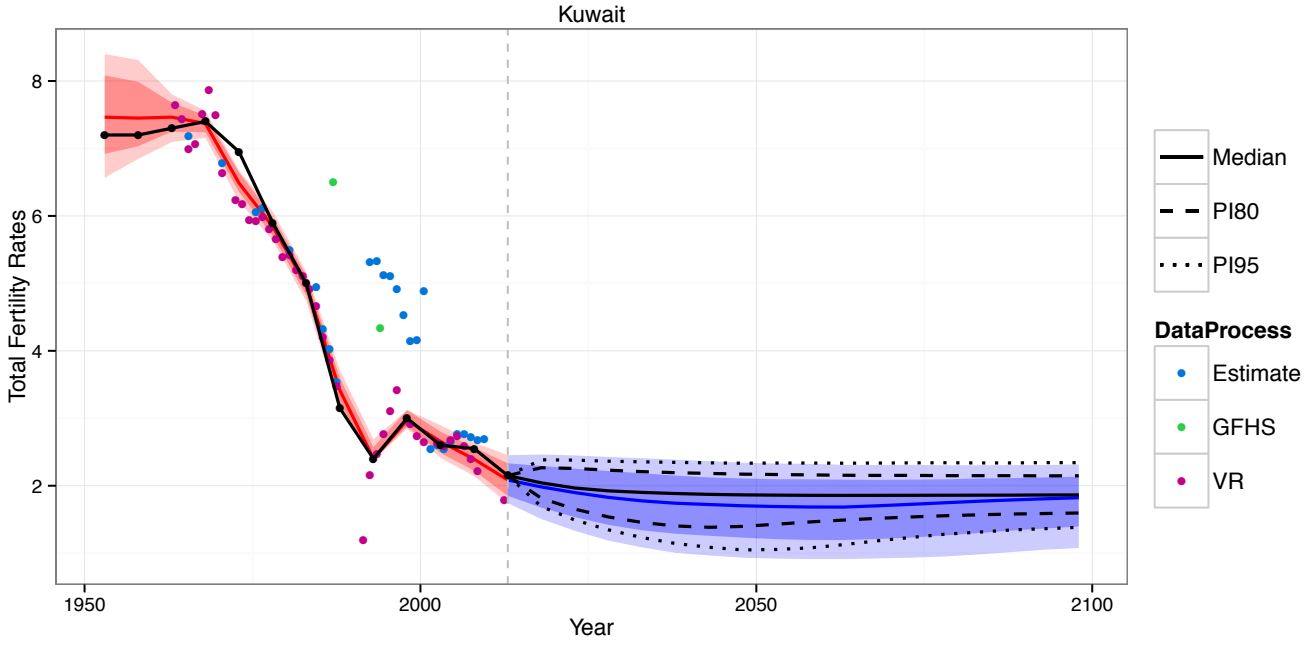

Japan
Total Fertility Rates
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

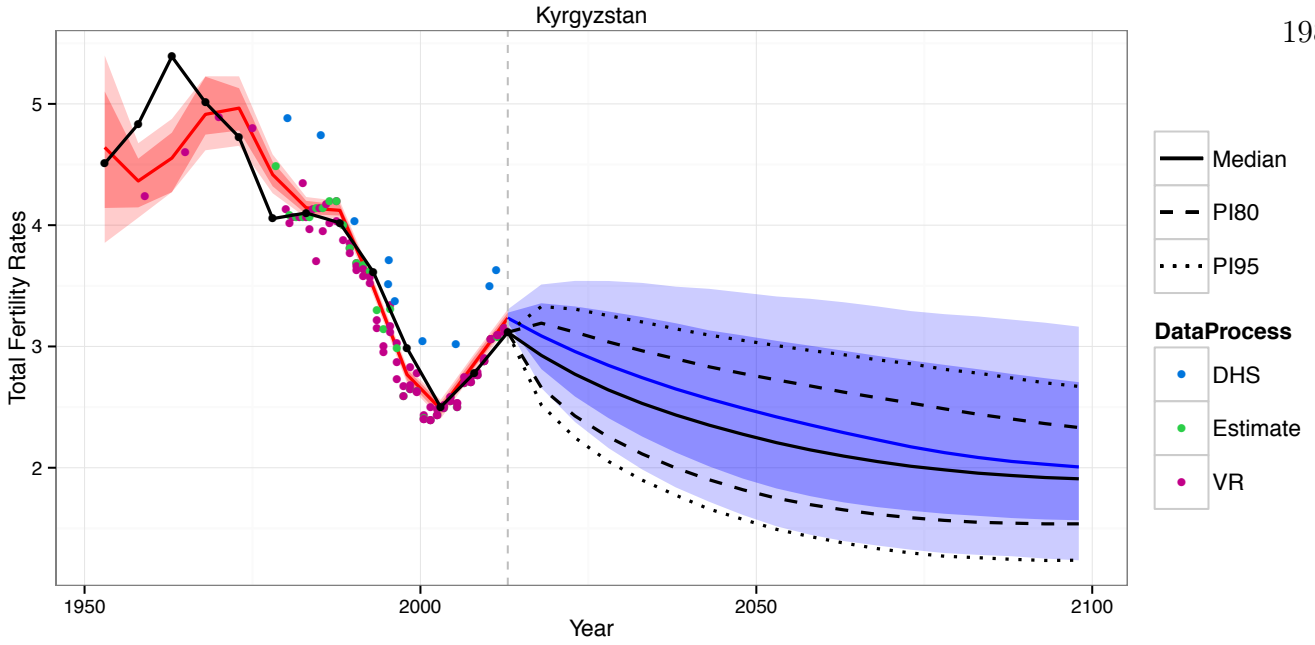

Jordan
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
DHS−NS
Estimate
VR

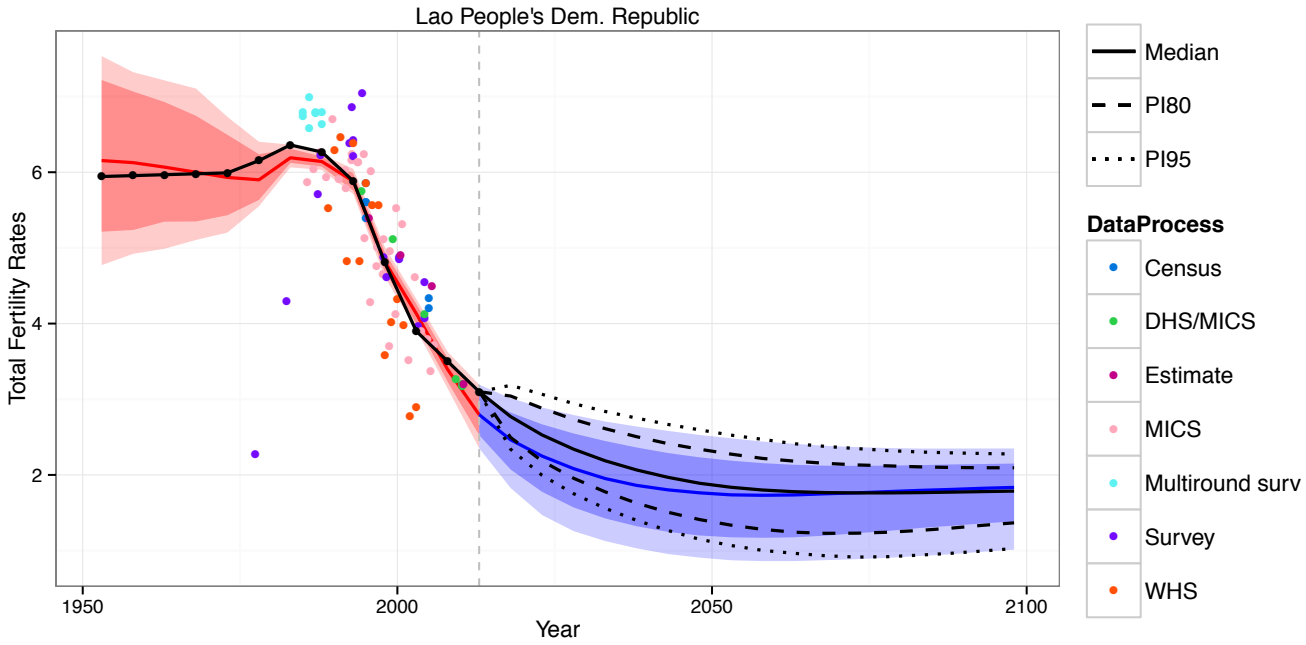

Kazakhstan
Total Fertility Rates
6
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS
VR

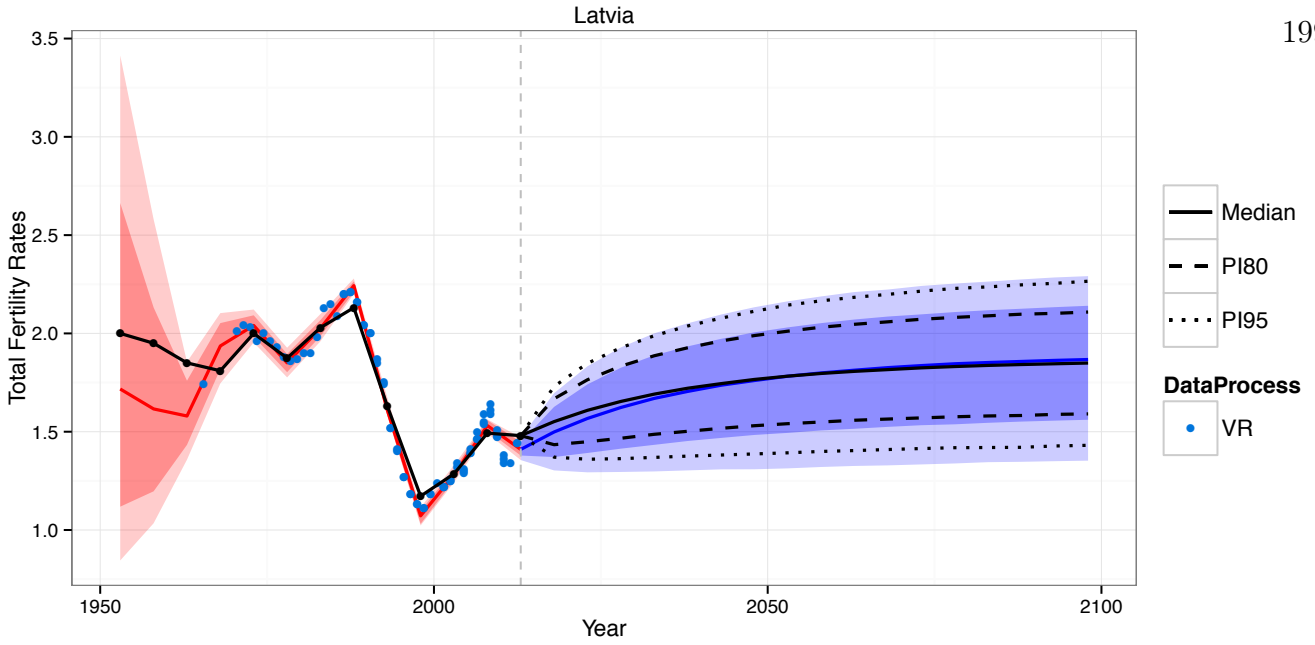

Kenya
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
DHS
Estimate
WFS

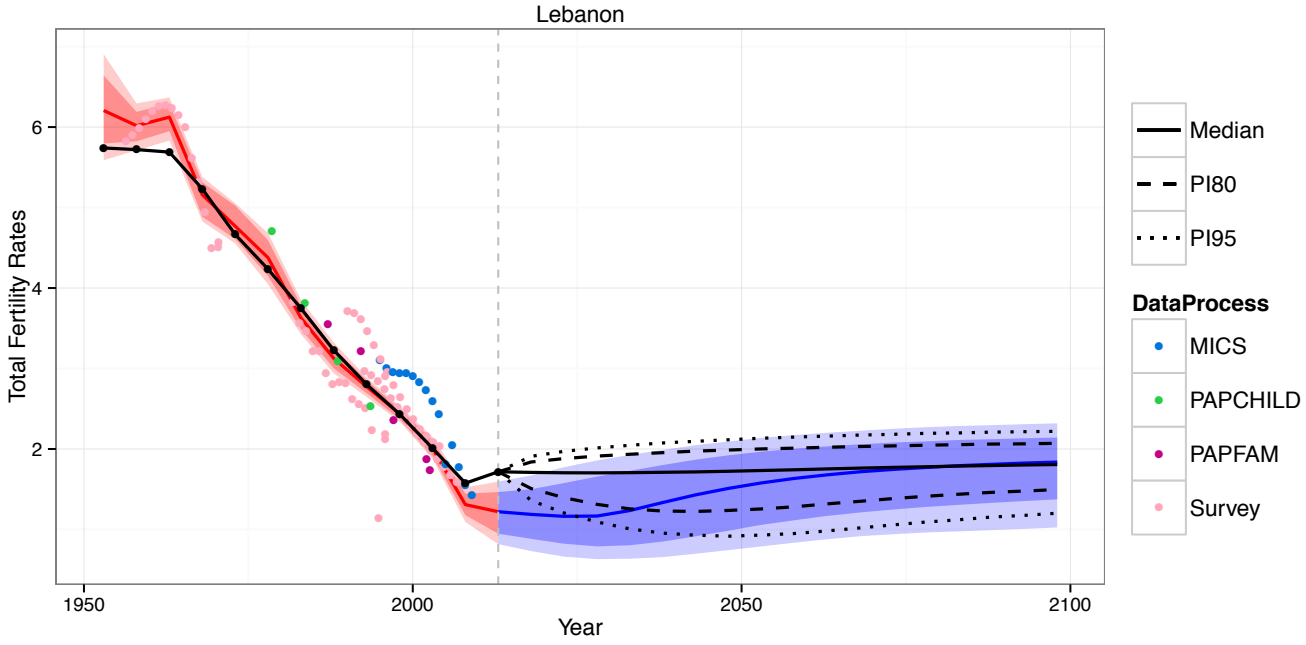

Kiribati
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate

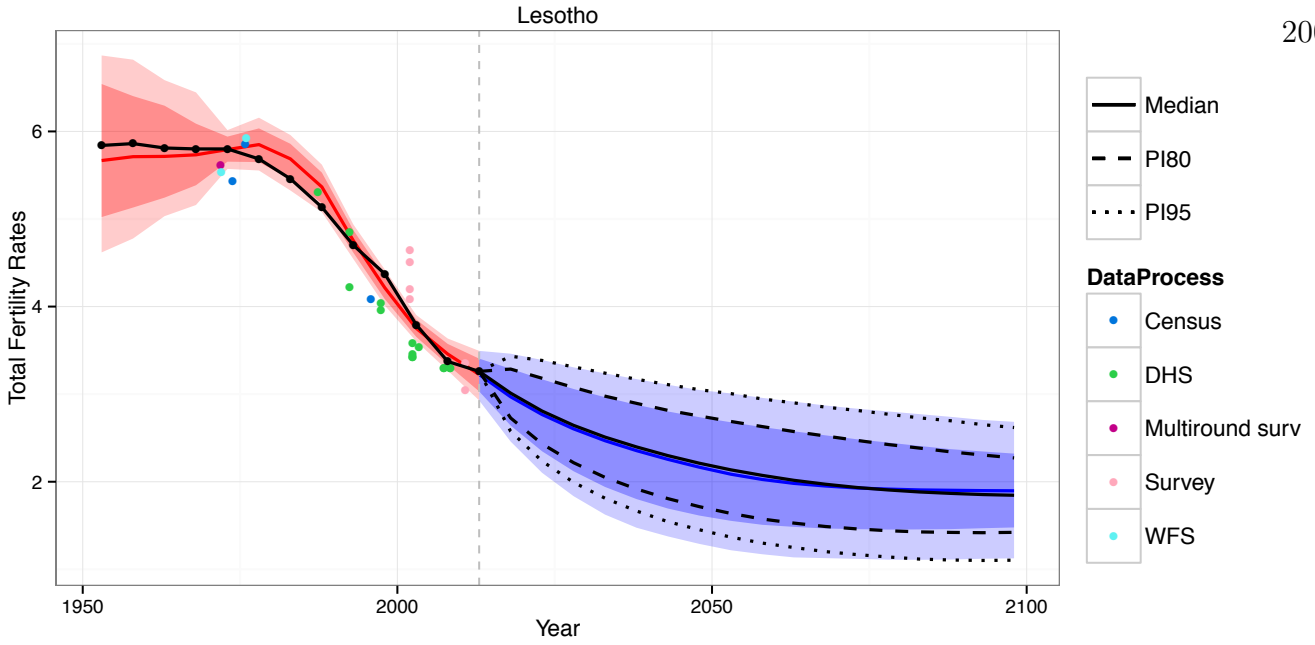

Kuwait
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
GFHS
VR

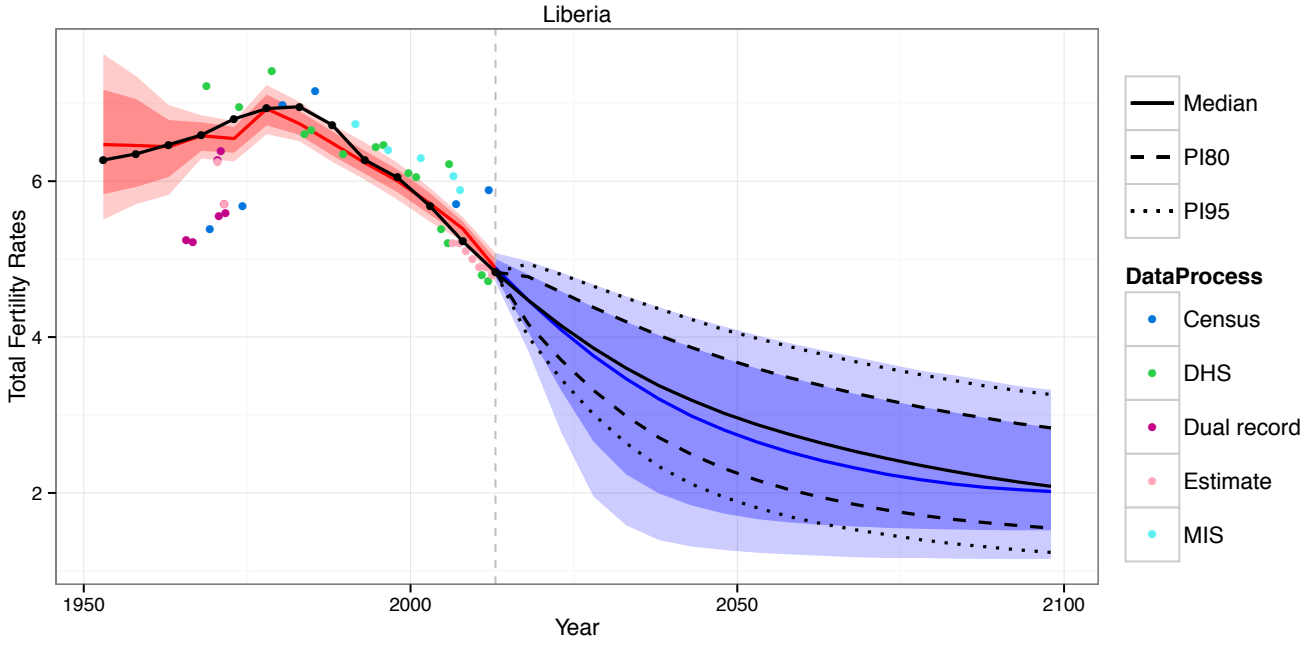

Kyrgyzstan
Total Fertility Rates
5
4
3
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
VR

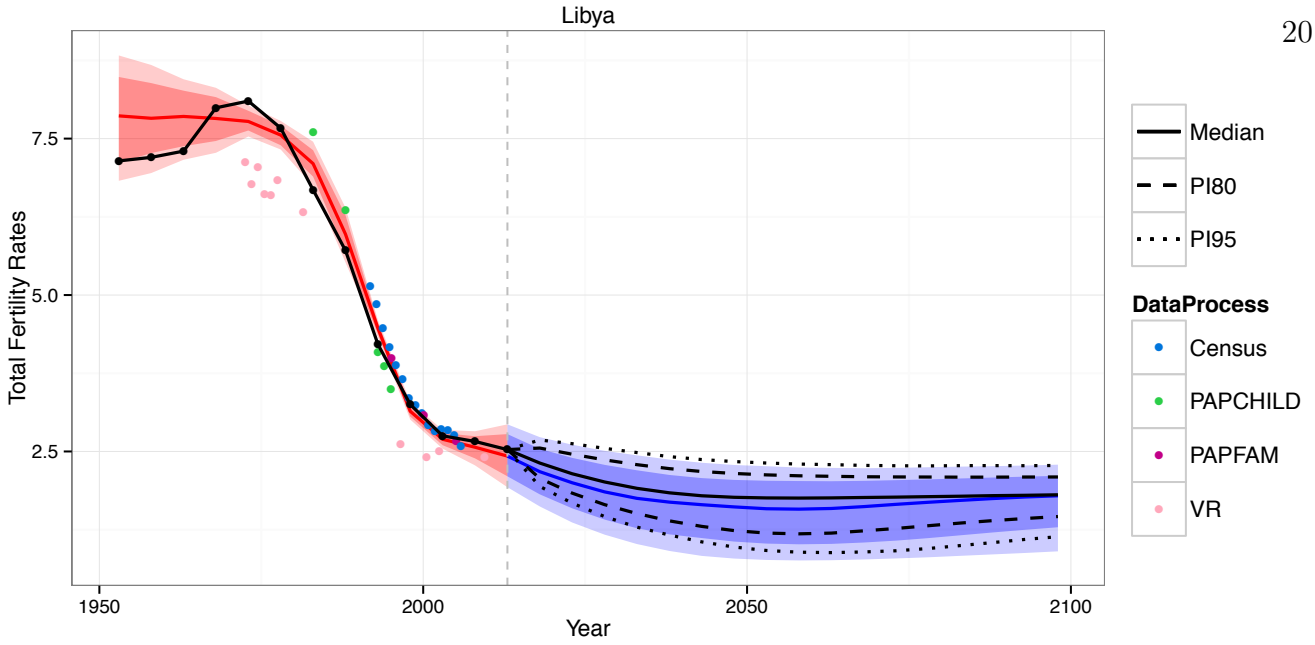

Lao People's Dem. Republic
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS/MICS
Estimate
MICS
Multiround surv
Survey
WHS

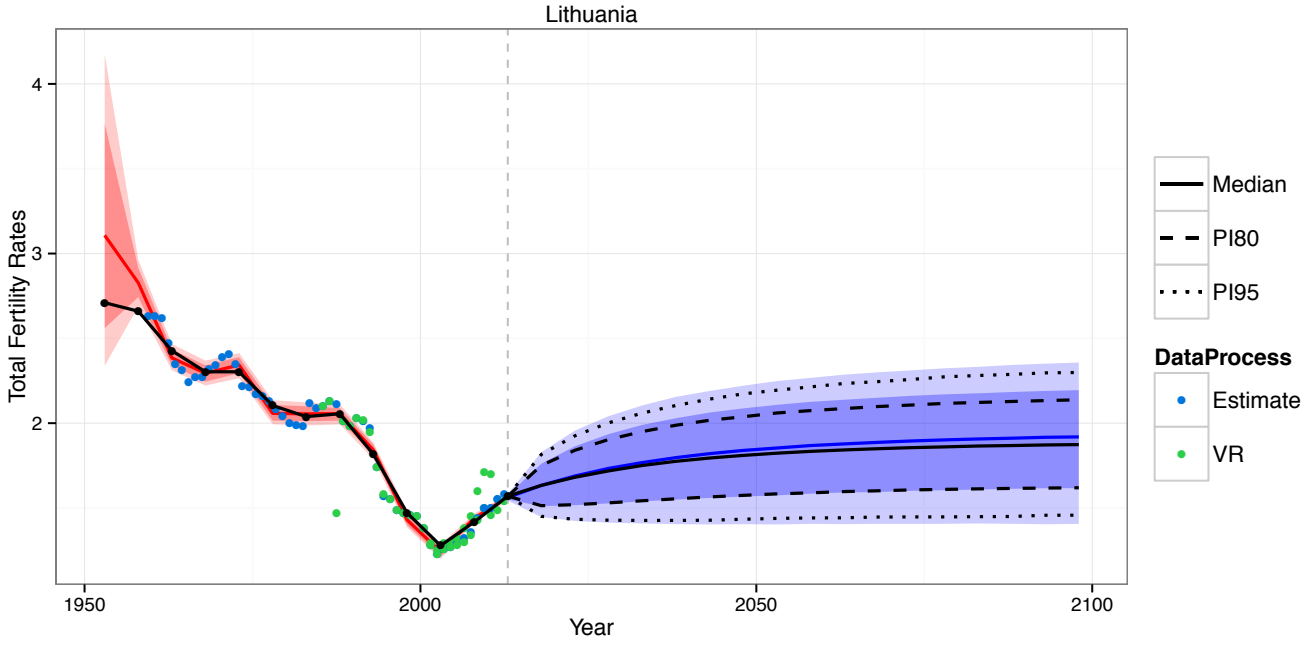

Latvia
Total Fertility Rates
3.5
3.0
2.5
2.0
1.5
1.0
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
VR

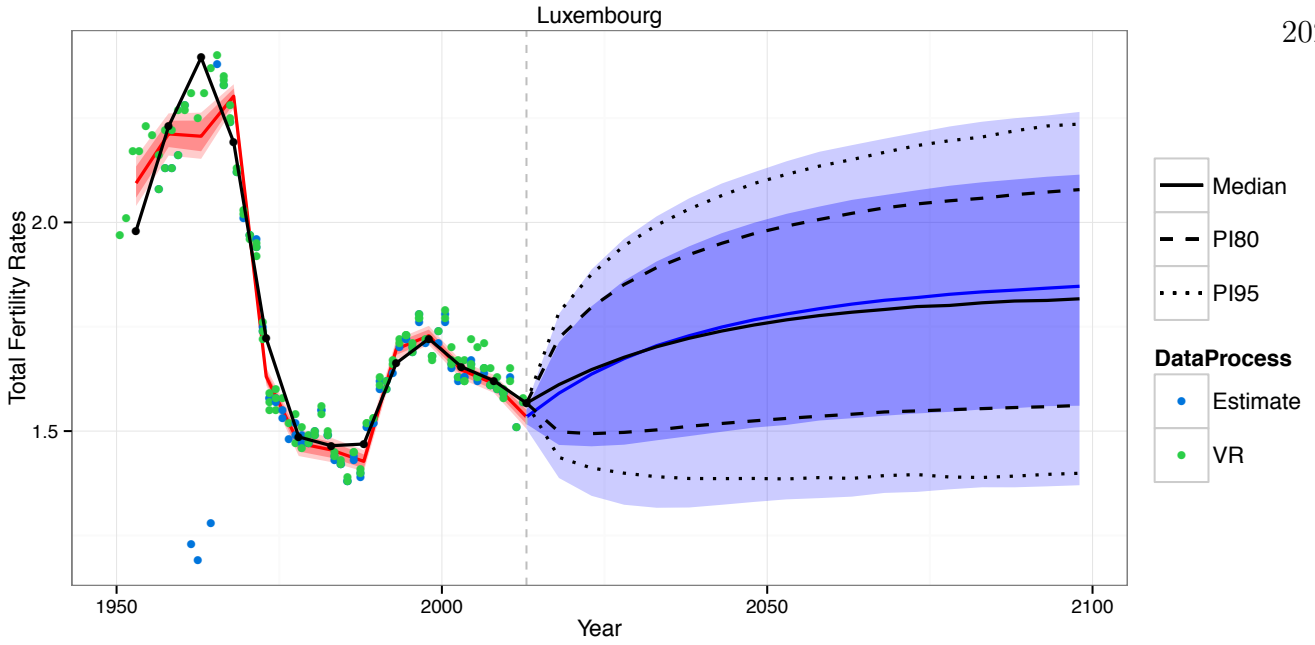

Lebanon
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
MICS
PAPCHILD
PAPFAM
Survey

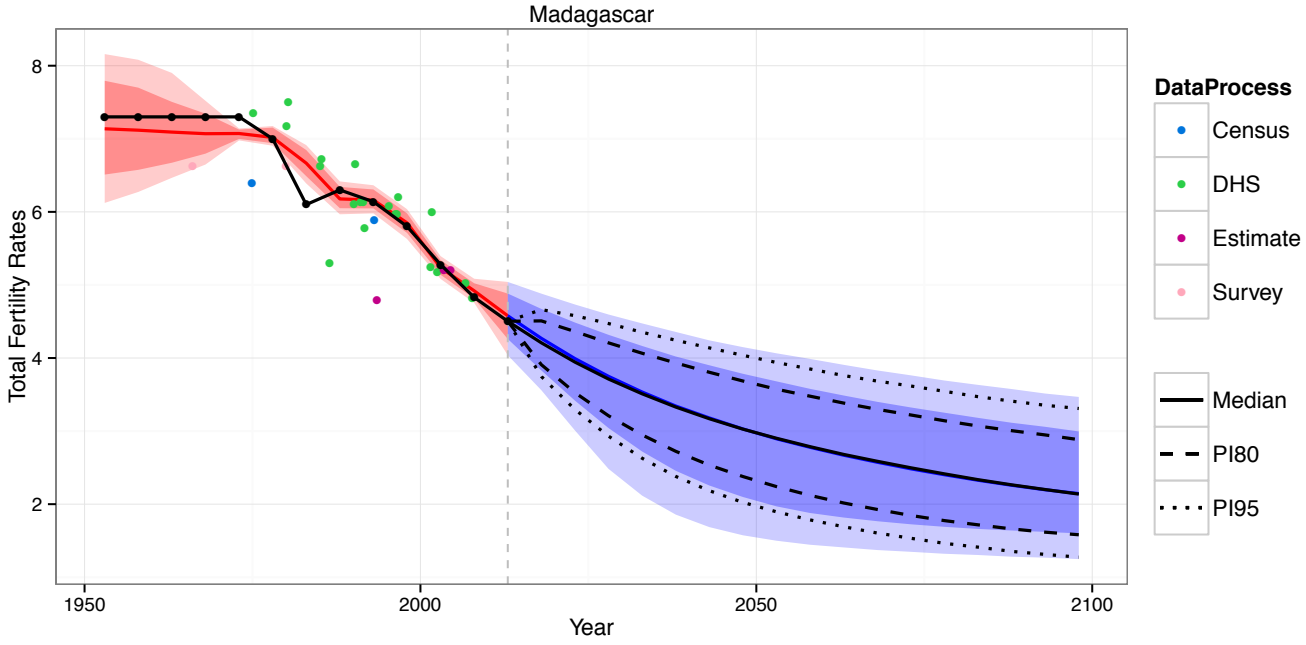

Lesotho
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Multiround surv
Survey
WFS

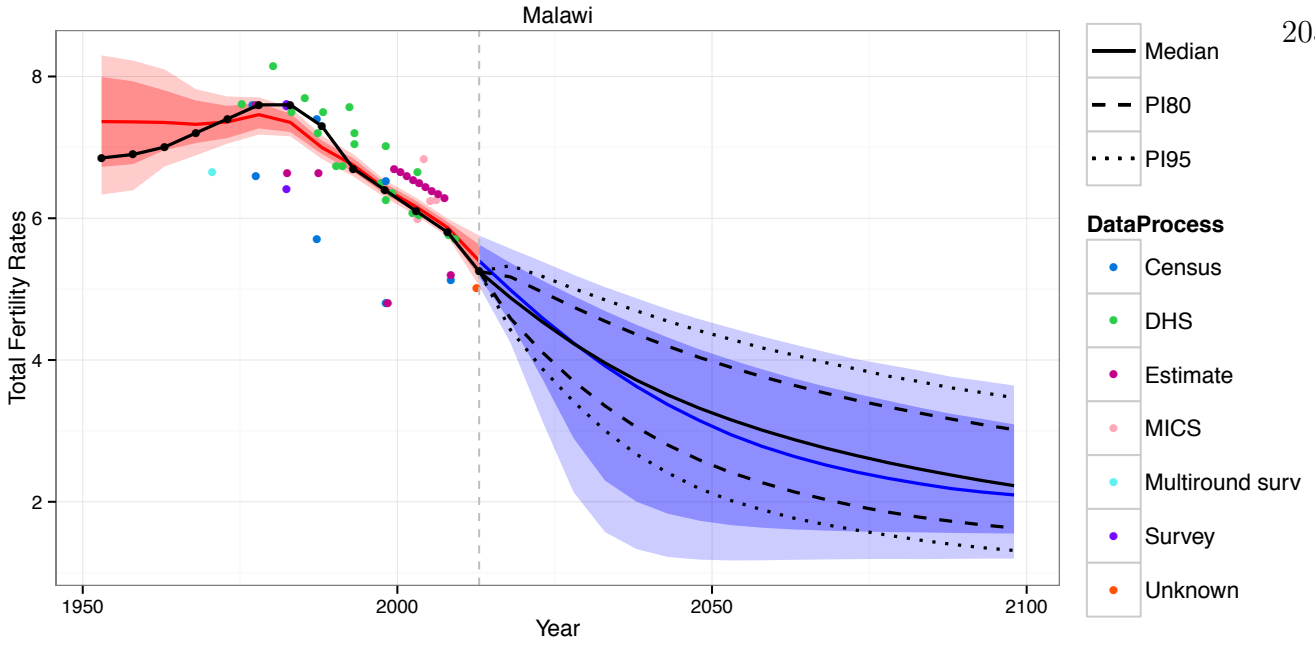

Liberia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Dual record
Estimate
MIS

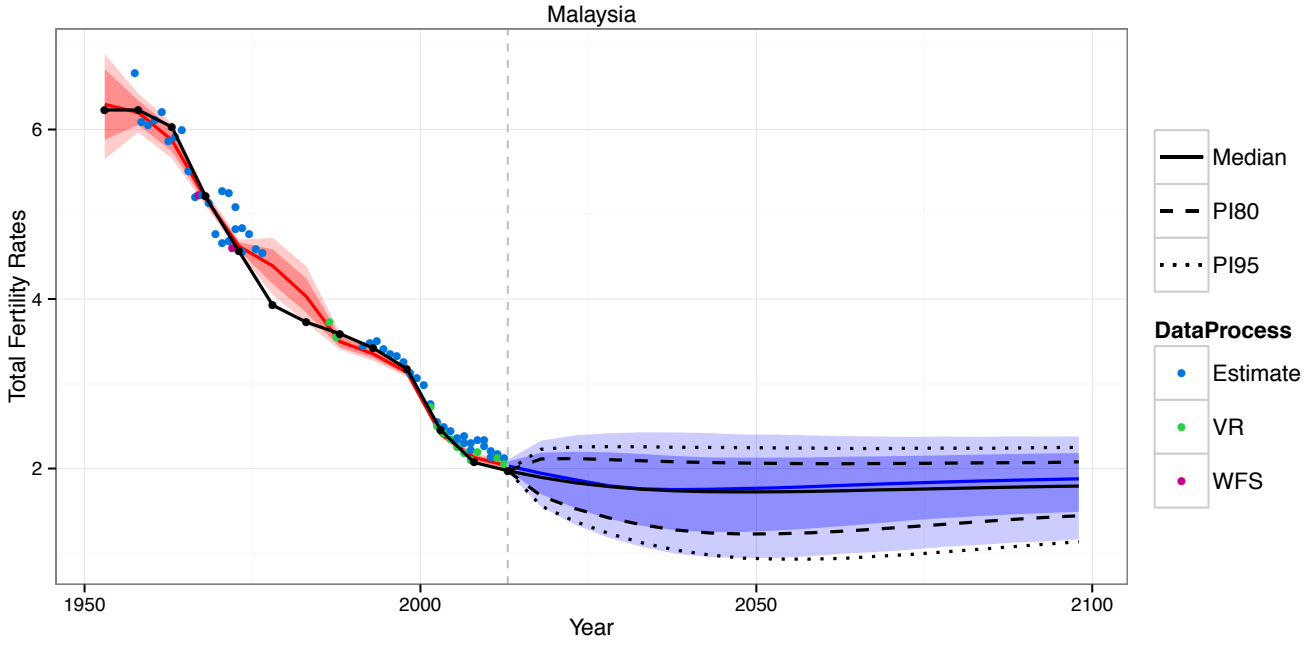

Libya
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
PAPCHILD
PAPFAM
VR

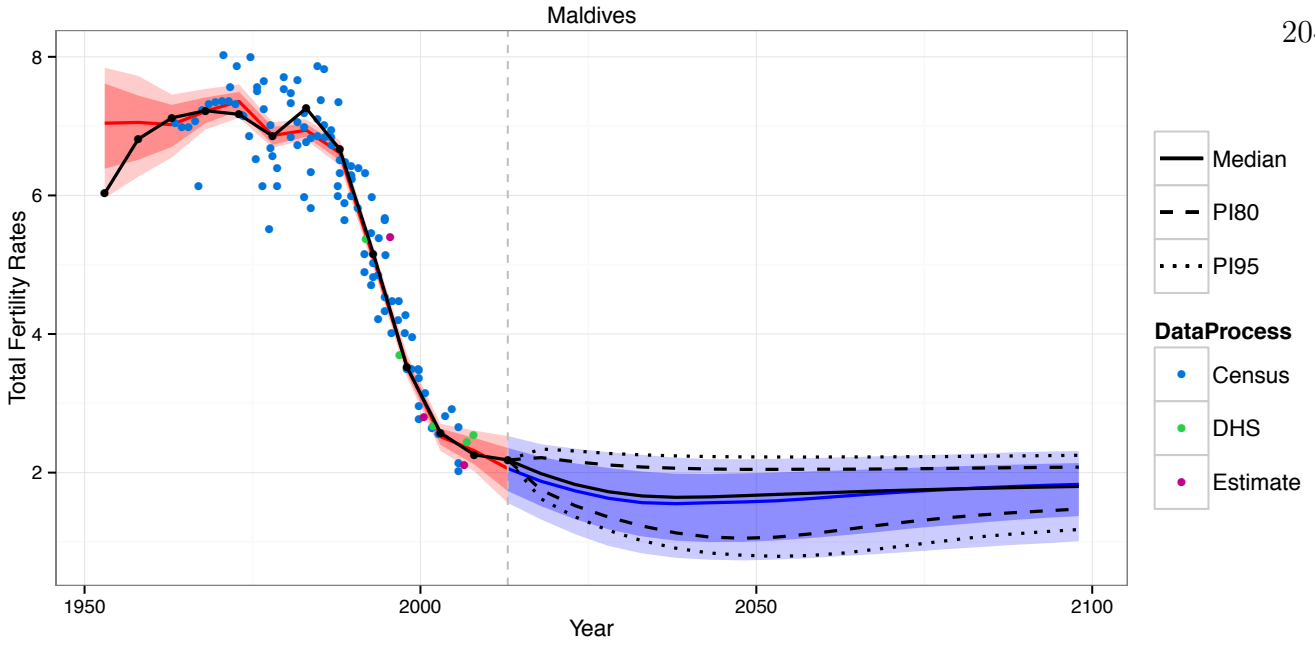

Lithuania
Total Fertility Rates
4
3
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

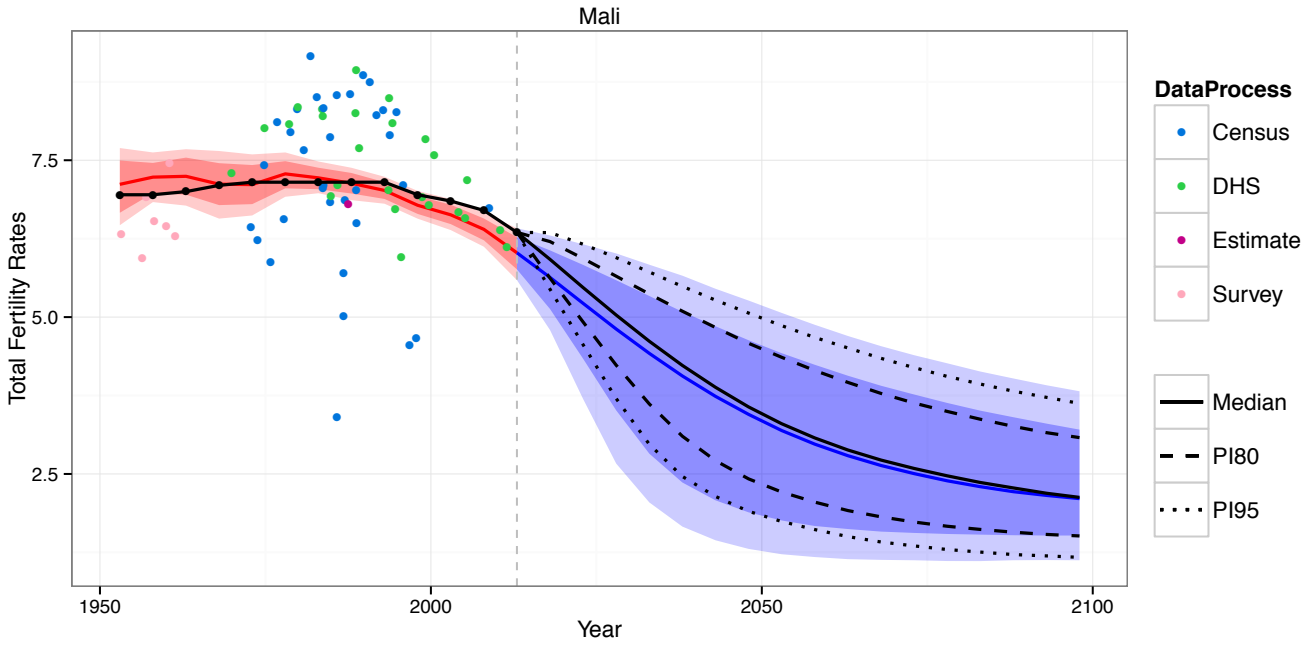

Luxembourg
Total Fertility Rates
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

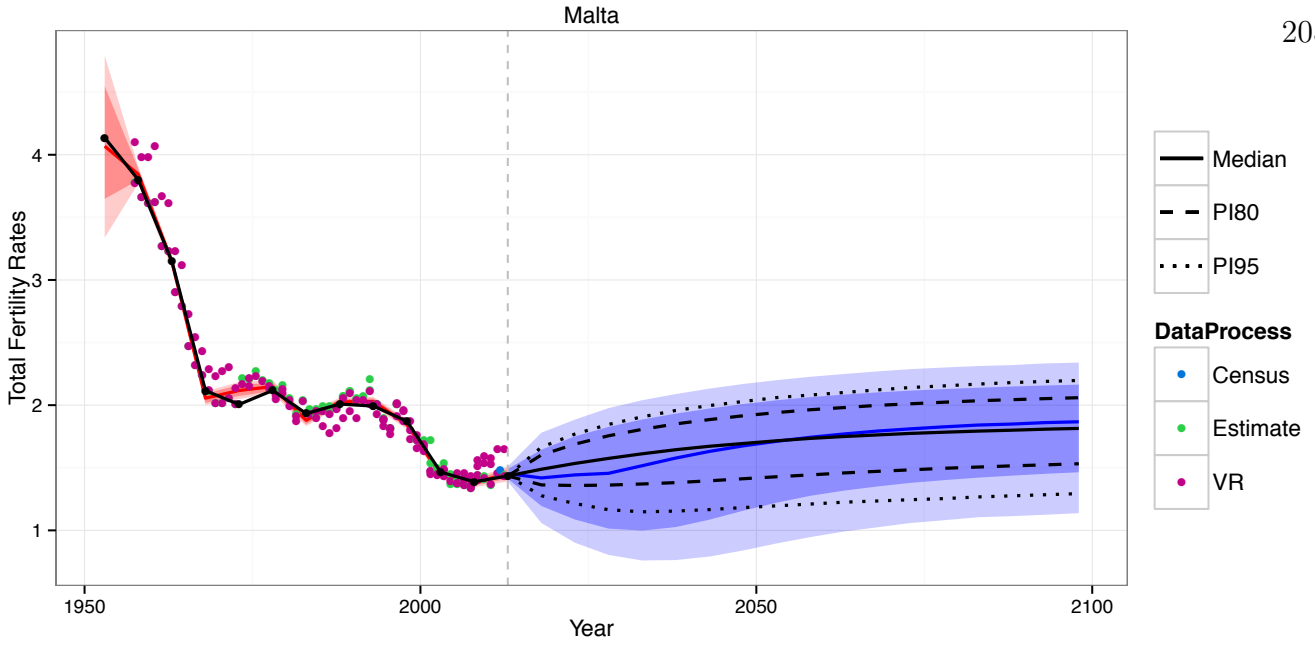

Madagascar
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Estimate
Survey
Median
PI80
PI95

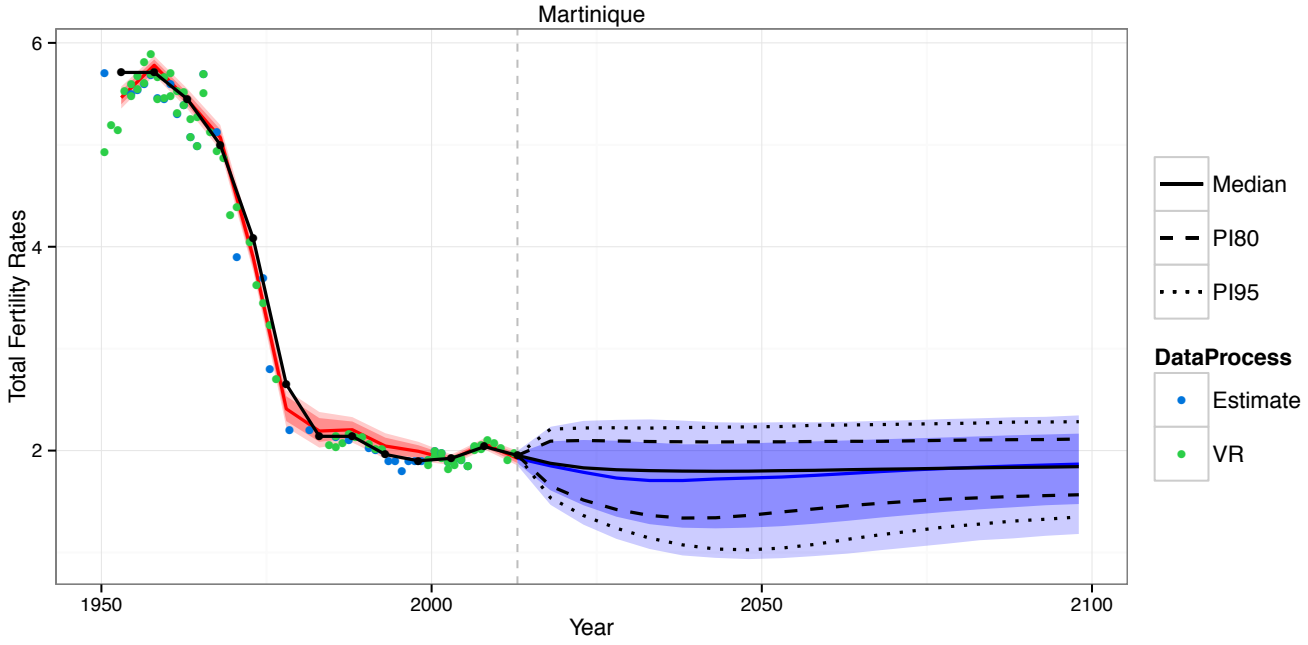

Malawi
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS
Multiround surv
Survey
Unknown

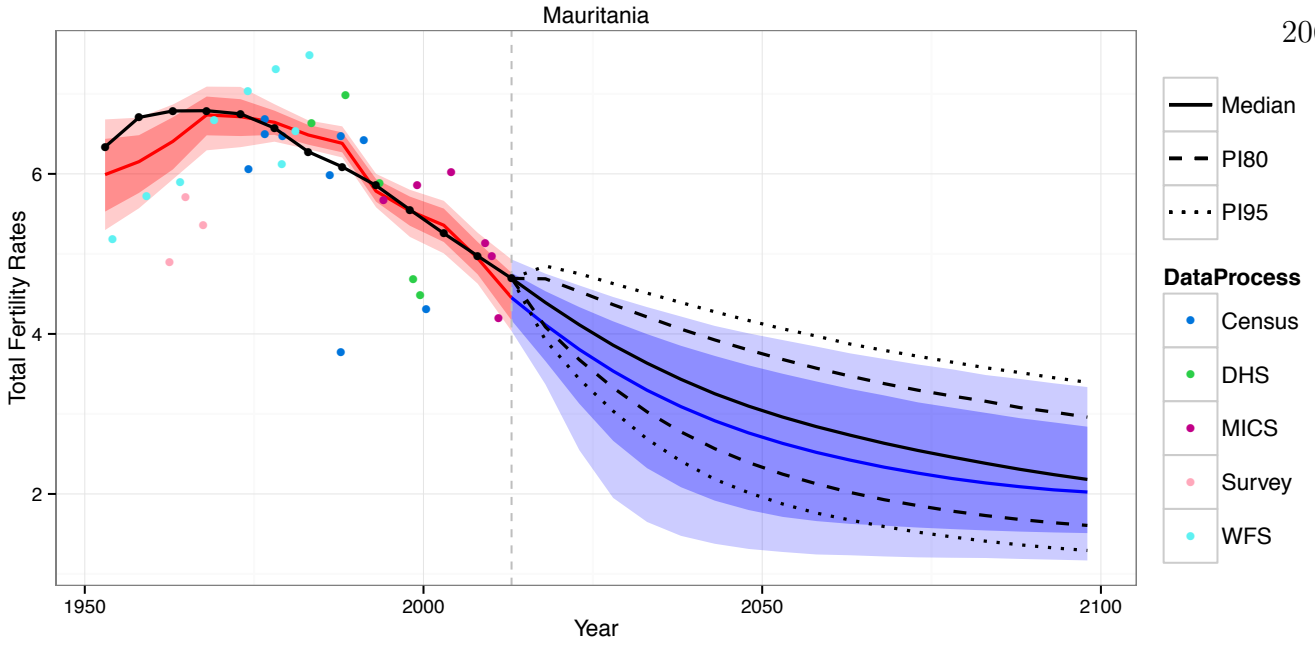

Malaysia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR
WFS

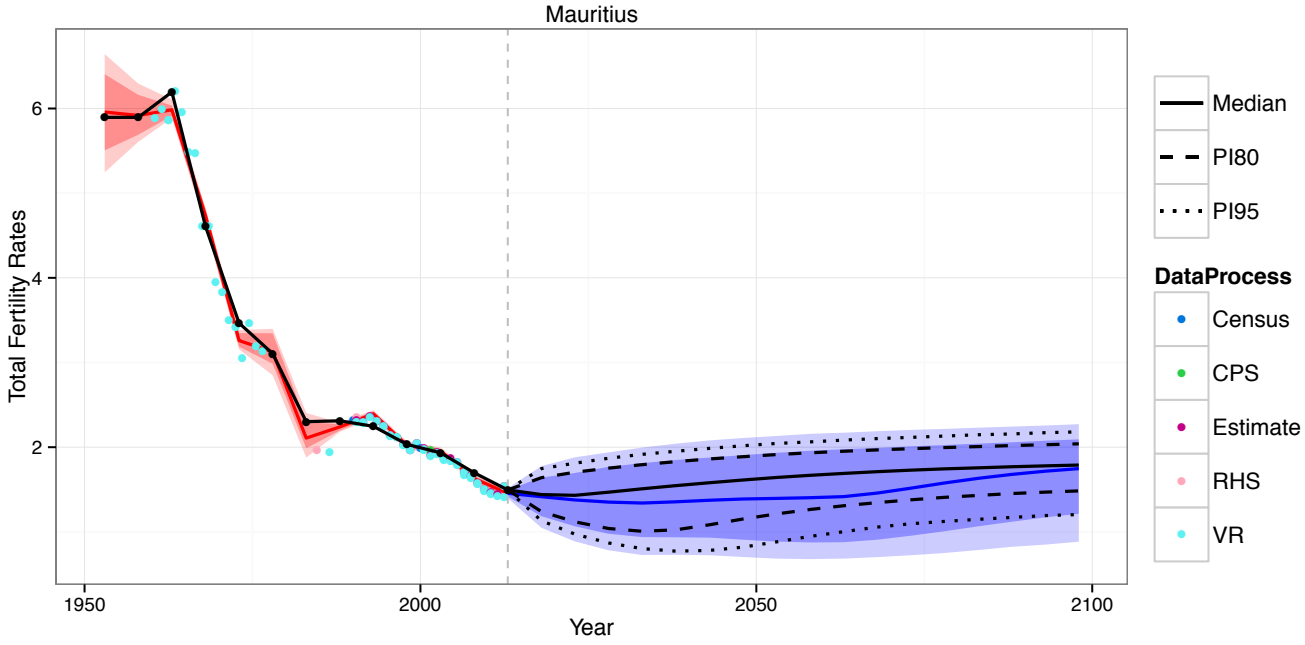

Maldives
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate

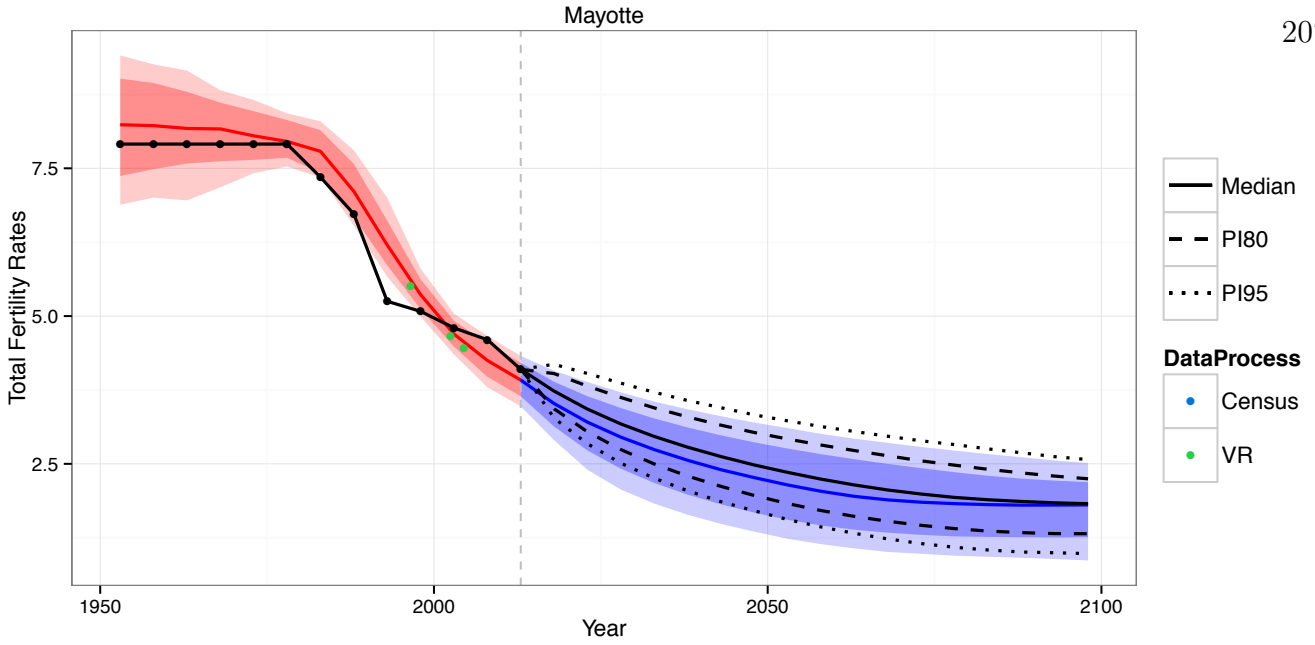

Mali
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Estimate
Survey
Median
PI80
PI95

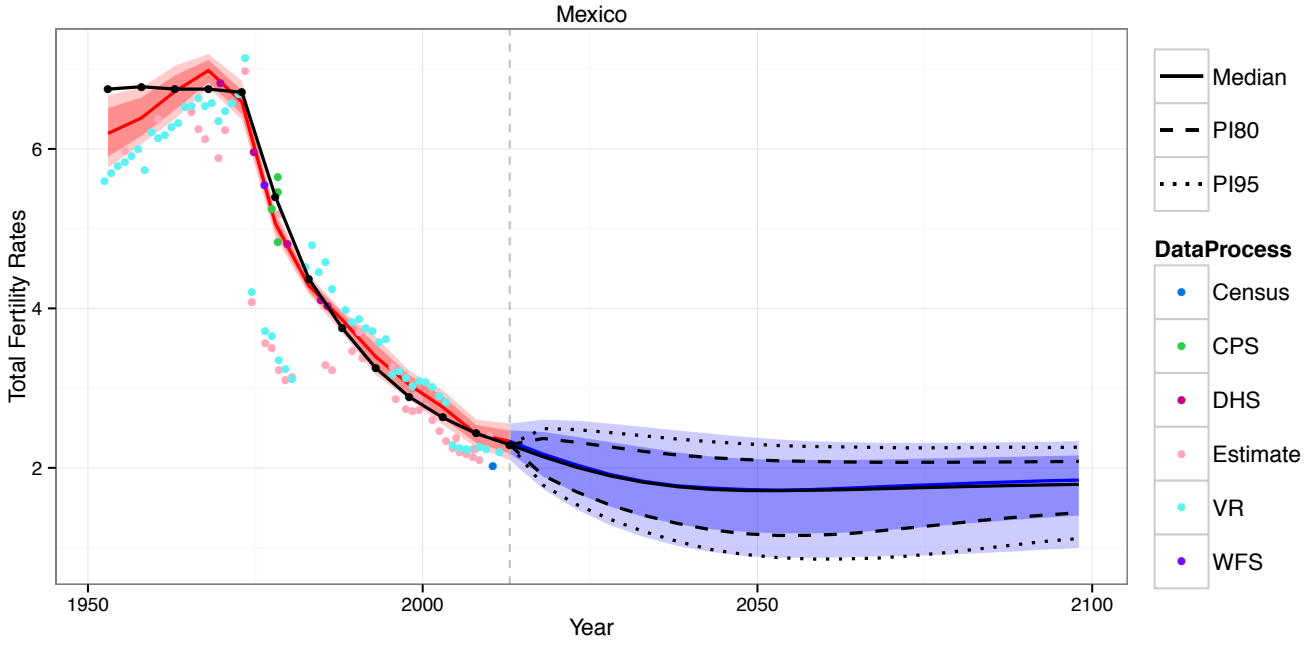

Malta
Total Fertility Rates
1
2
3
4
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

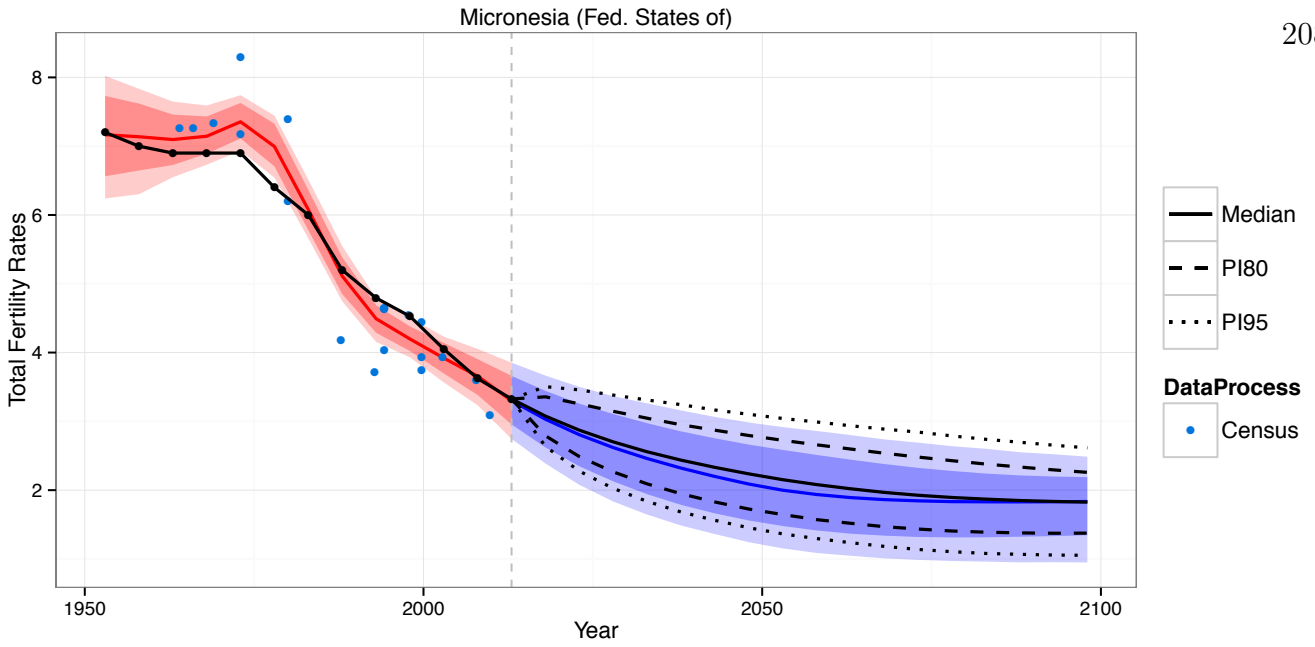

Martinique
Total Fertility Rates
2
4
6
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

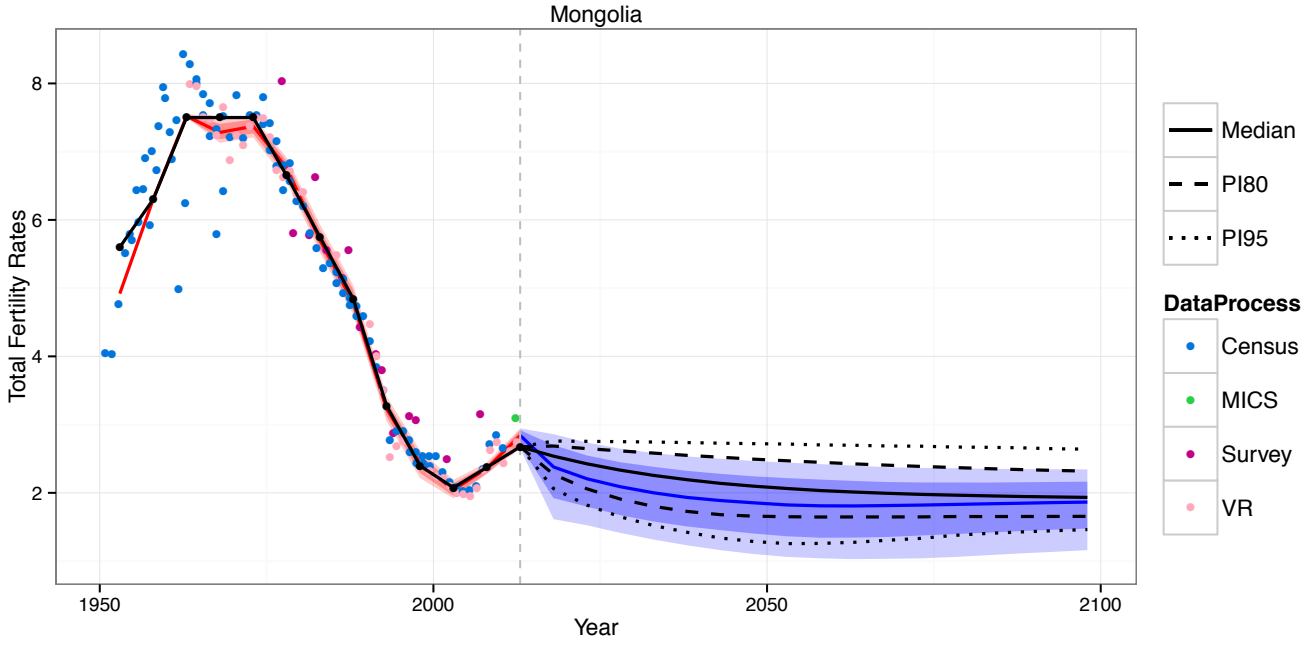

Mauritania
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
MICS
Survey
WFS

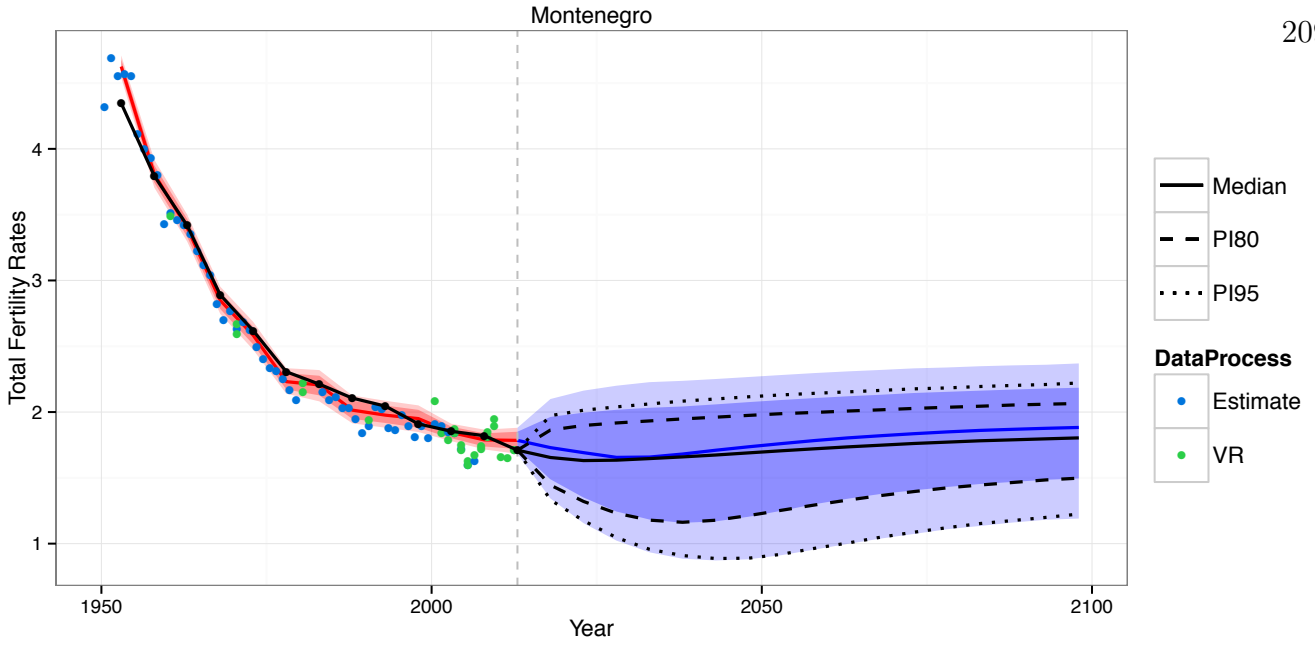

Mauritius
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
Estimate
RHS
VR

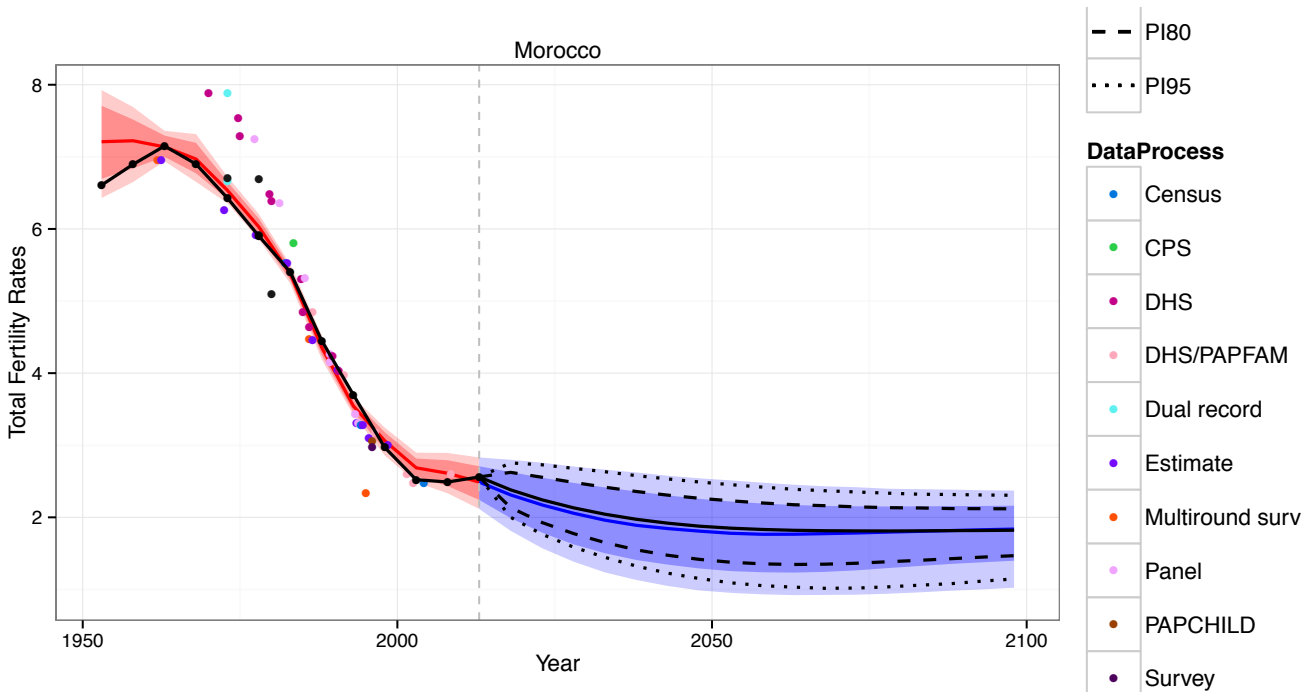

Mayotte
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
VR

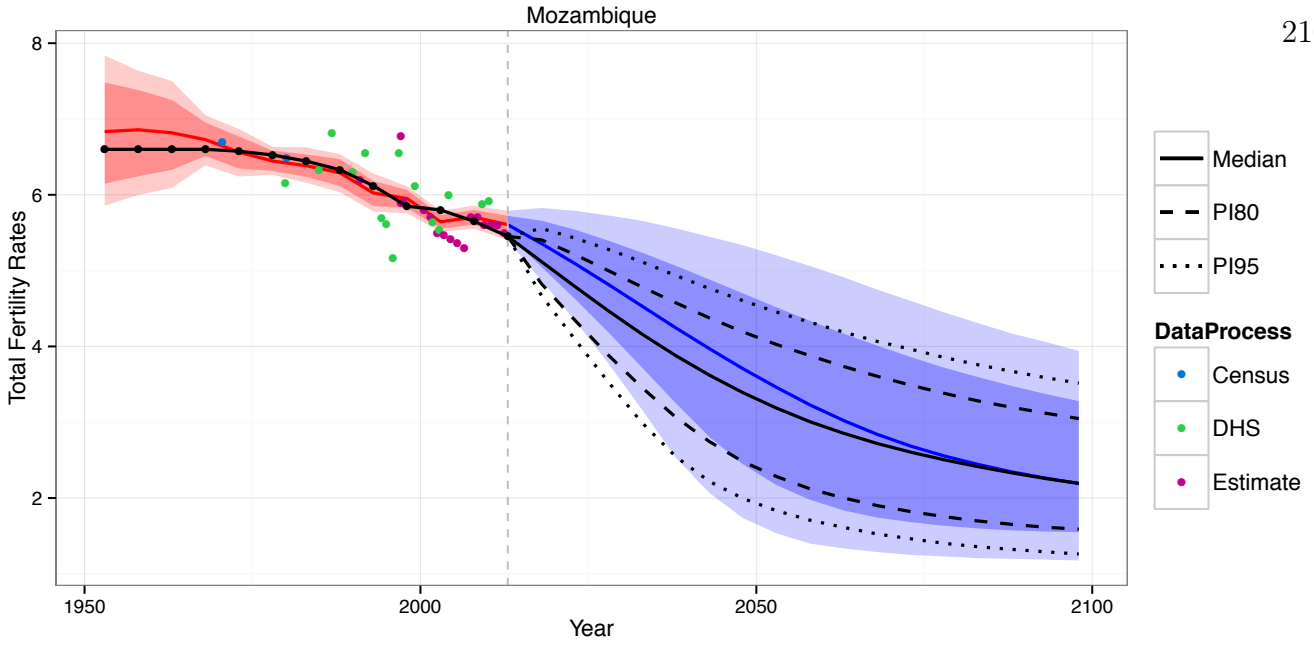

Mexico
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
DHS
Estimate
VR
WFS

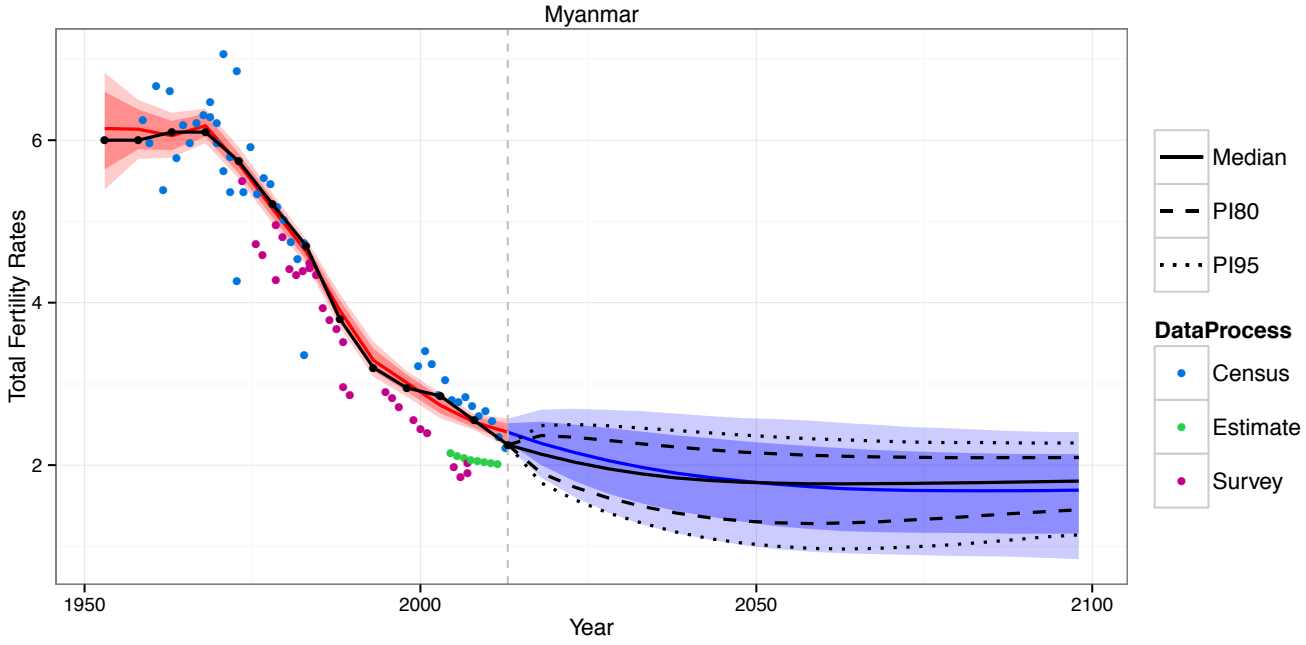

Micronesia (Fed. States of)
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census

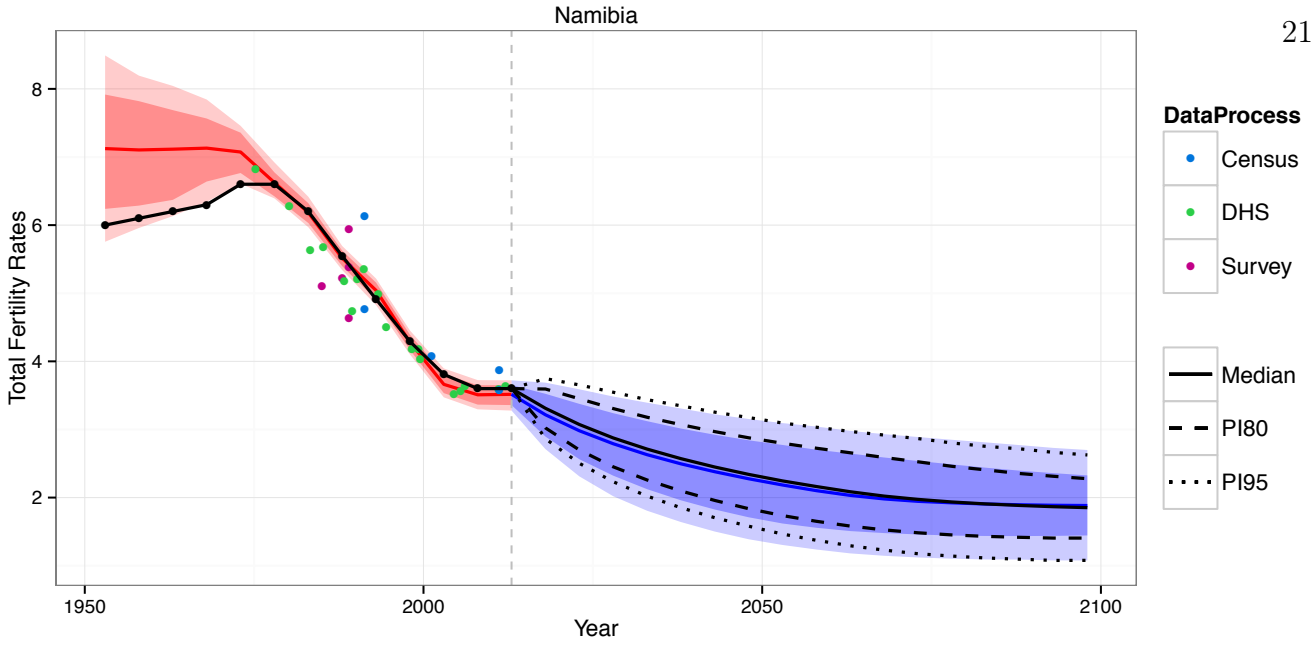

Mongolia
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
MICS
Survey
VR

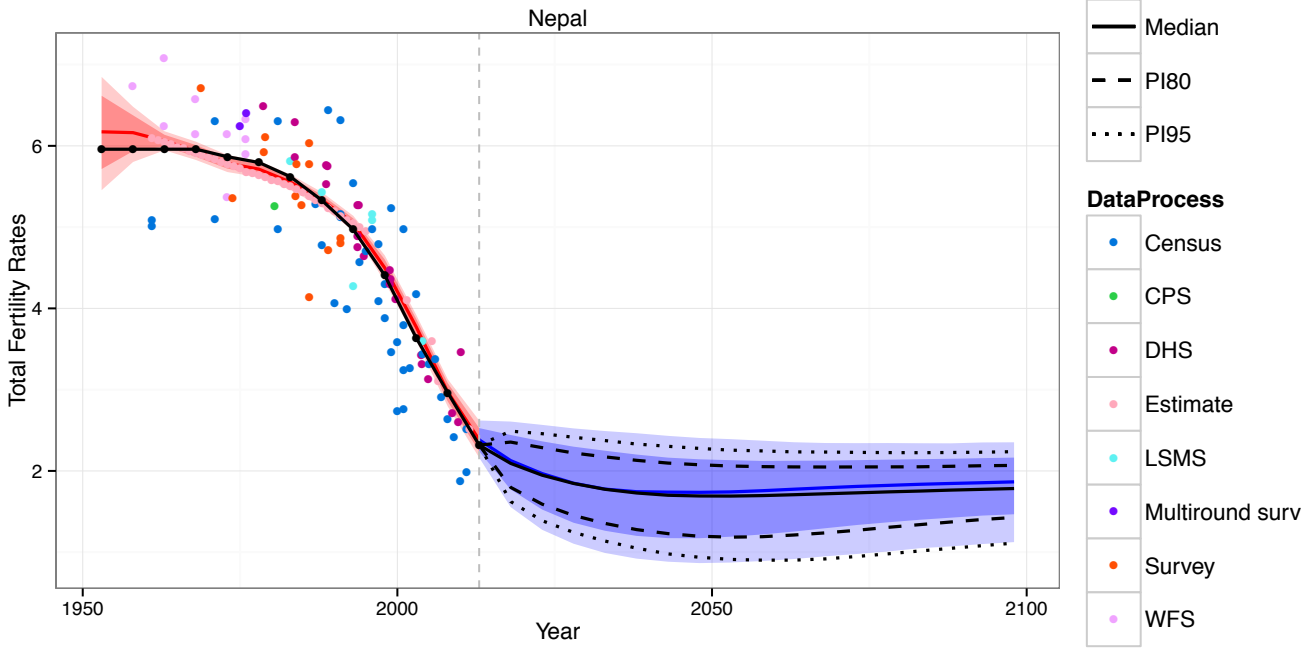

Montenegro
Total Fertility Rates
1
2
3
4
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

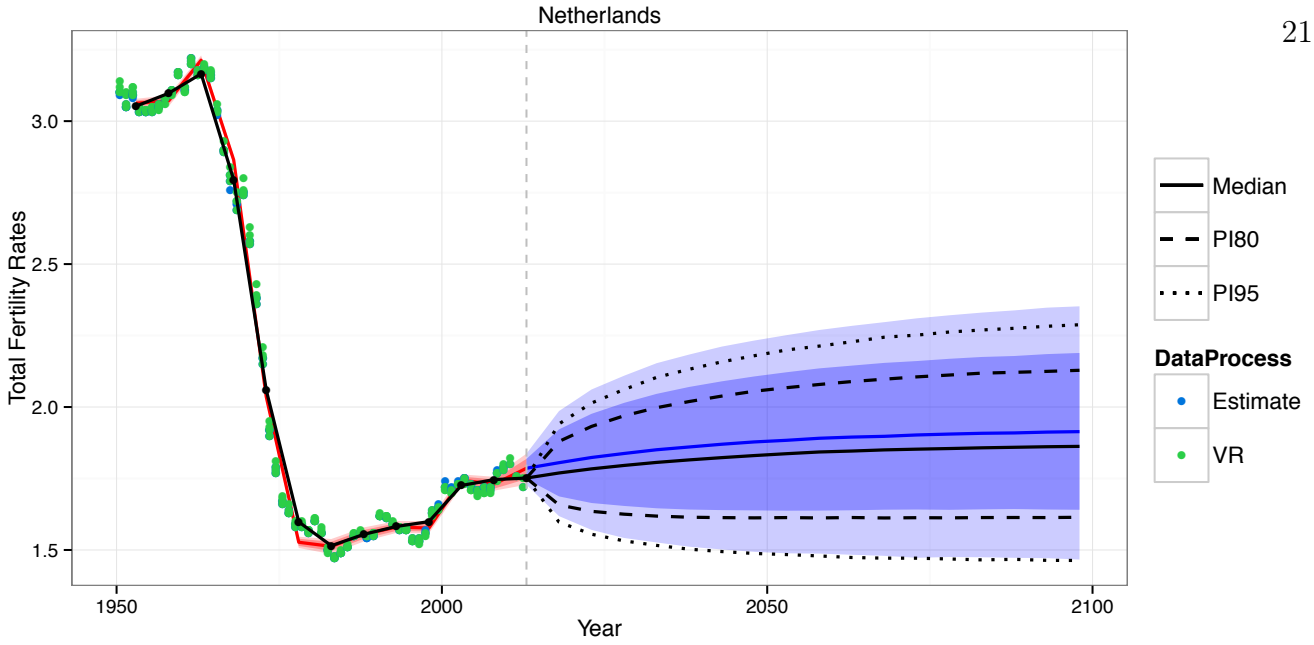

Morocco
Total Fertility Rates
2
4
6
8
1950
2000
2050
2100
Year
PI80
PI95
DataProcess
Census
CPS
DHS
DHS/PAPFAM
Dual record
Estimate
Multiround surv
Panel
PAPCHILD
Survey

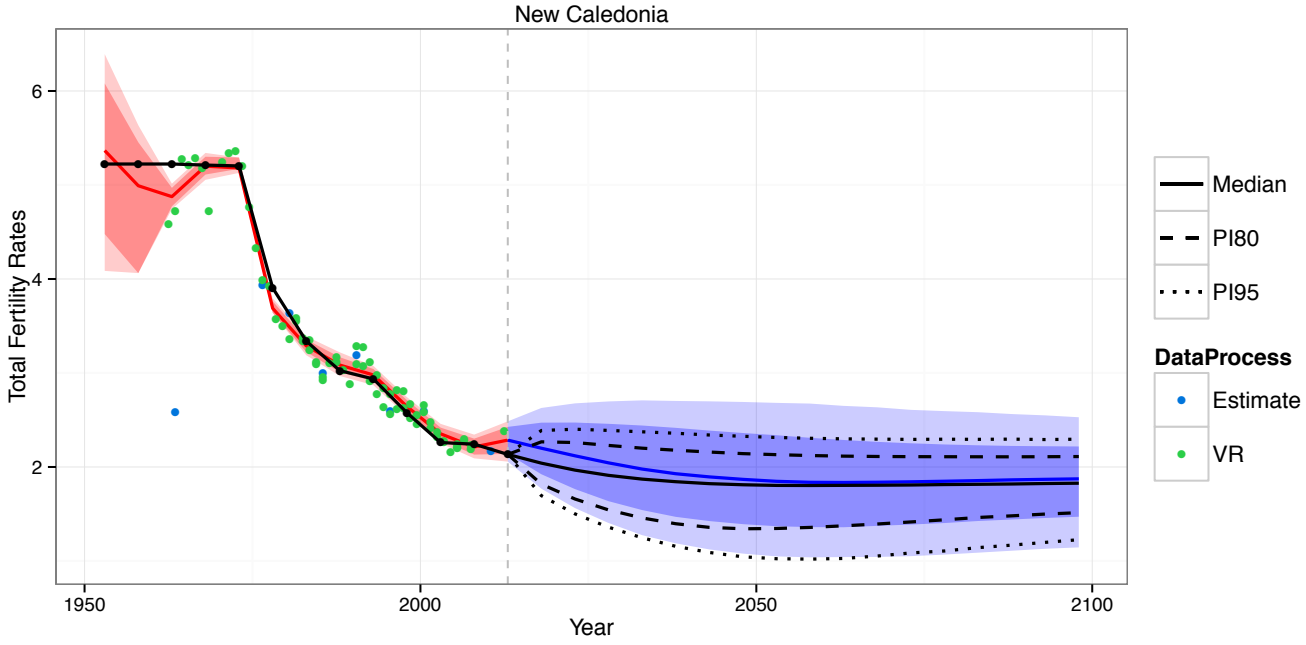

Mozambique
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate

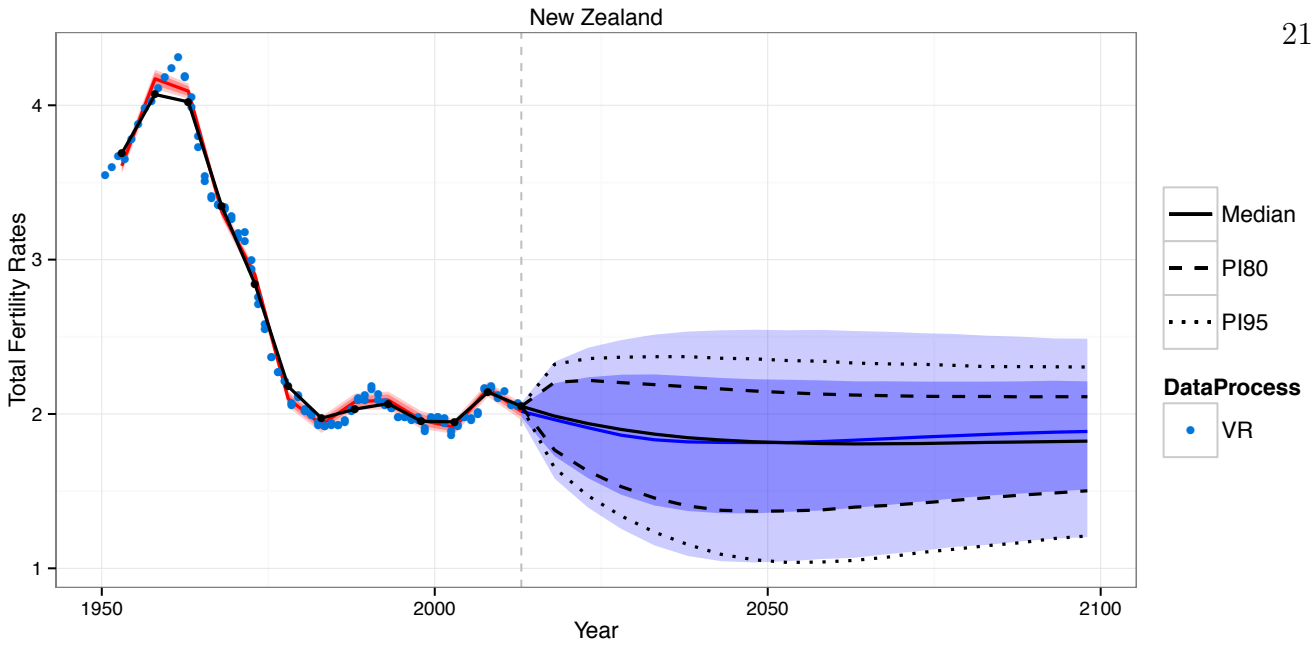

Myanmar
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
Survey

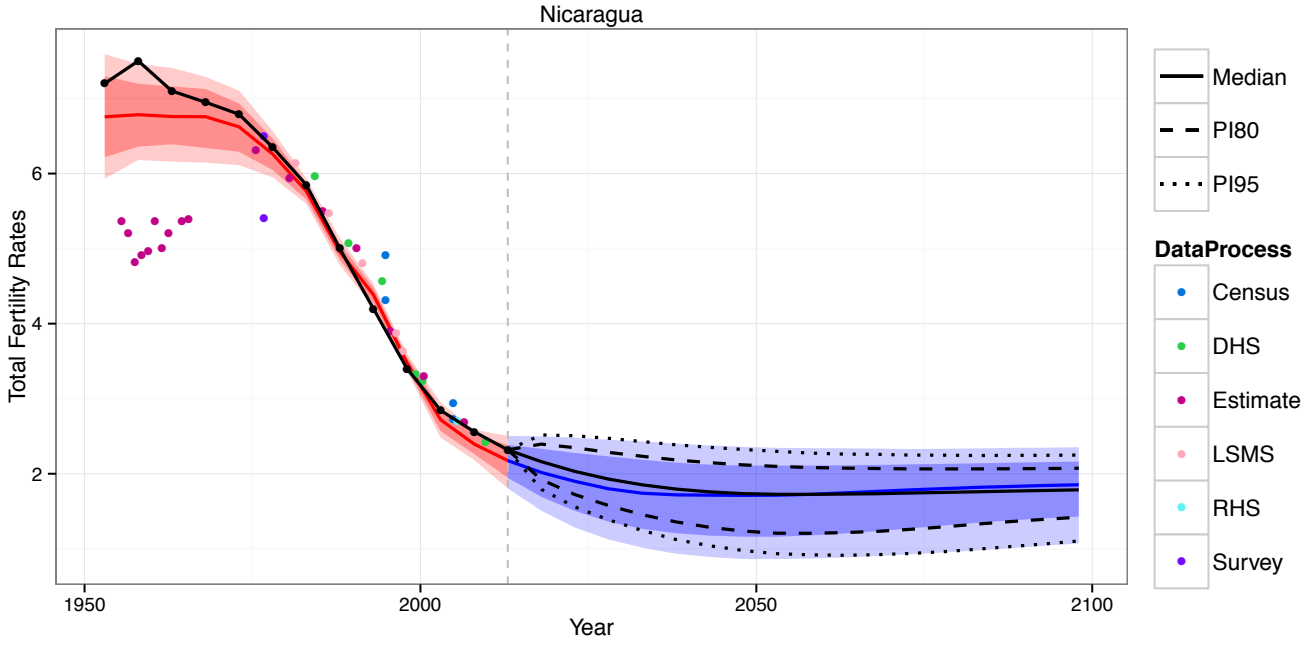

Namibia
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Survey
Median
PI80
PI95

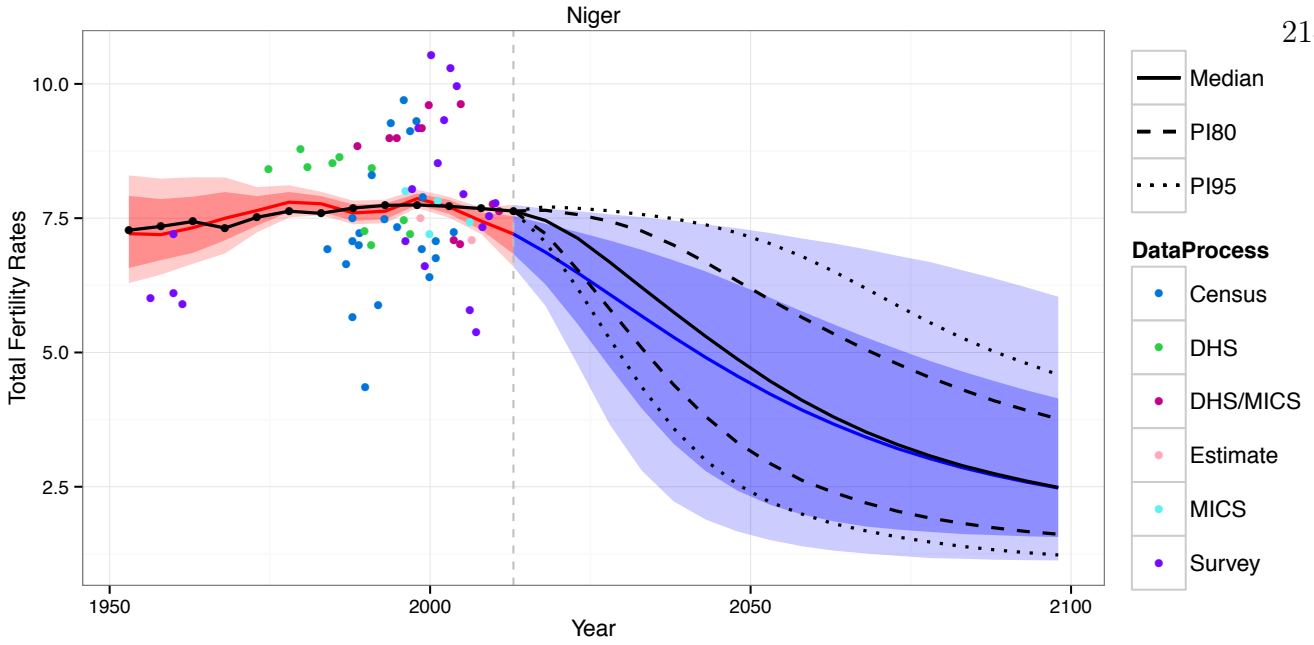

Nepal
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
DHS
Estimate
LSMS
Multiround surv
Survey
WFS

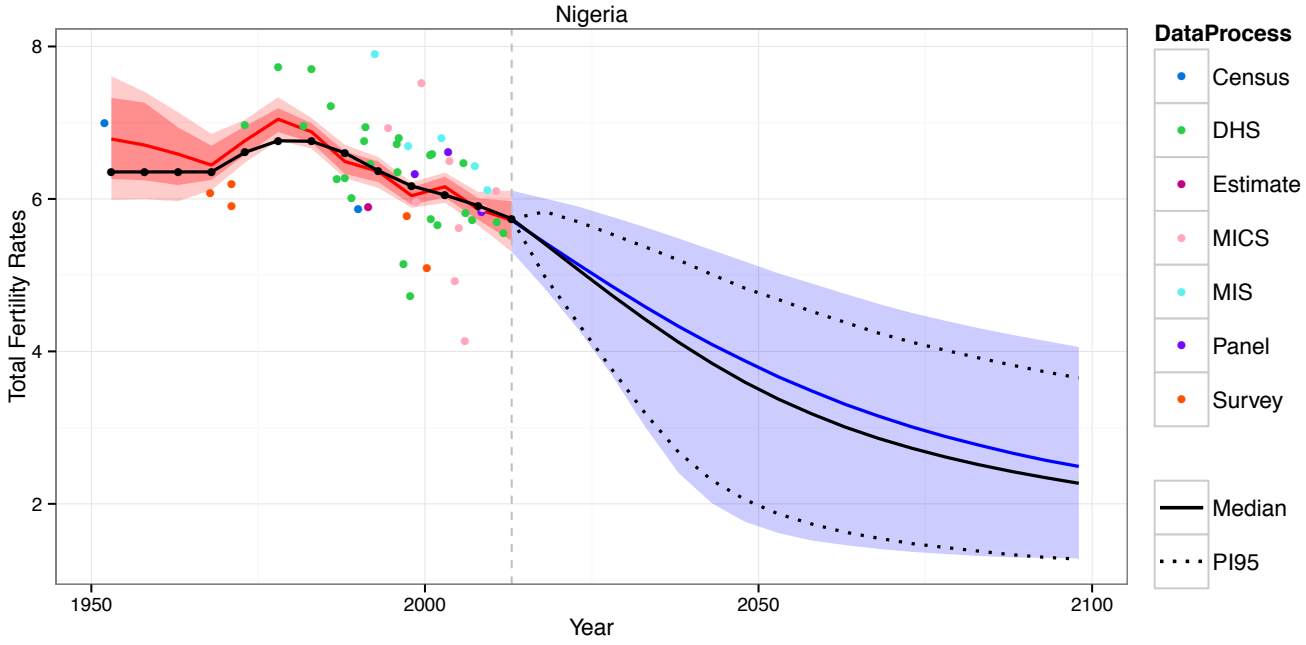

Netherlands
Total Fertility Rates
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

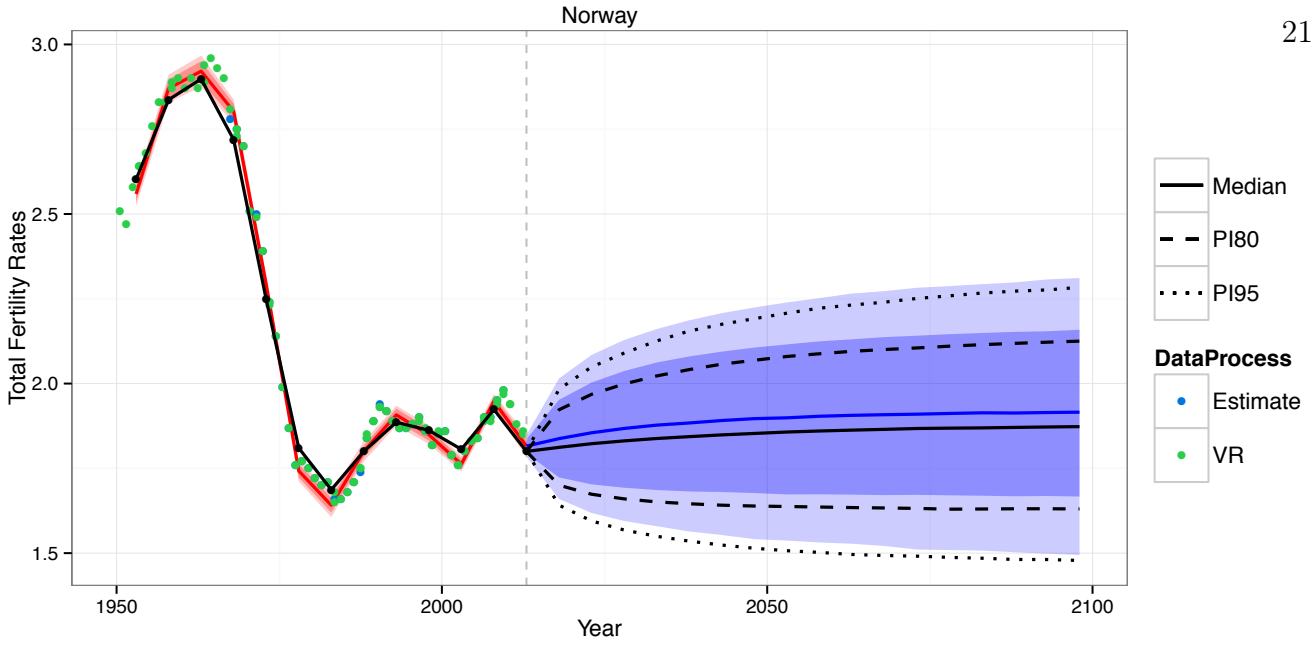

New Caledonia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

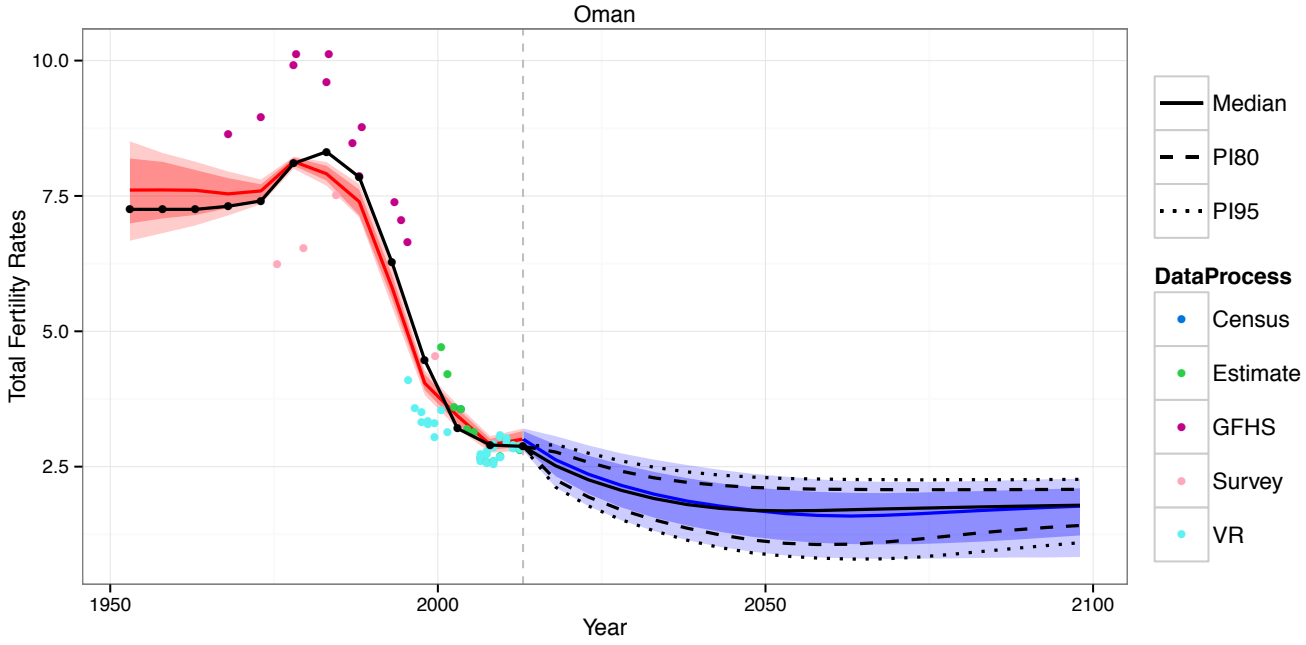

New Zealand
Total Fertility Rates
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
VR

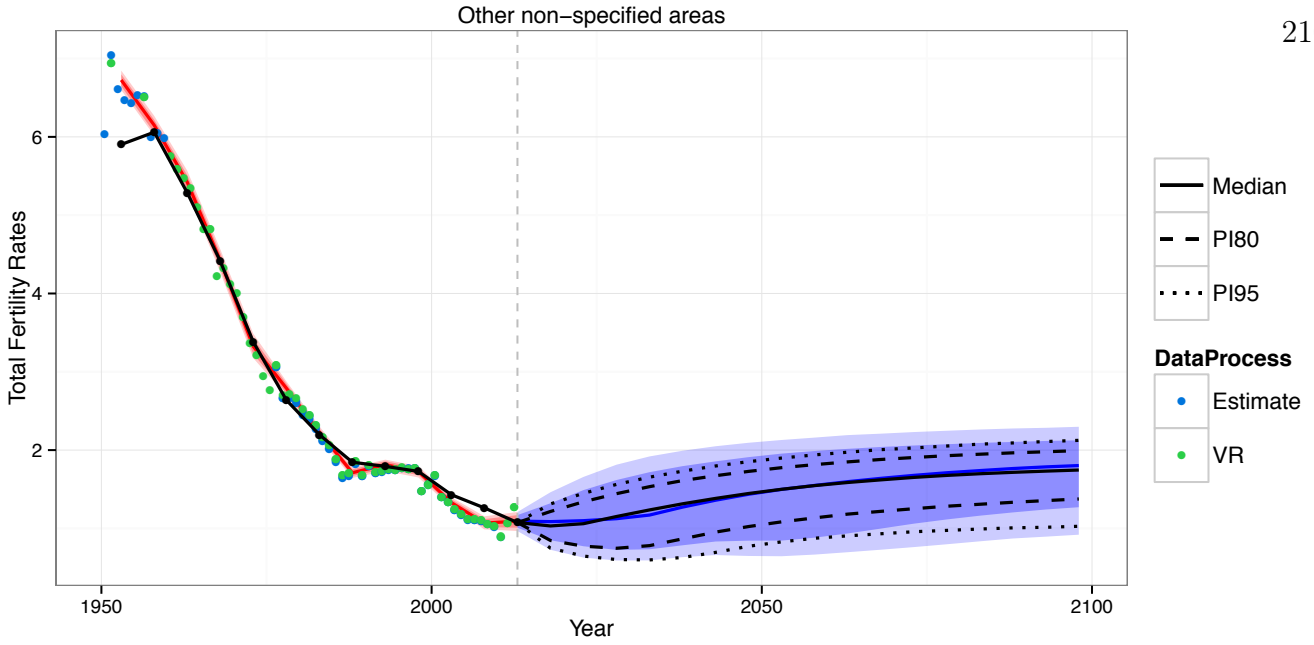

Nicaragua
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
LSMS
RHS
Survey

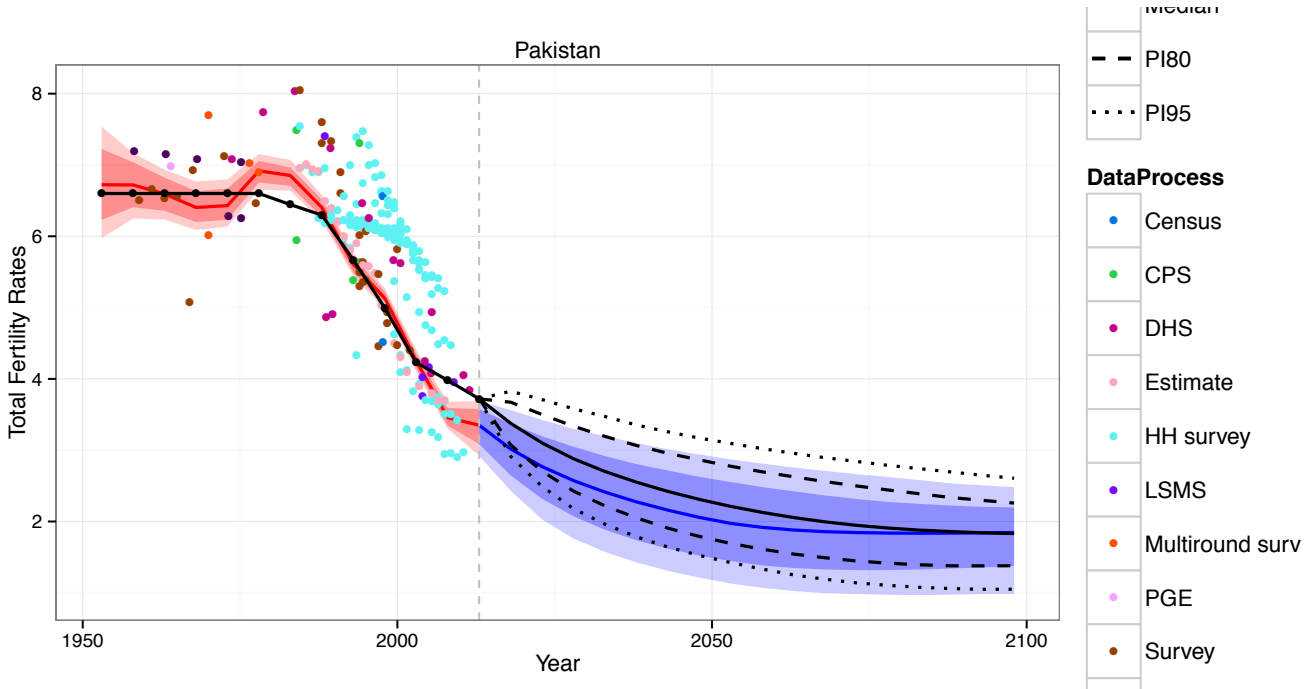

Niger
Total Fertility Rates
10.0
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
DHS/MICS
Estimate
MICS
Survey

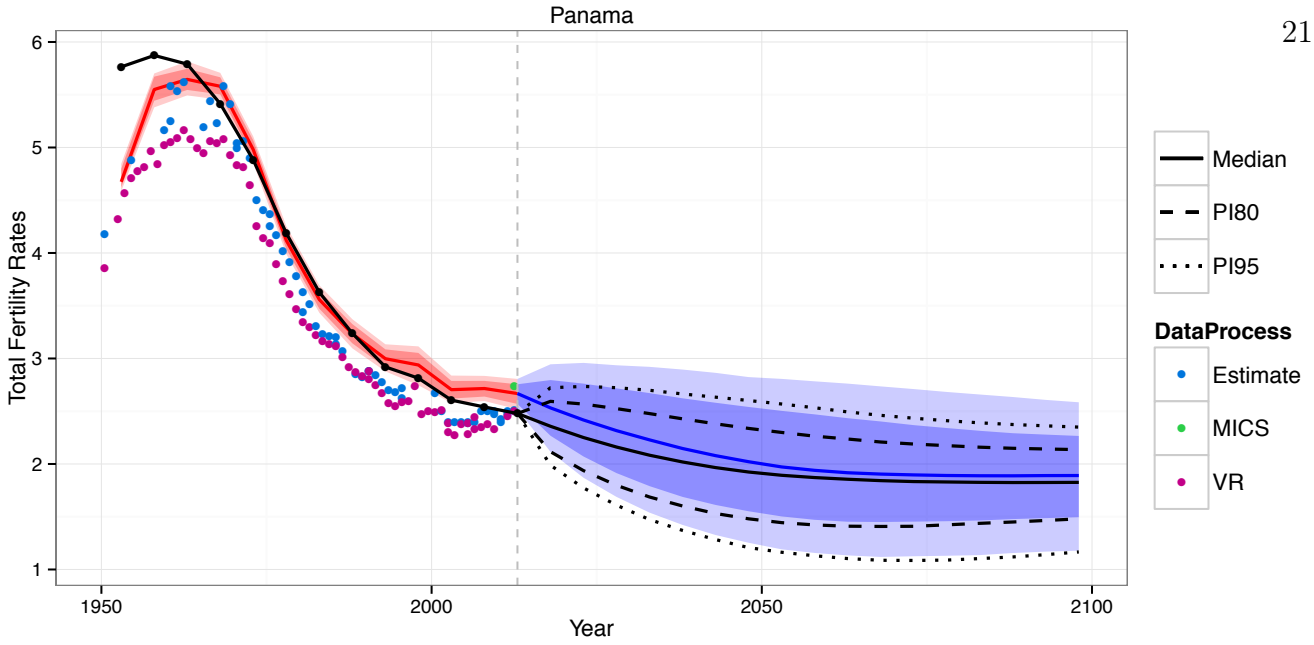

Nigeria
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Estimate
MICS
MIS
Panel
Survey
Median
PI95

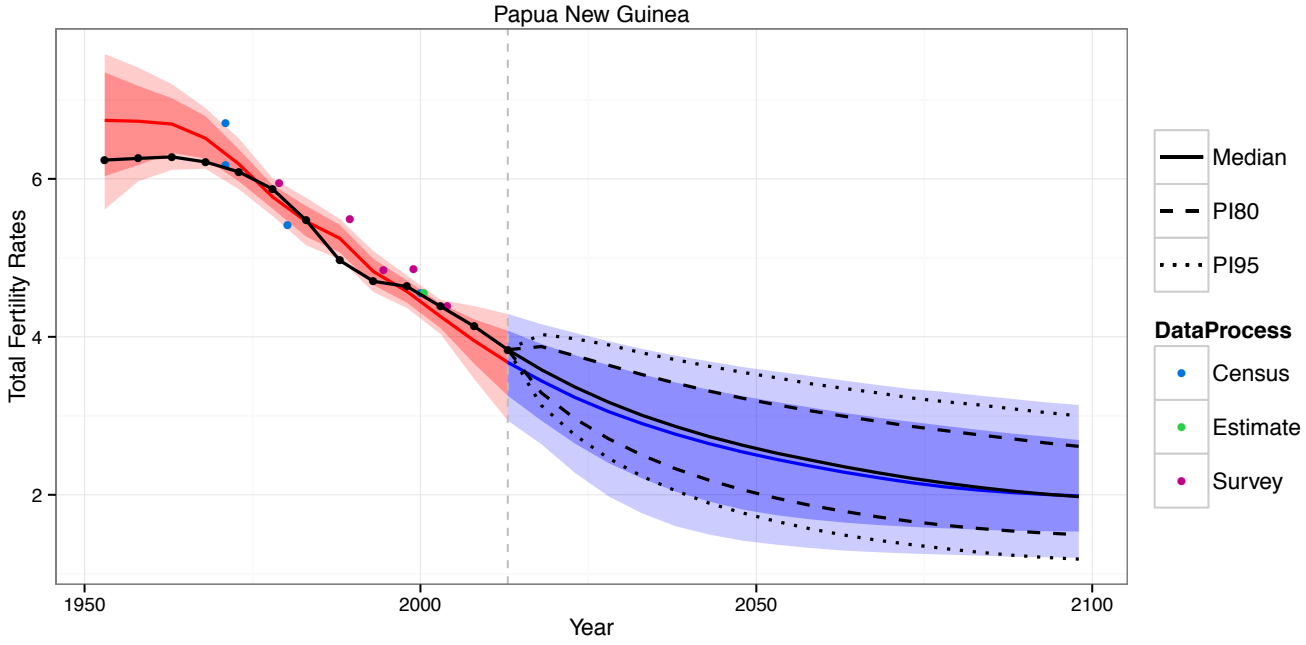

Norway
Total Fertility Rates
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

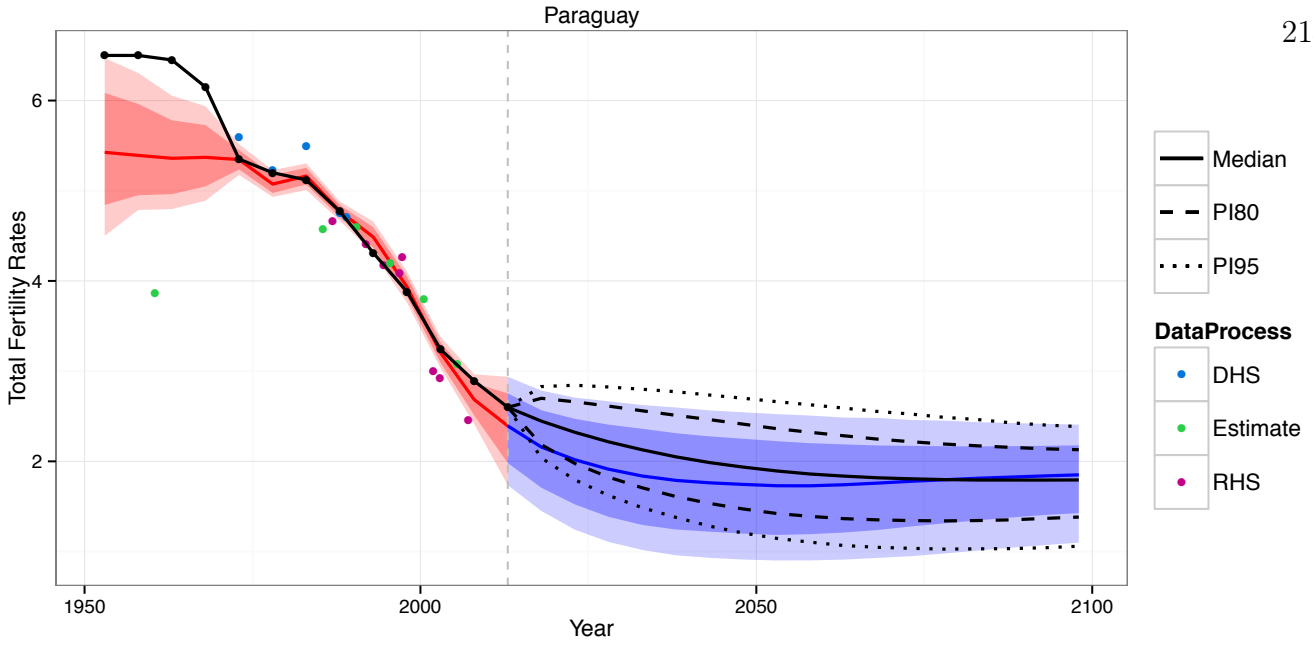

Oman
Total Fertility Rates
10.0
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
GFHS
Survey
VR

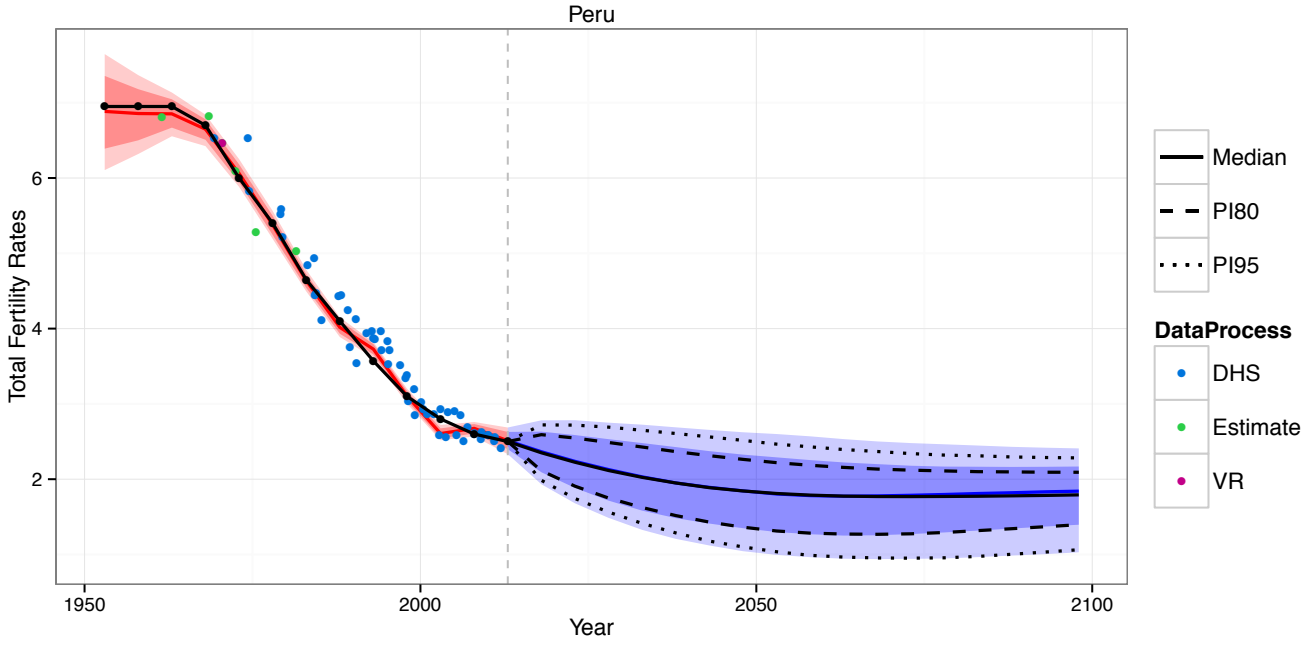

Other non−specified areas
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

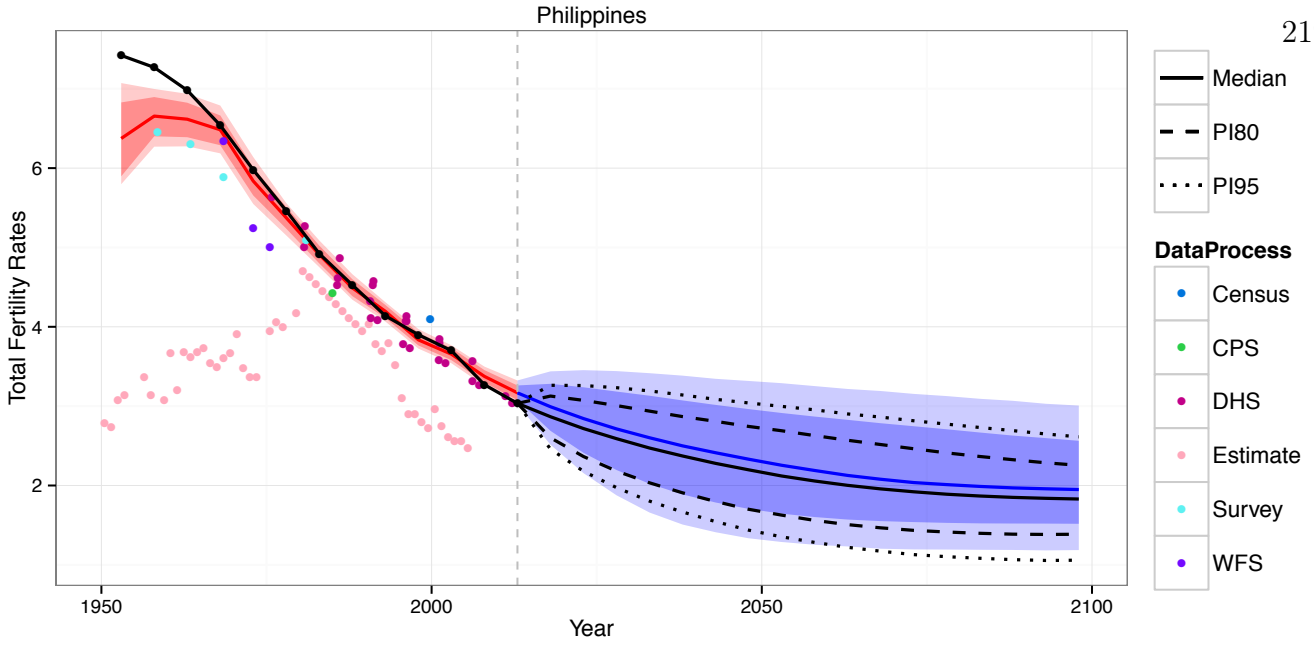

Pakistan
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
PI80
PI95
DataProcess
Census
CPS
DHS
Estimate
HH survey
LSMS
Multiround surv
PGE
Survey

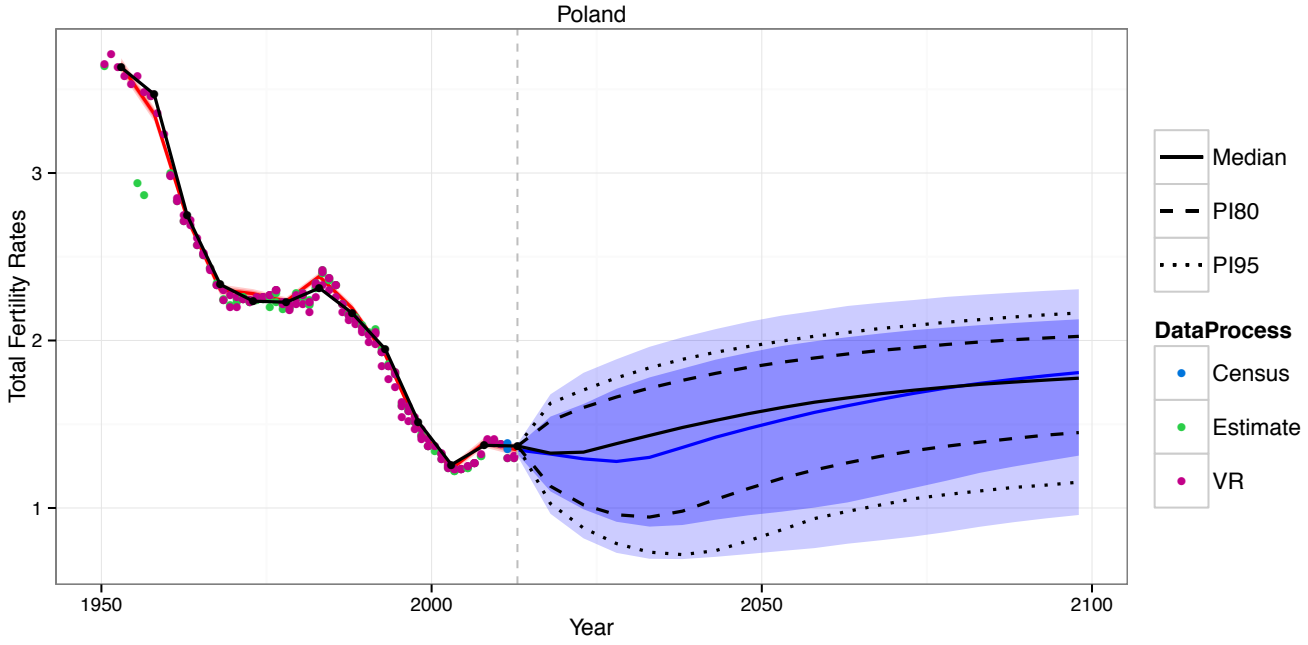

Panama
Total Fertility Rates
Year
1950
2000
2050
2100
1
2
3
4
5
6
Median
PI80
PI95
DataProcess
Estimate
MICS
VR

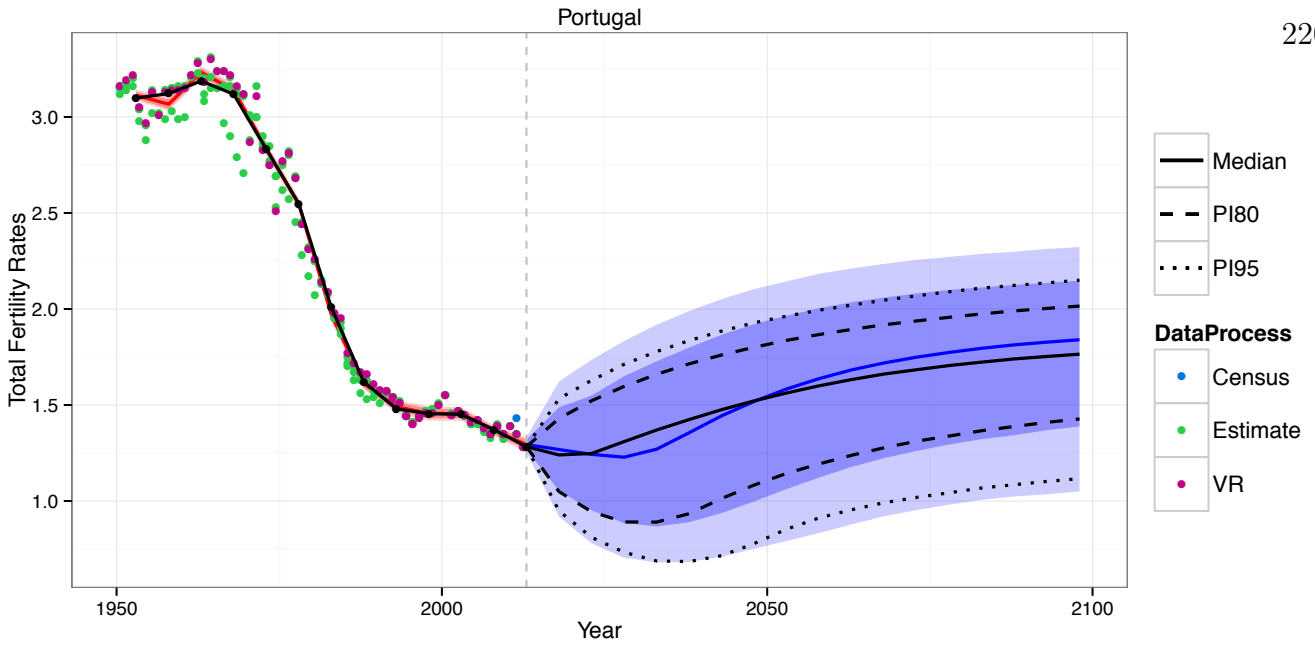

Papua New Guinea
Total Fertility Rates
Year
1950
2000
2050
2100
2
4
6
Median
PI80
PI95
DataProcess
Census
Estimate
Survey

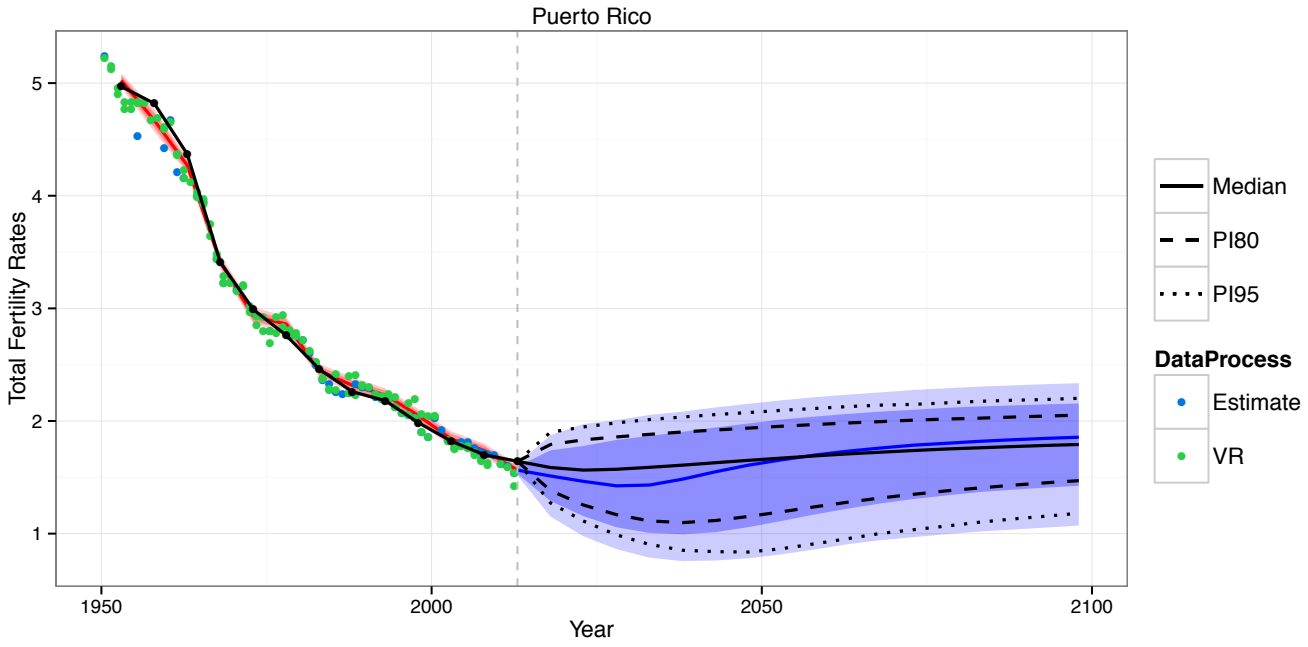

Paraguay
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
RHS

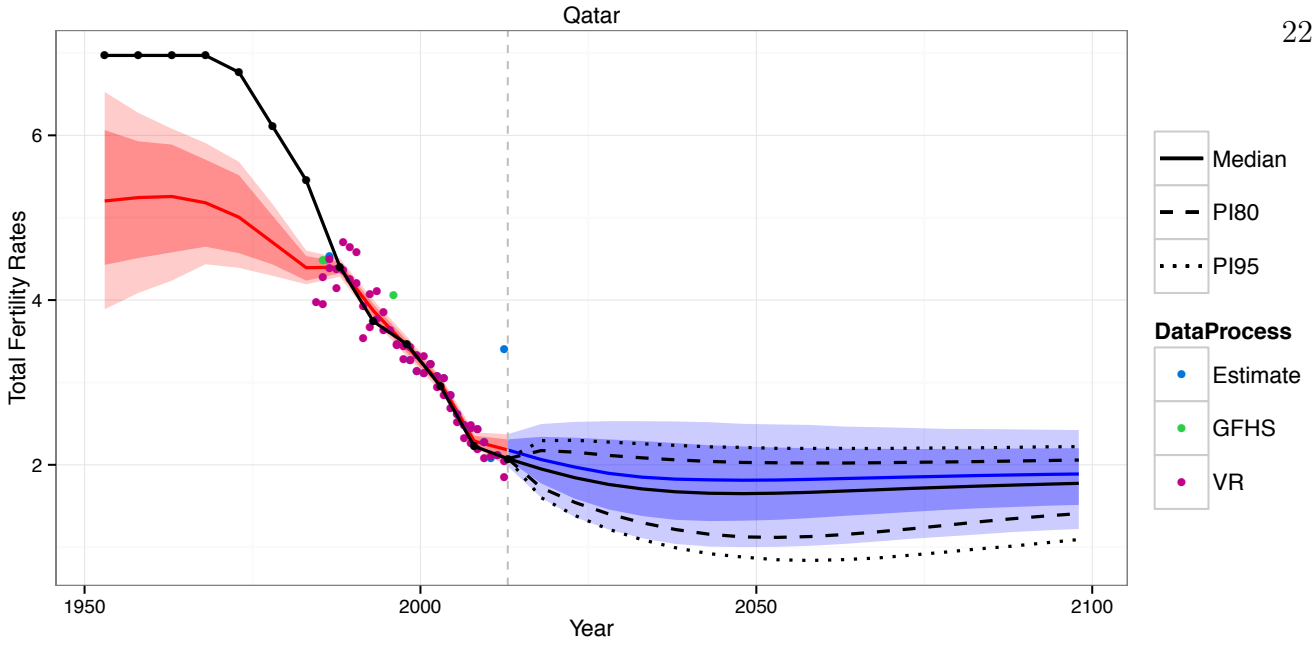

Peru
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
VR

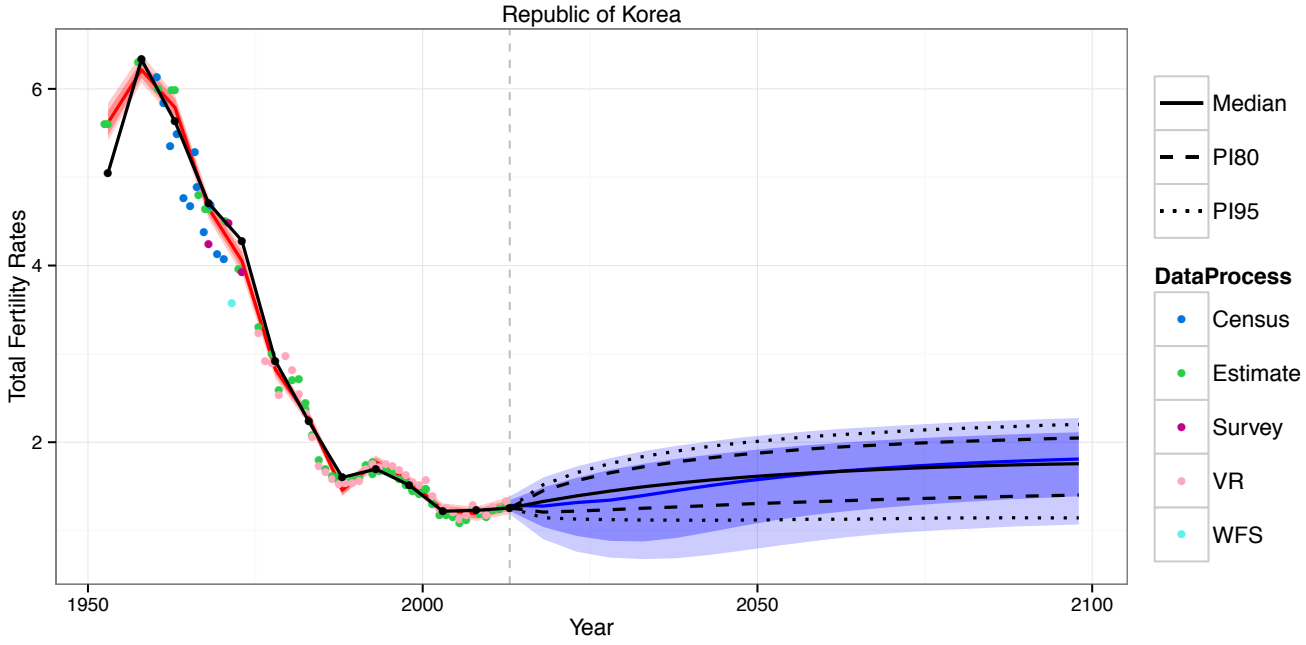

Philippines
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
DHS
Estimate
Survey
WFS

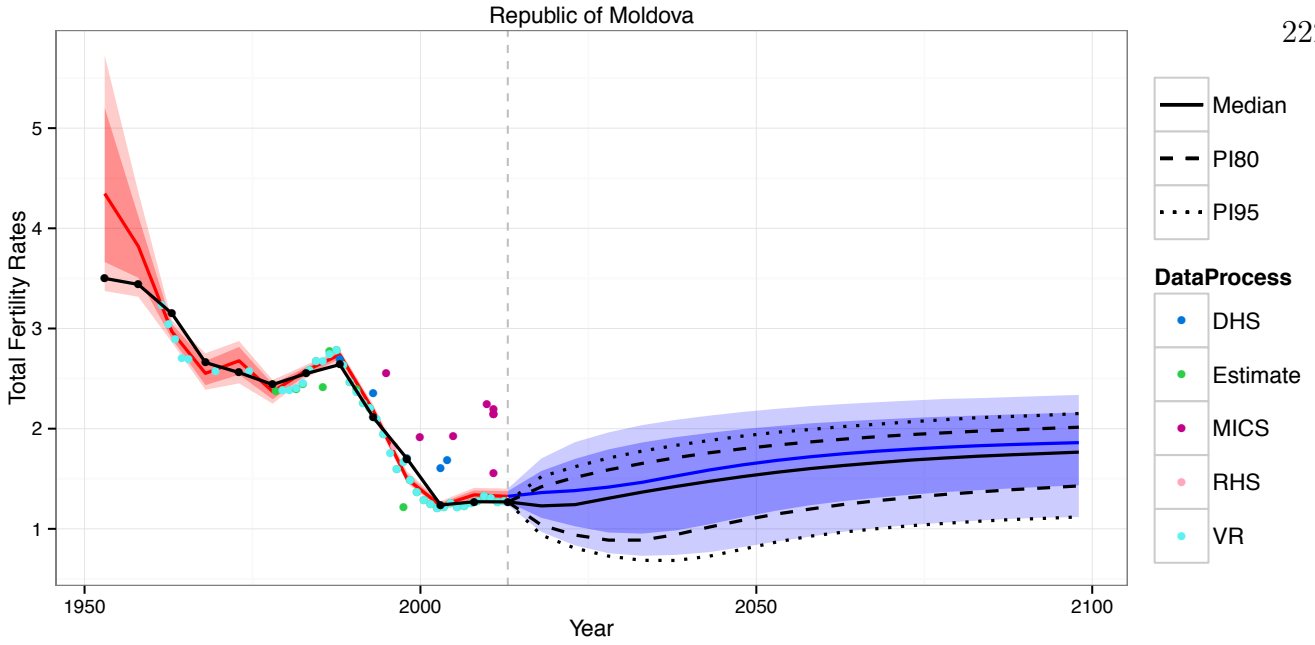

Poland
Total Fertility Rates
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

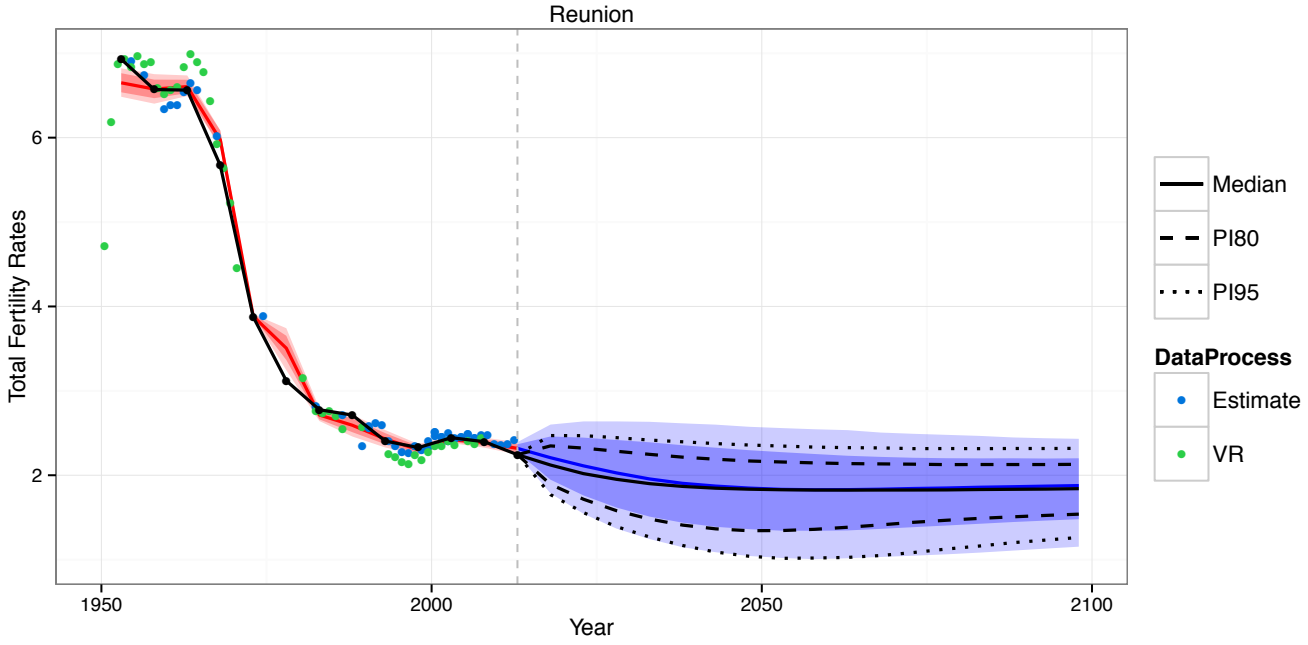

Portugal
Total Fertility Rates
3.0
2.5
2.0
1.5
1.0
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

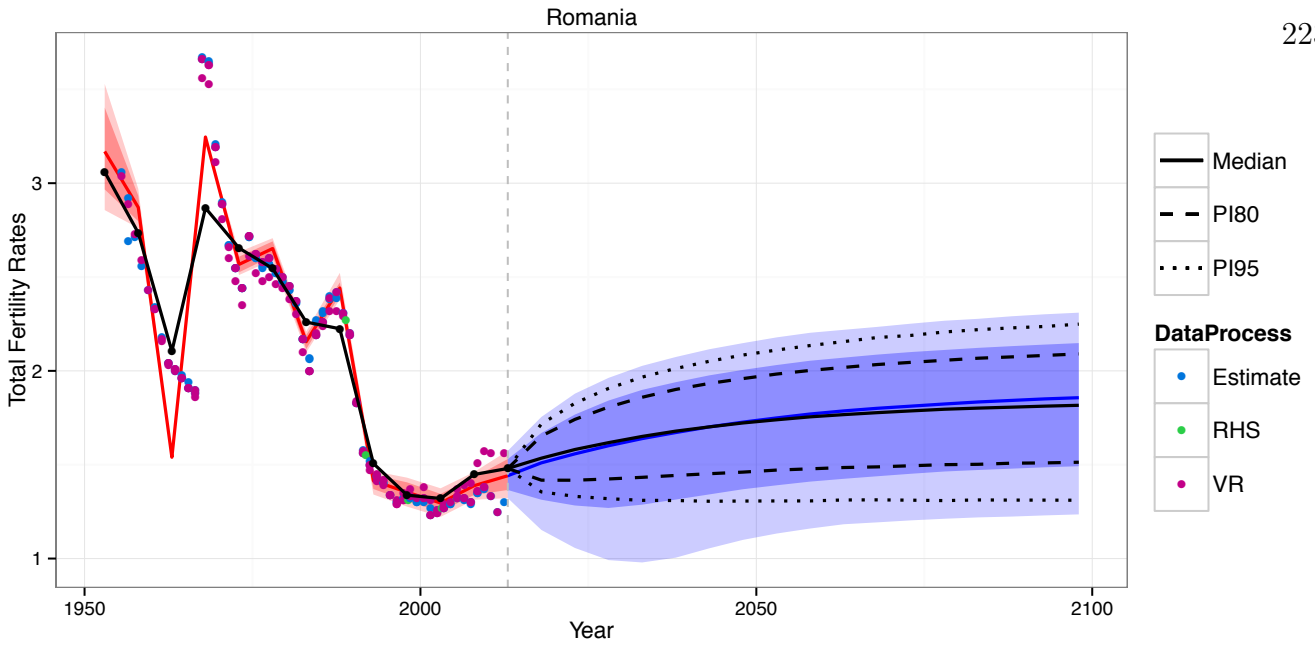

Puerto Rico
Total Fertility Rates
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

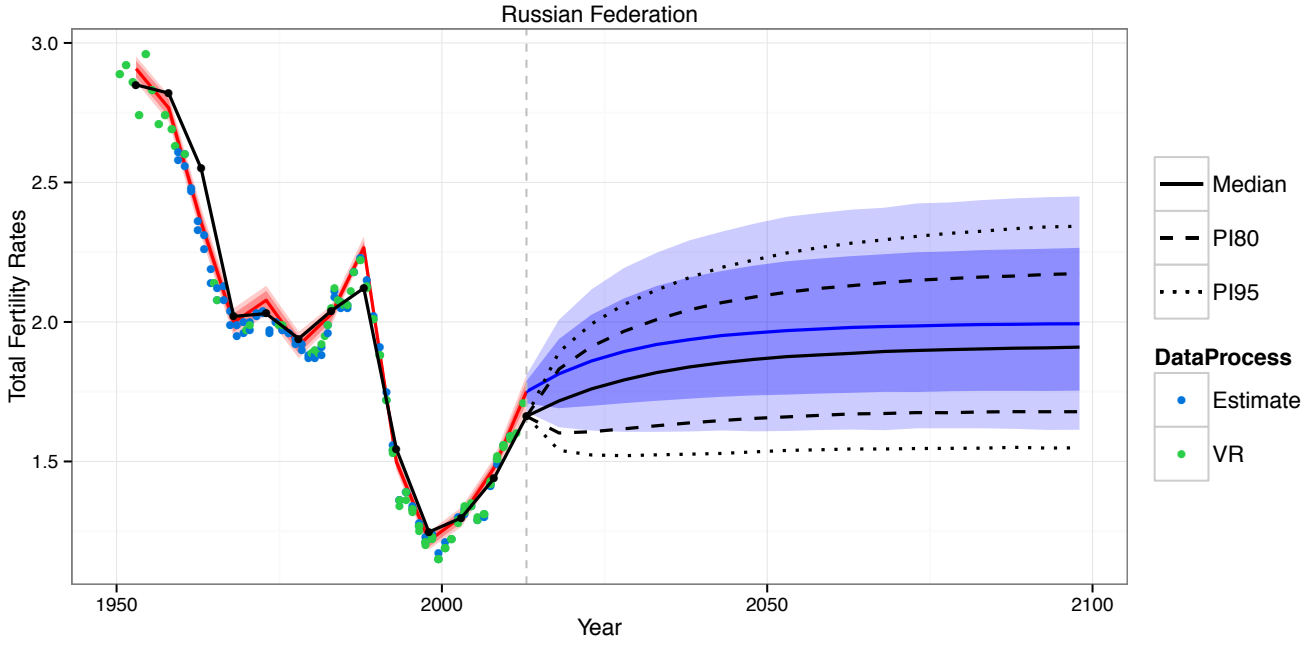

Qatar
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
GFHS
VR

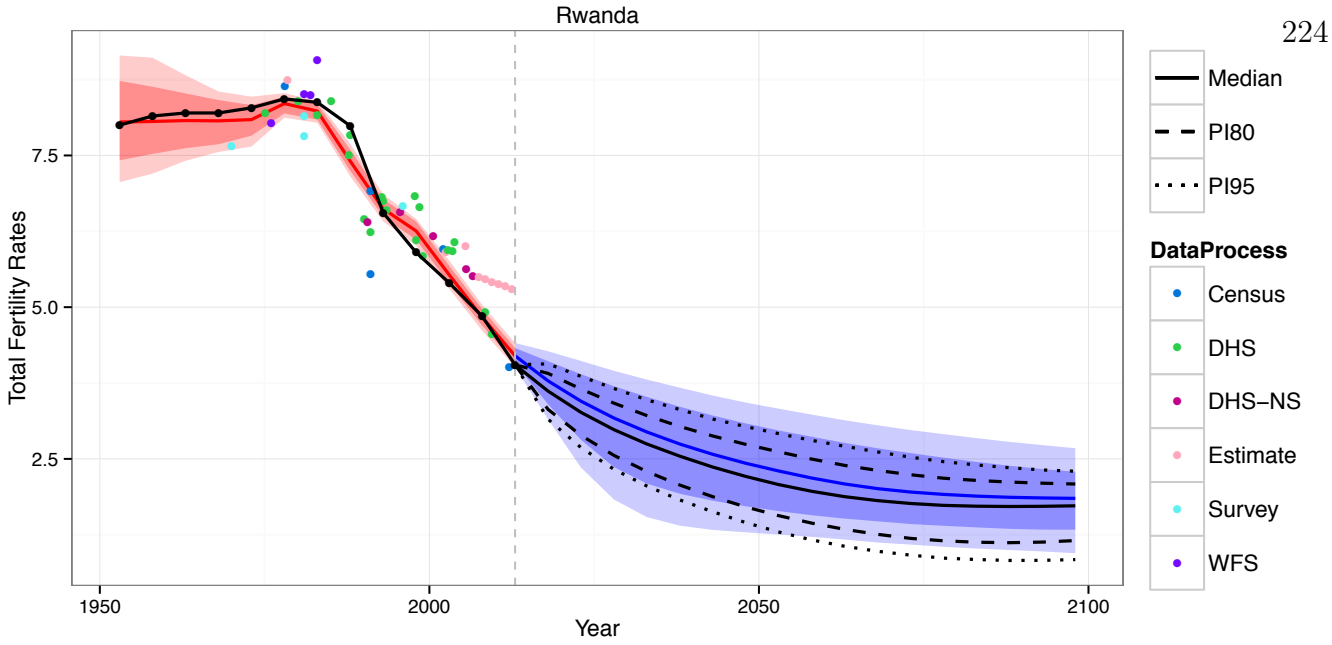

Republic of Korea
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
Survey
VR
WFS

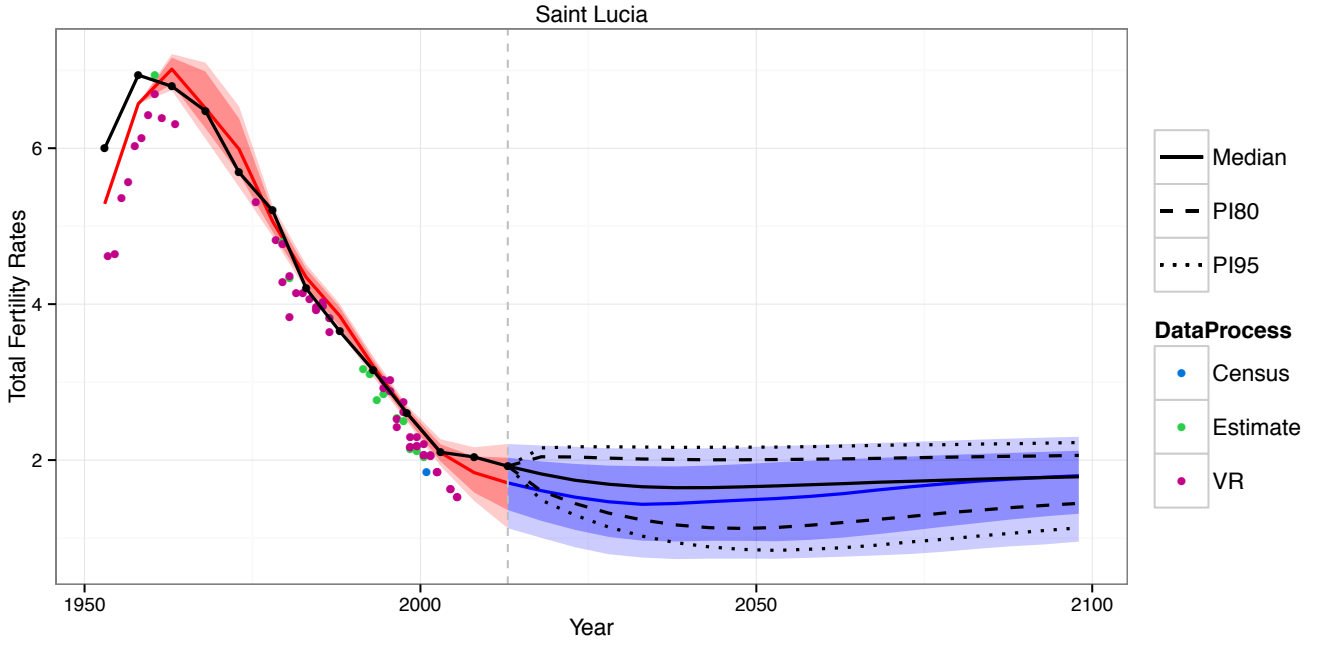

Republic of Moldova
Total Fertility Rates
Year
1950
2000
2050
2100
1
2
3
4
5
Median
PI80
PI95
DataProcess
DHS
Estimate
MICS
RHS
VR

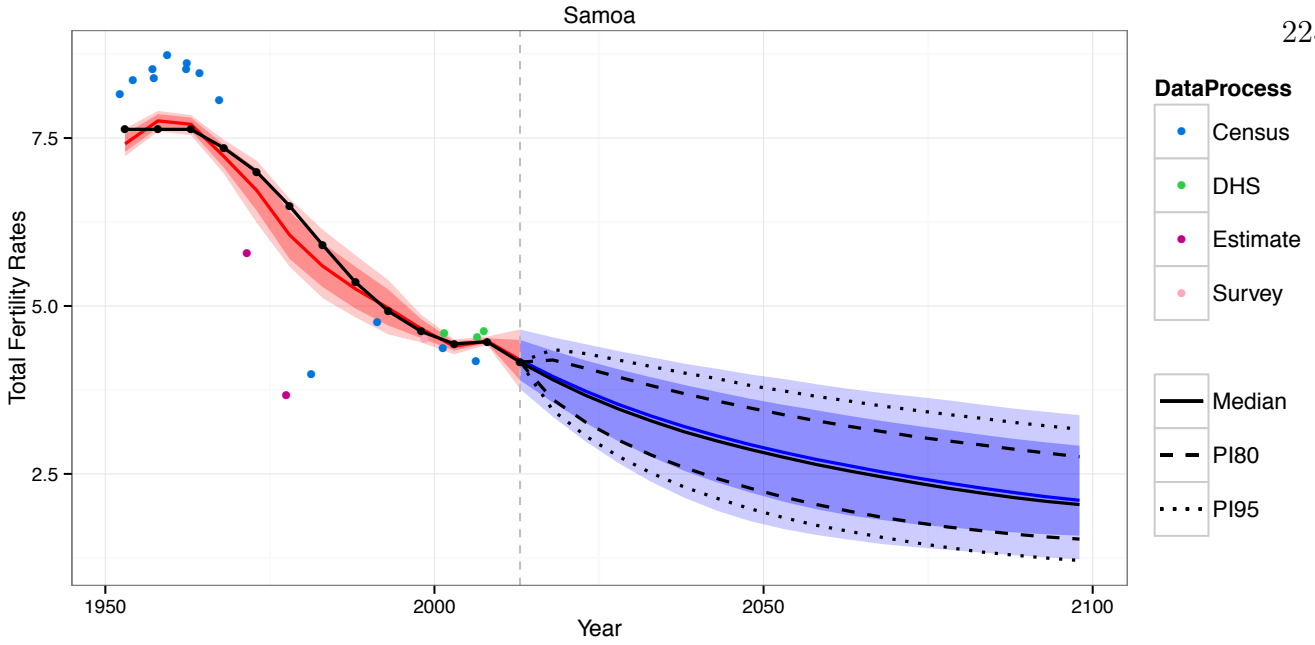

Reunion
Total Fertility Rates
Year
1950
2000
2050
2100
2
4
6
Median
PI80
PI95
DataProcess
Estimate
VR

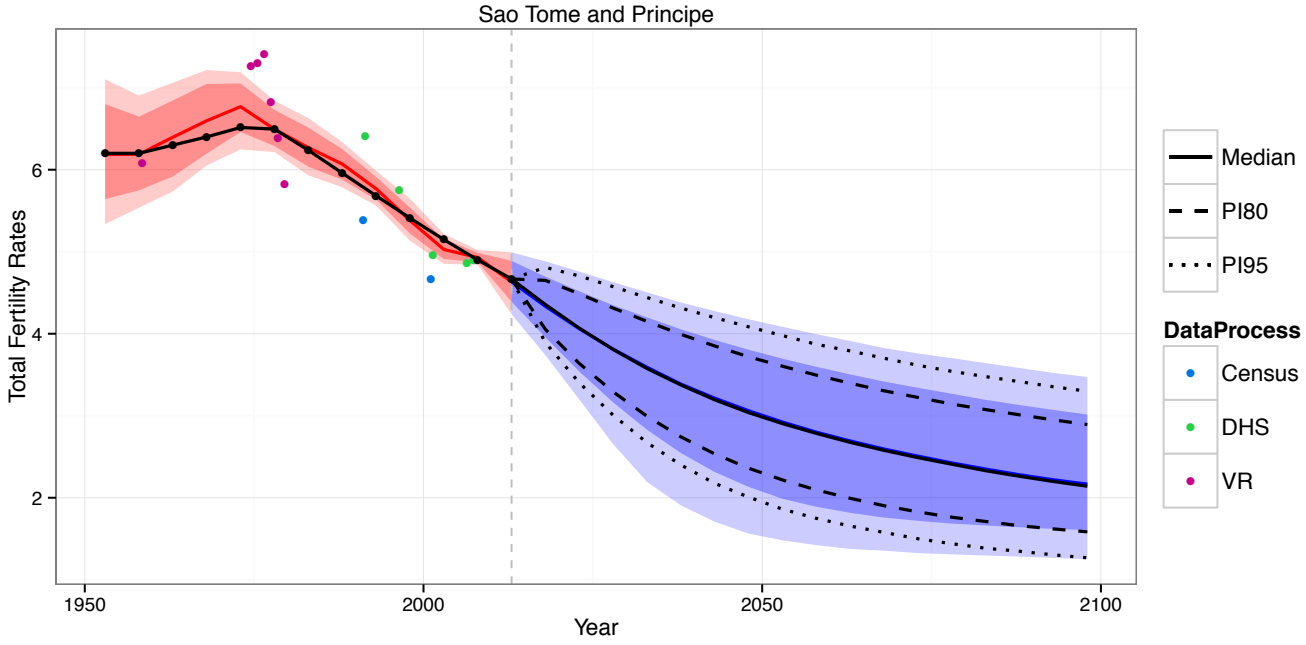

Romania
Total Fertility Rates
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
RHS
VR

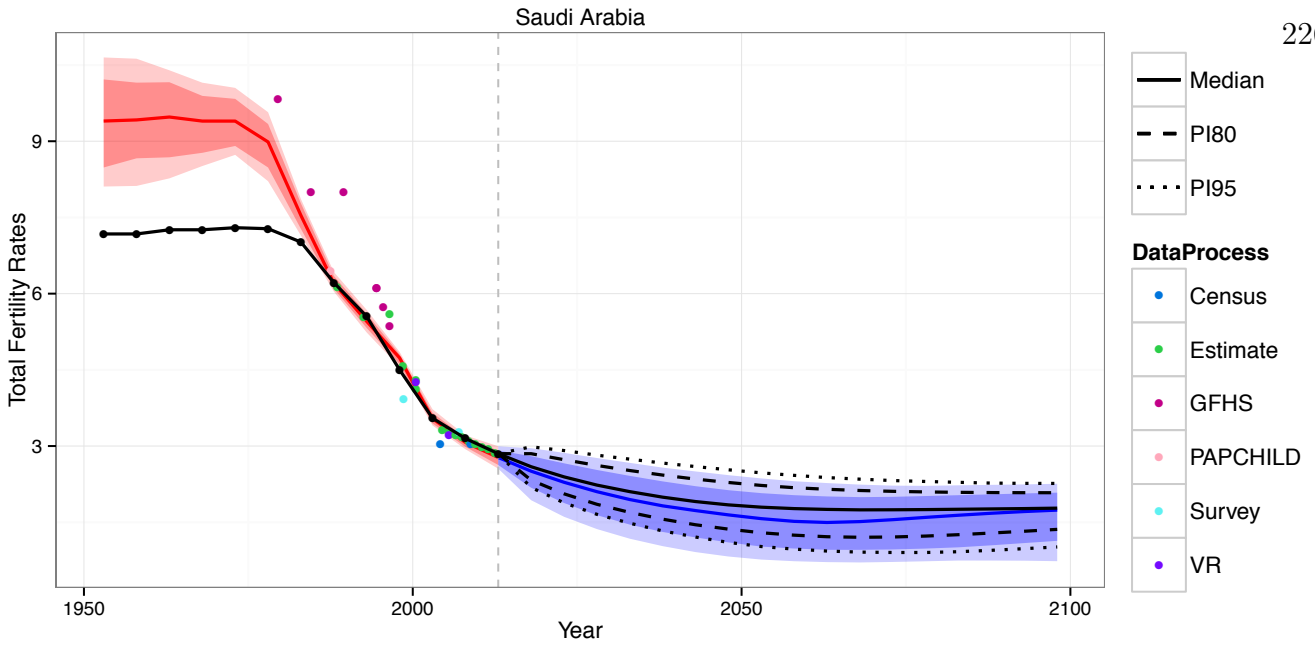

Russian Federation
Total Fertility Rates
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

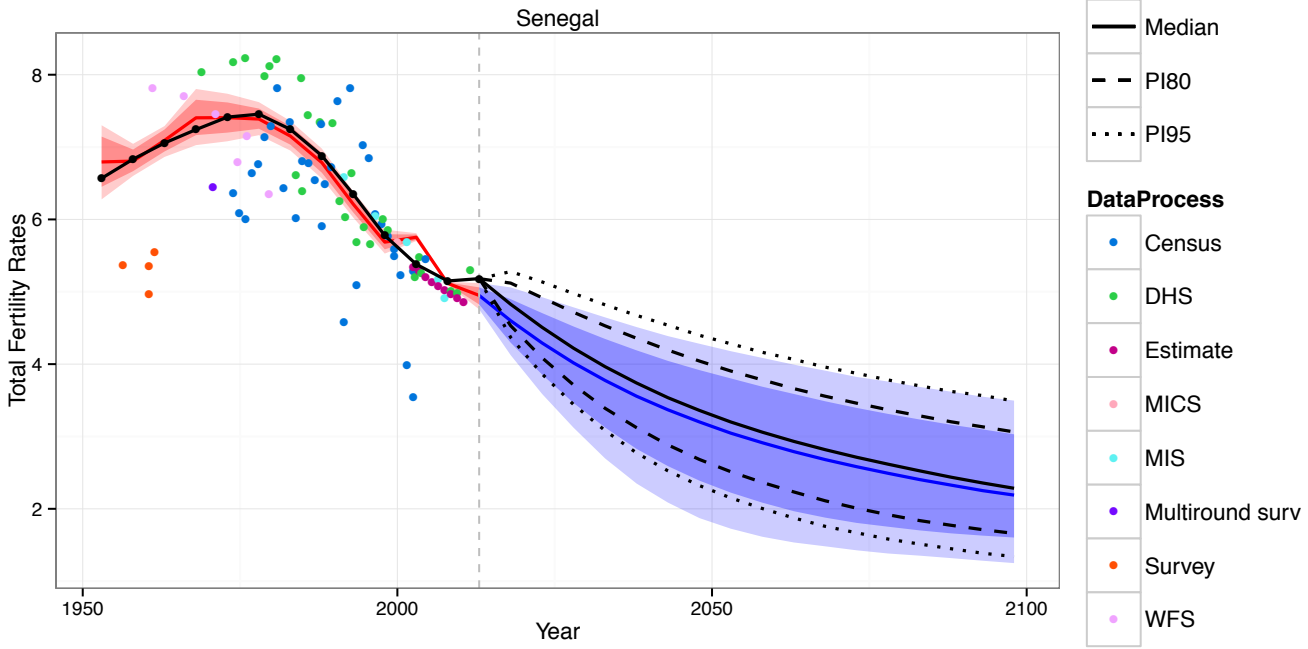

Rwanda
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
DHS−NS
Estimate
Survey
WFS

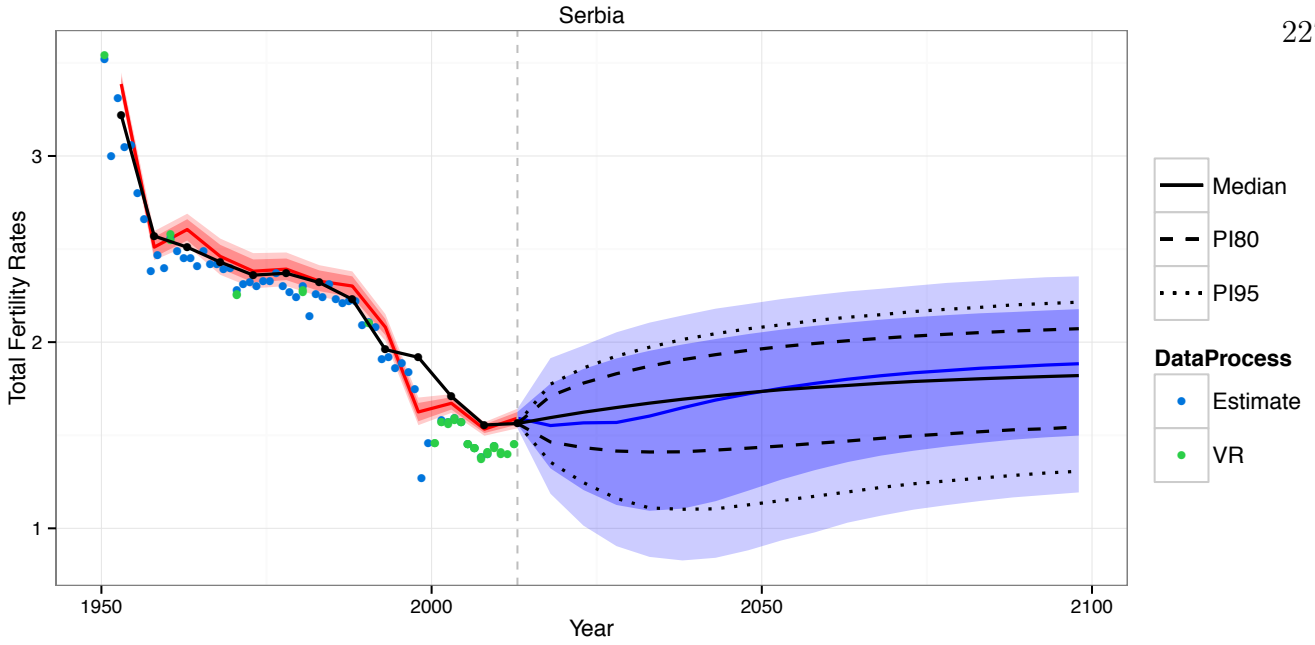

Saint Lucia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
VR

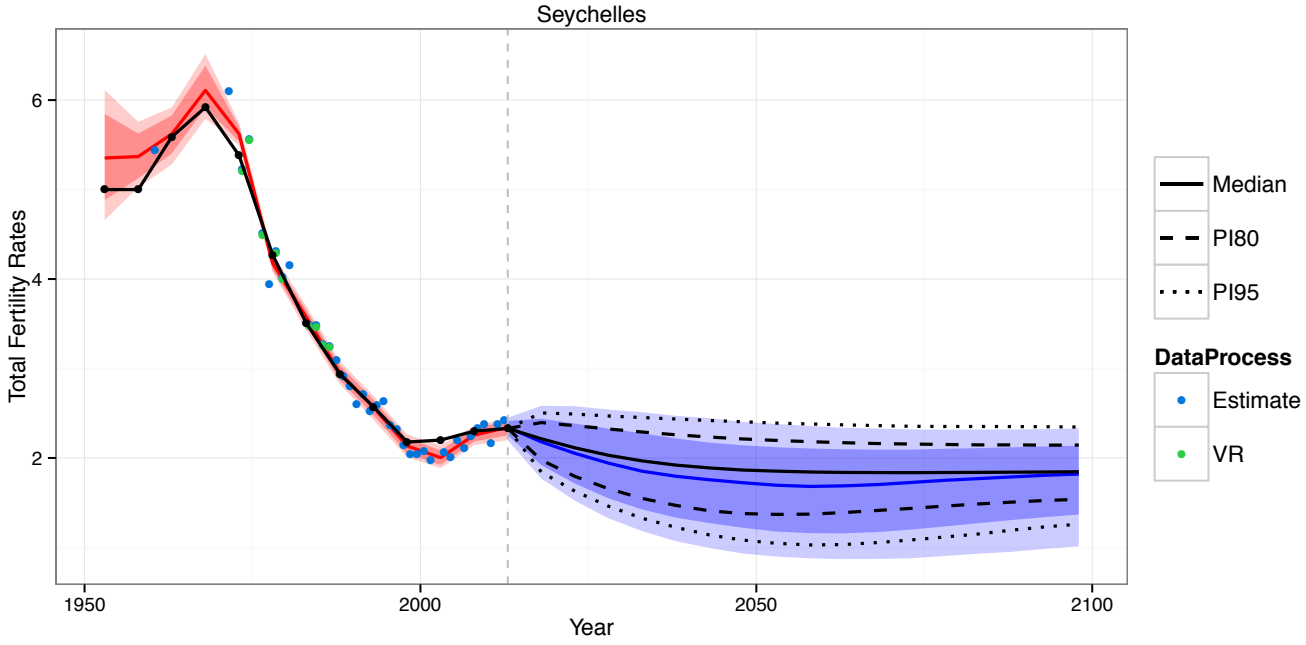

Samoa
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Estimate
Survey
Median
PI80
PI95

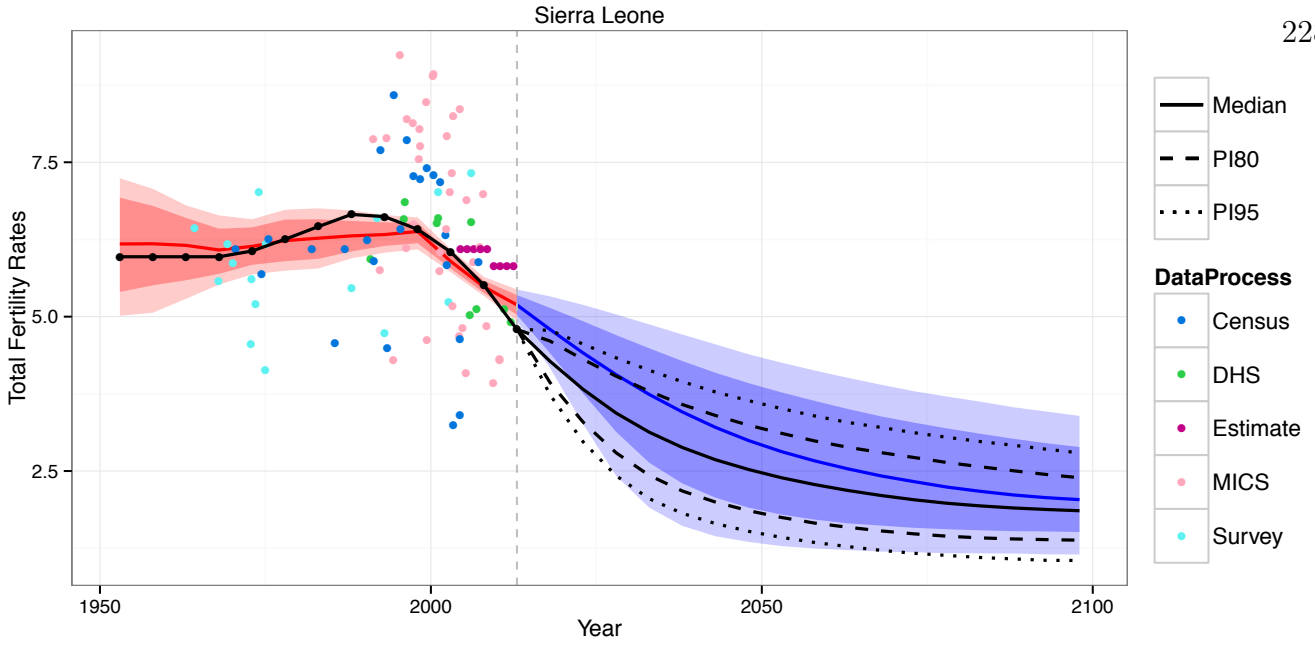

Sao Tome and Principe
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
VR

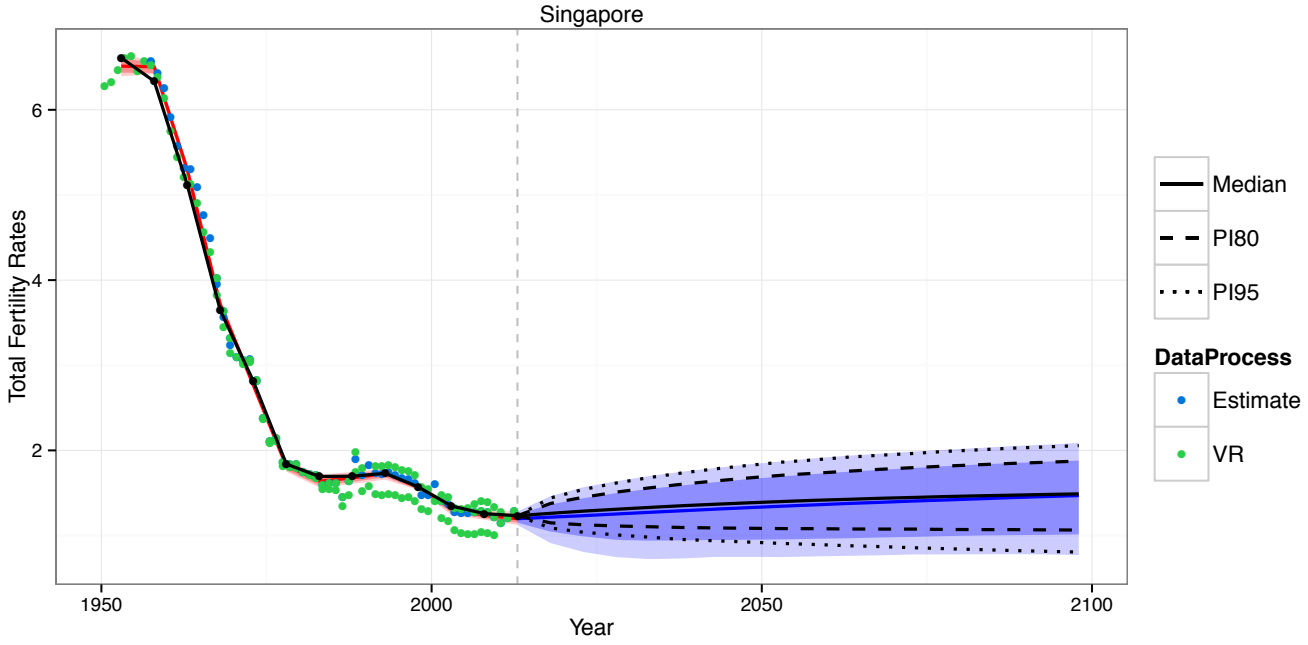

Saudi Arabia
Total Fertility Rates
9
6
3
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
GFHS
PAPCHILD
Survey
VR

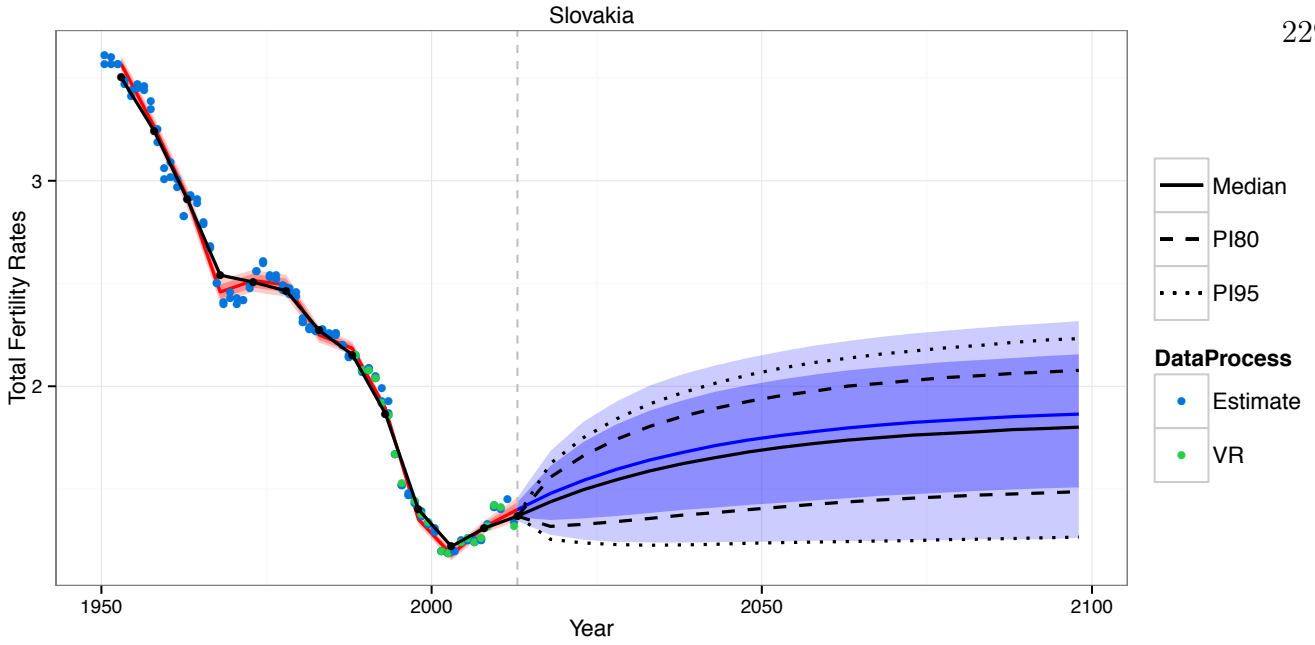

Senegal
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS
MIS
Multiround surv
Survey
WFS

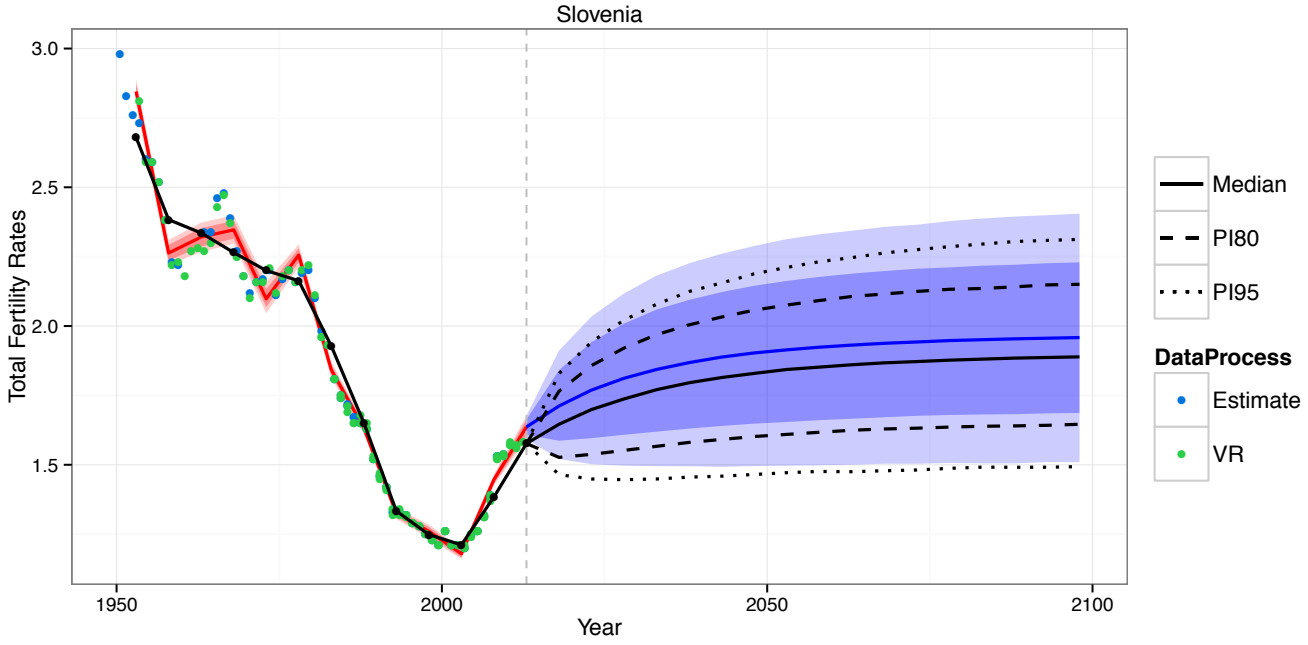

Serbia
Total Fertility Rates
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

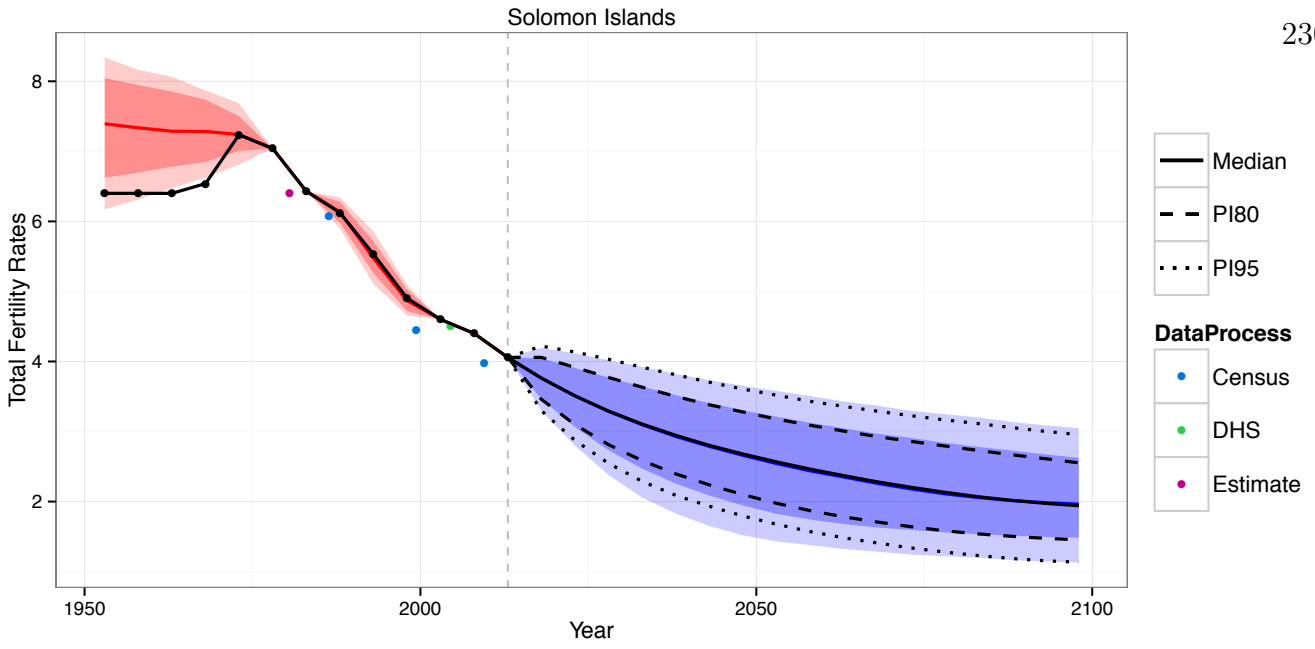

Seychelles
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

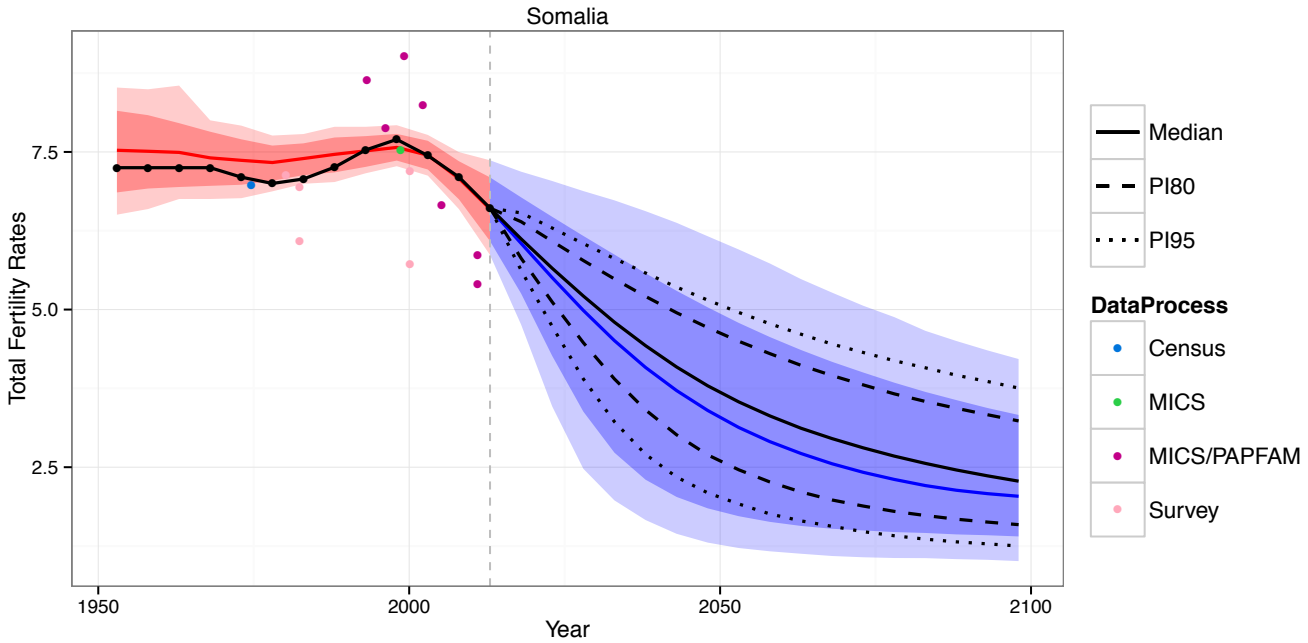

Sierra Leone
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS
Survey

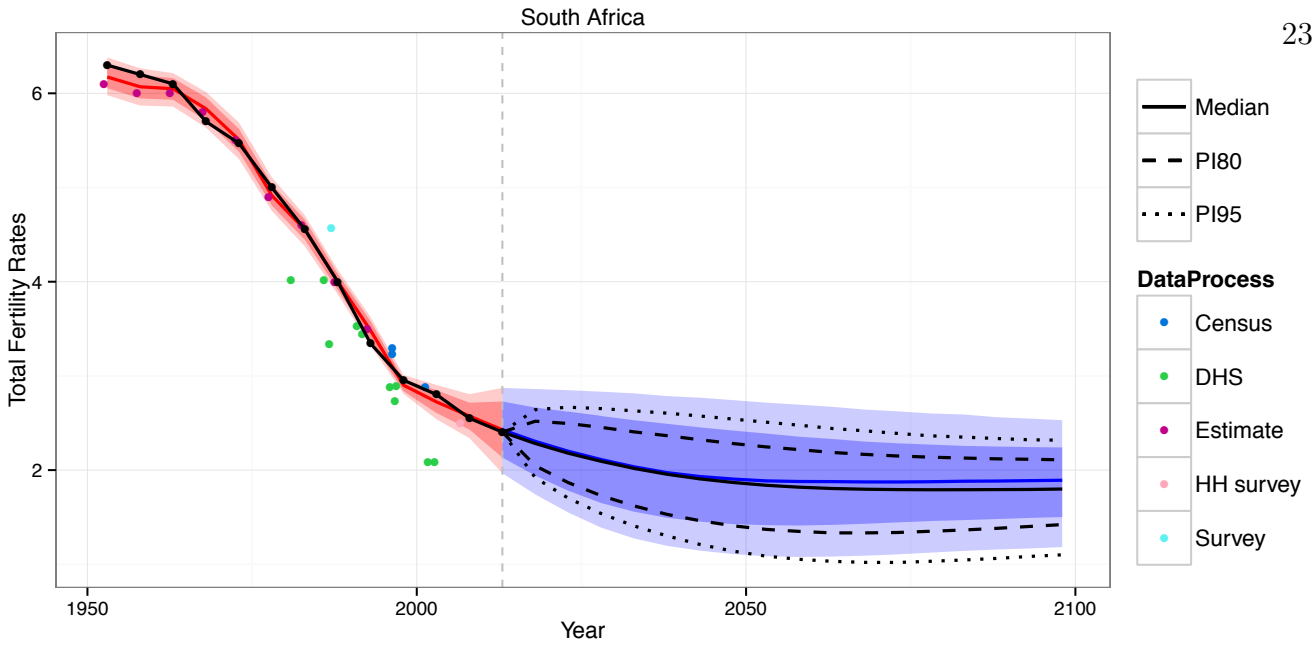

Singapore
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

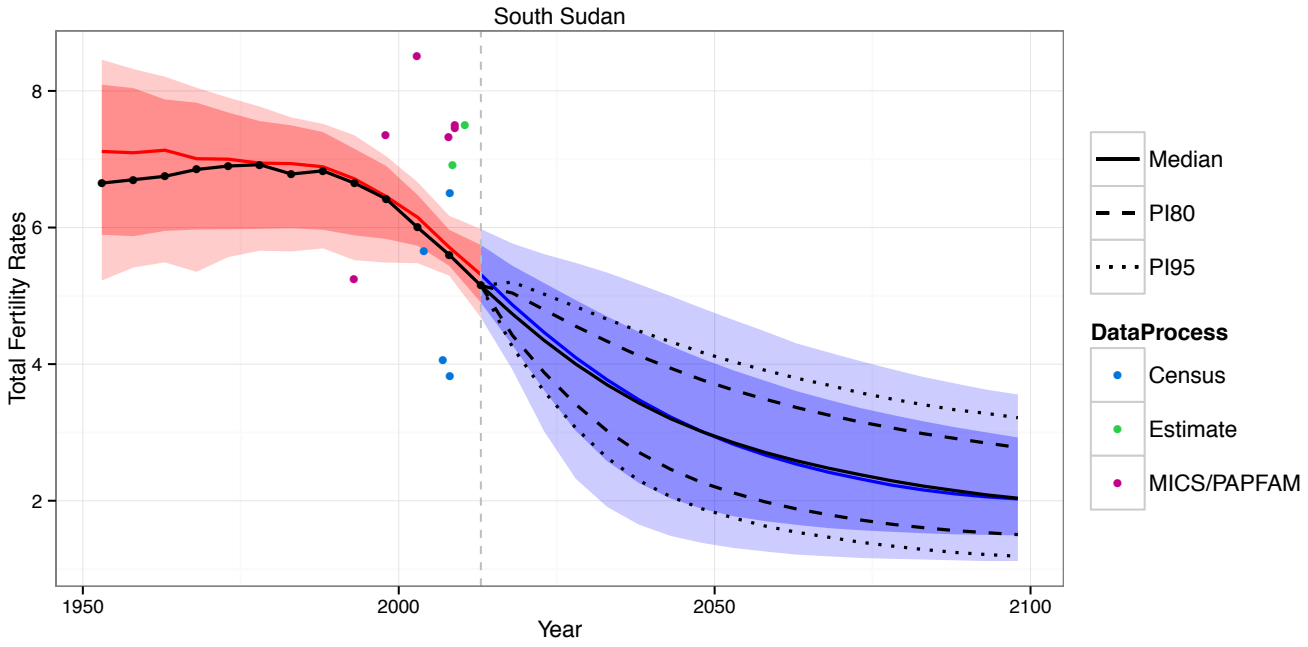

Slovakia
Total Fertility Rates
3
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

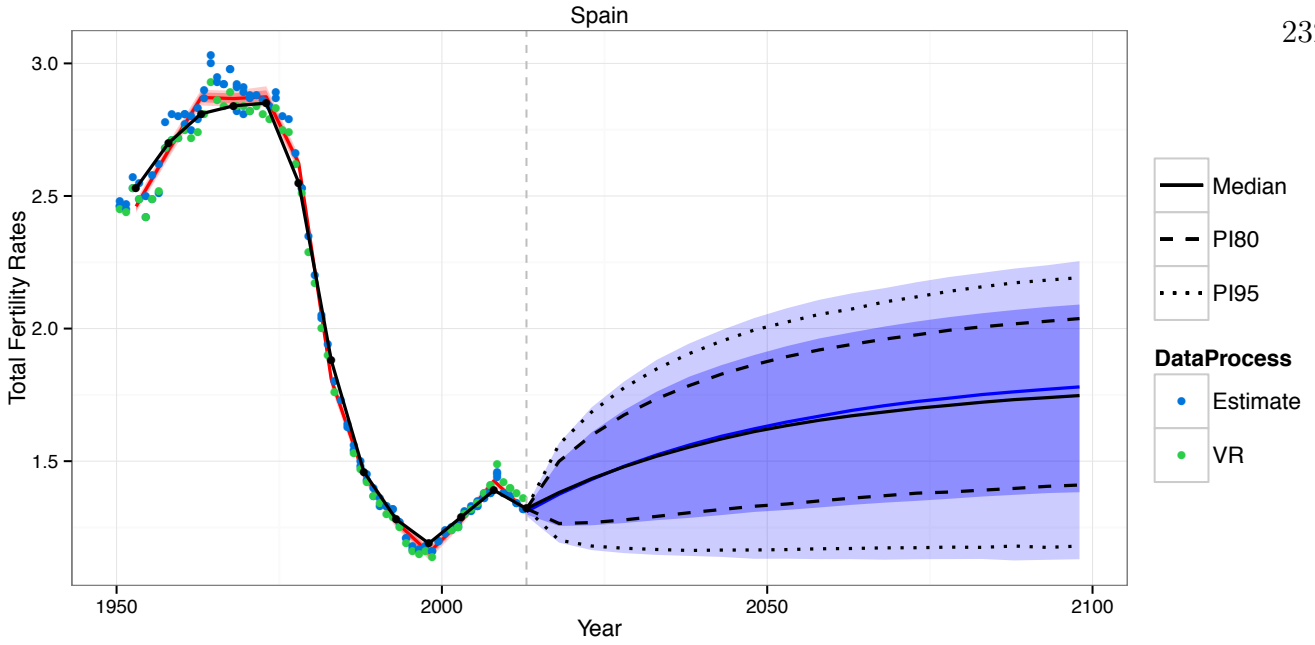

Slovenia
Total Fertility Rates
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

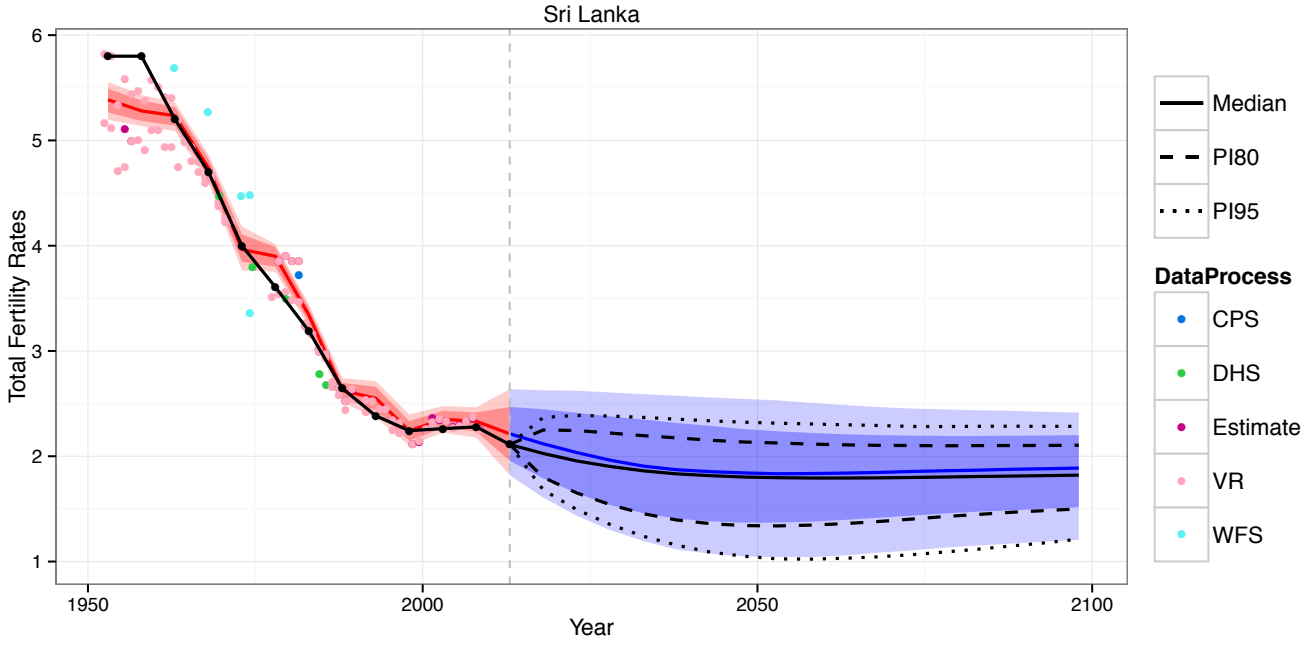

Solomon Islands
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate

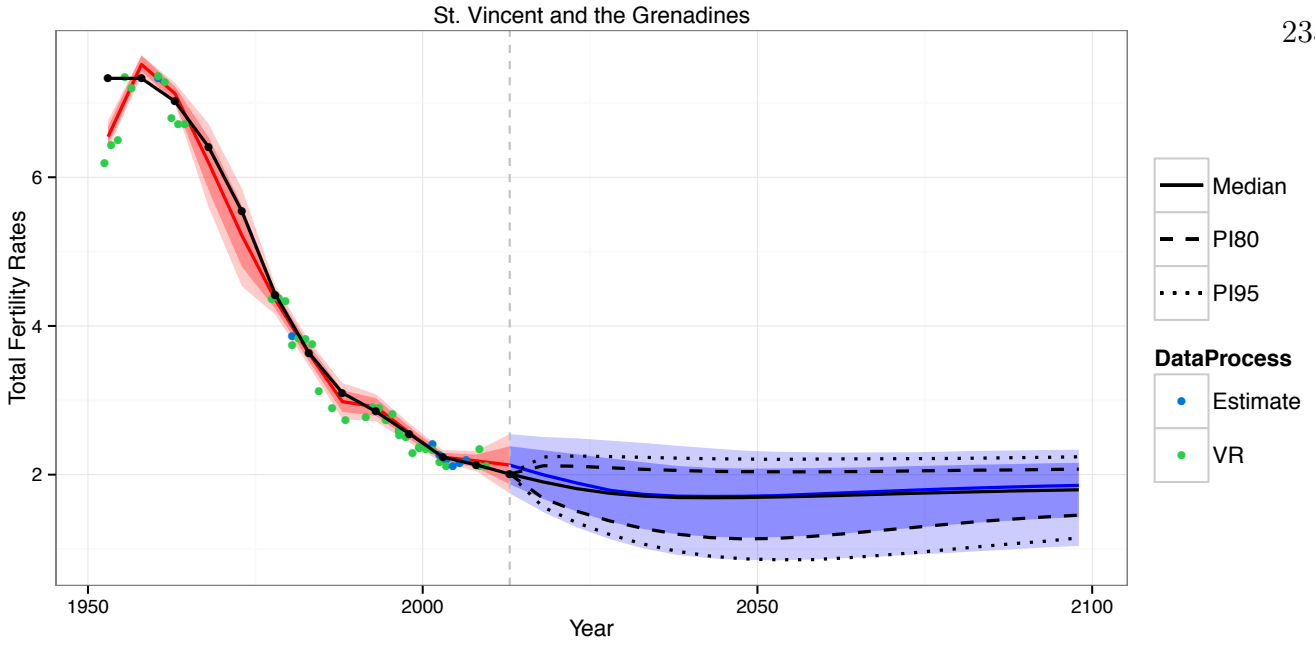

Somalia
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
MICS
MICS/PAPFAM
Survey

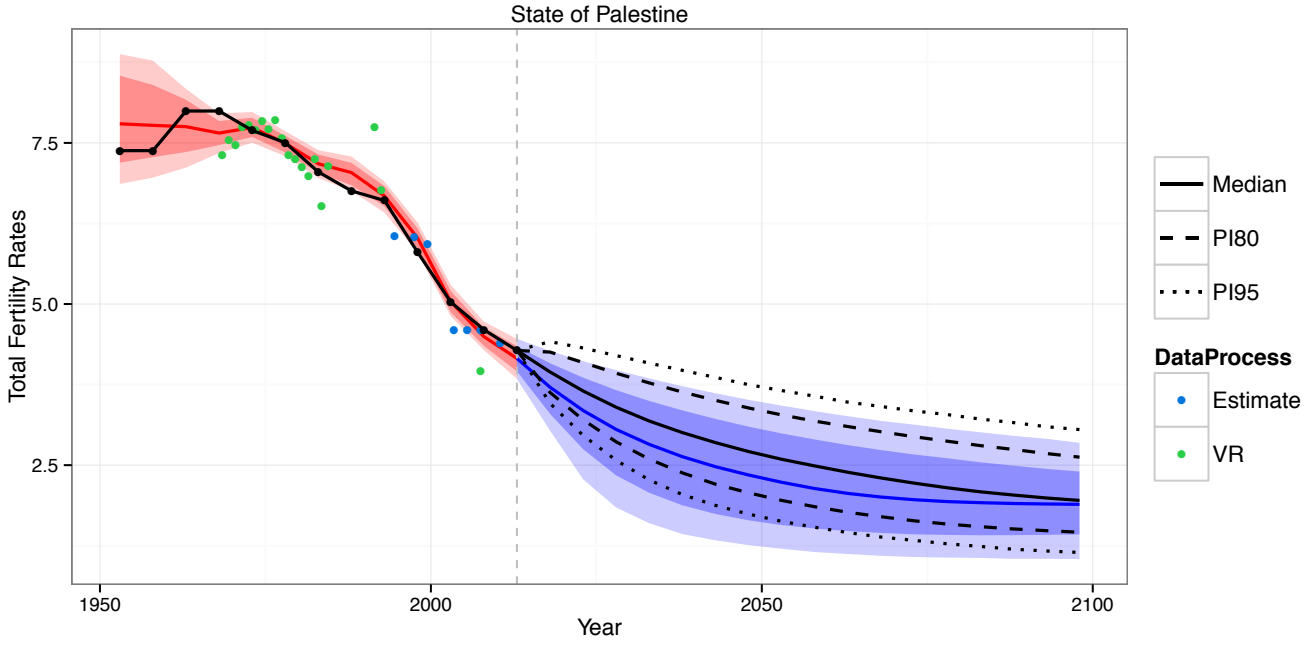

South Africa
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
HH survey
Survey

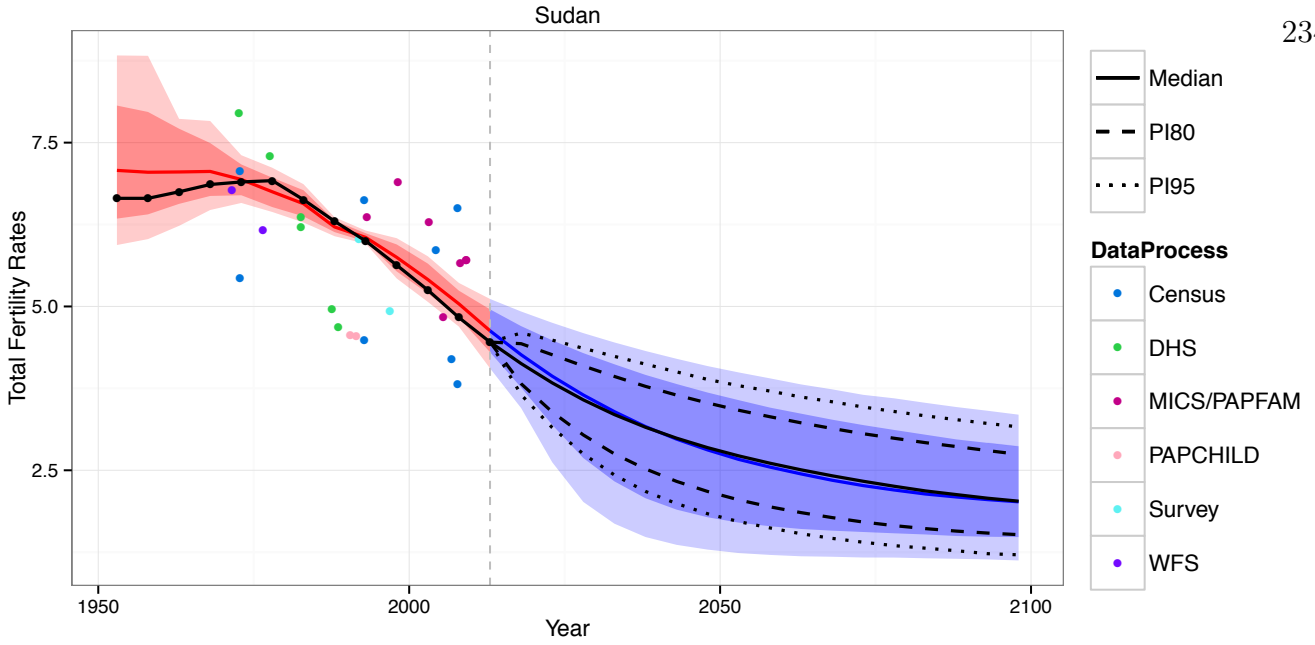

South Sudan
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
MICS/PAPFAM

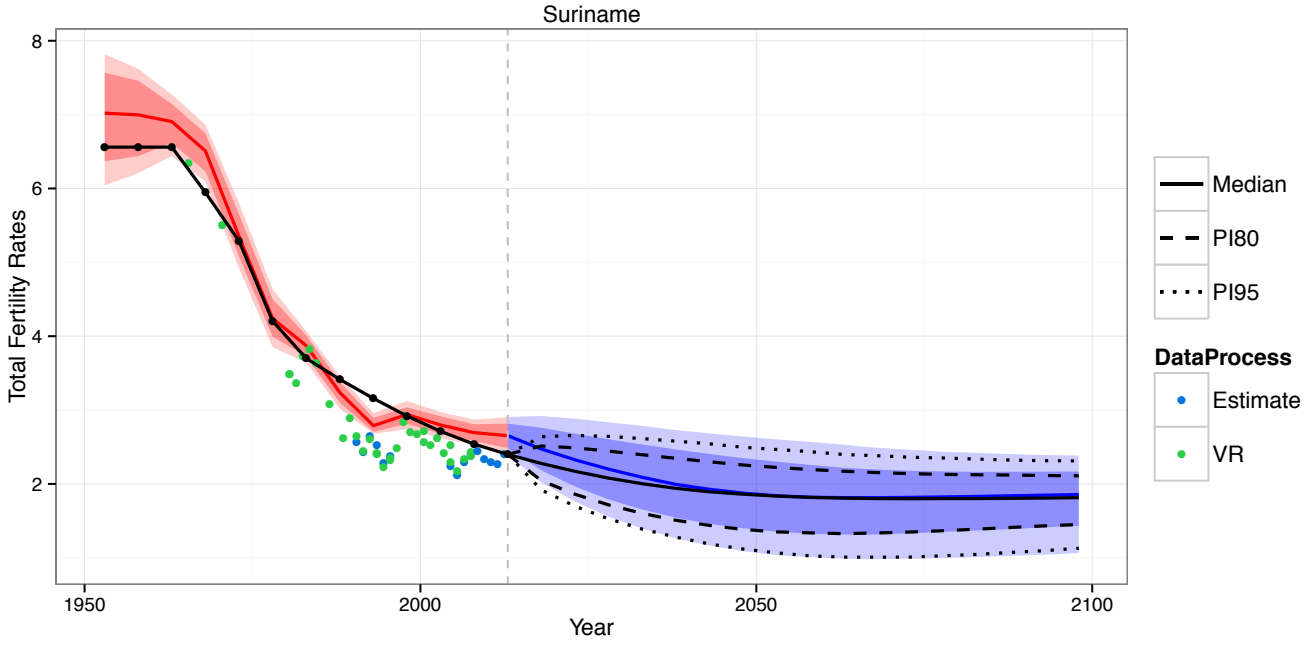

Spain
Total Fertility Rates
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

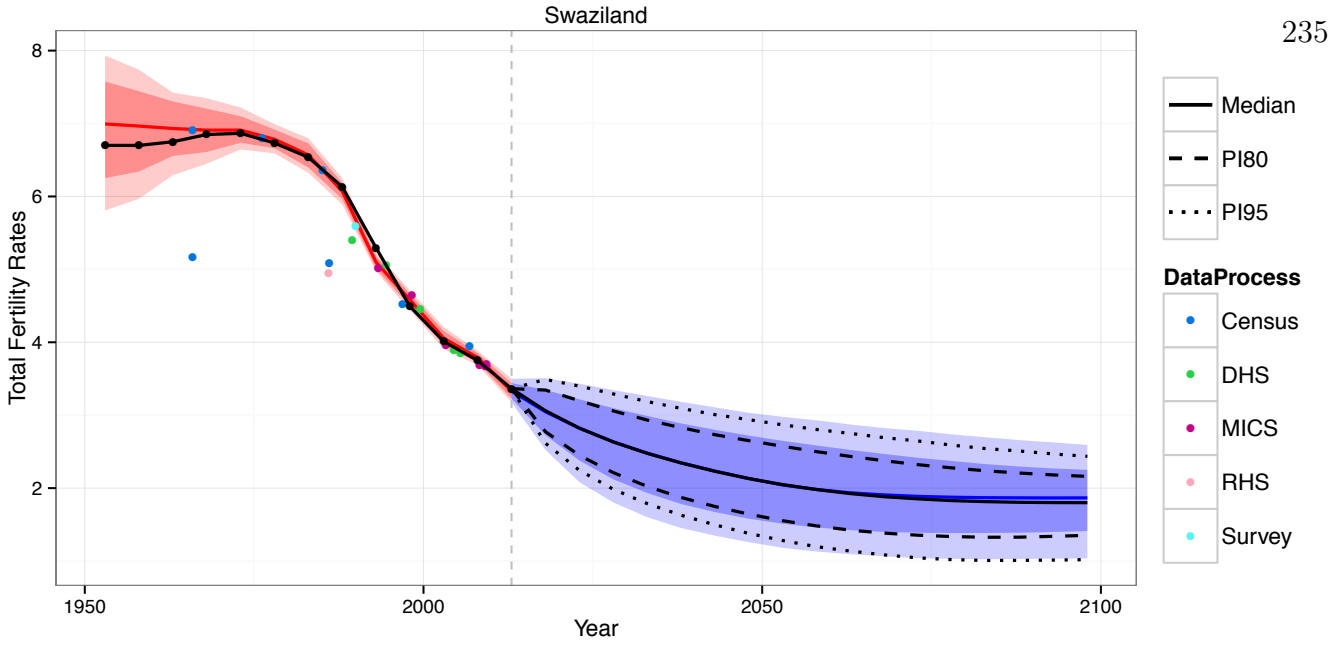

Sri Lanka
Total Fertility Rates
6
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
CPS
DHS
Estimate
VR
WFS

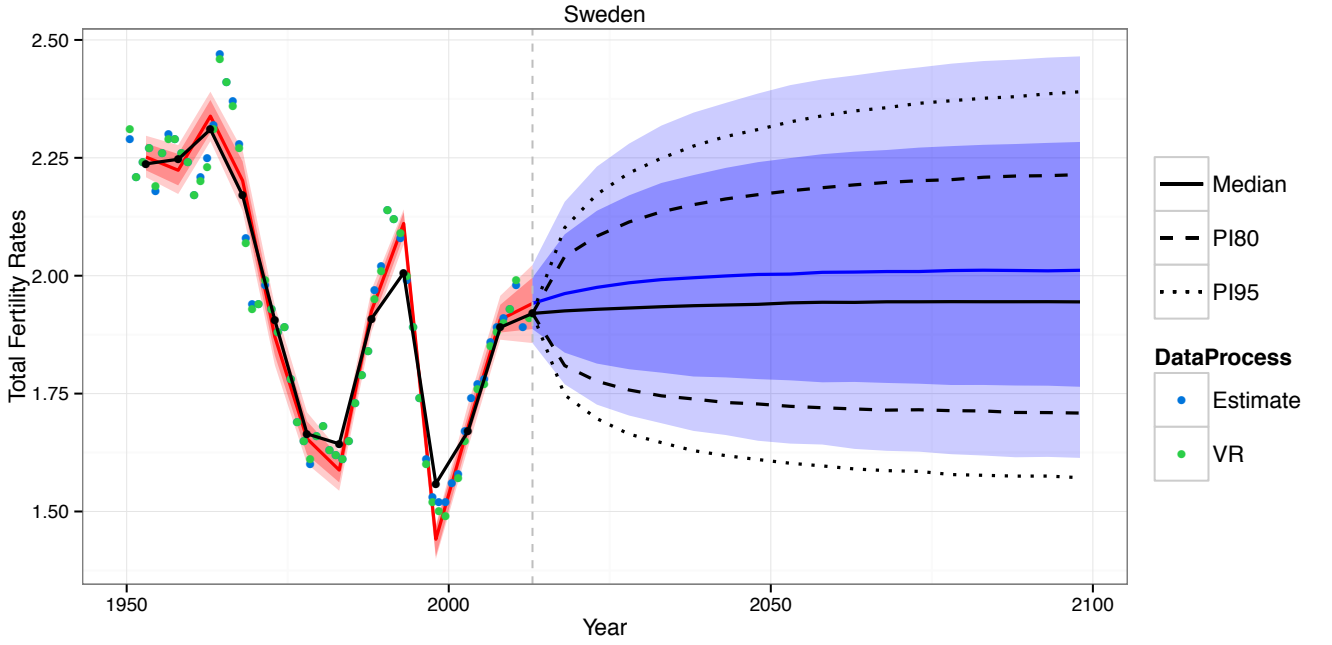

St. Vincent and the Grenadines
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

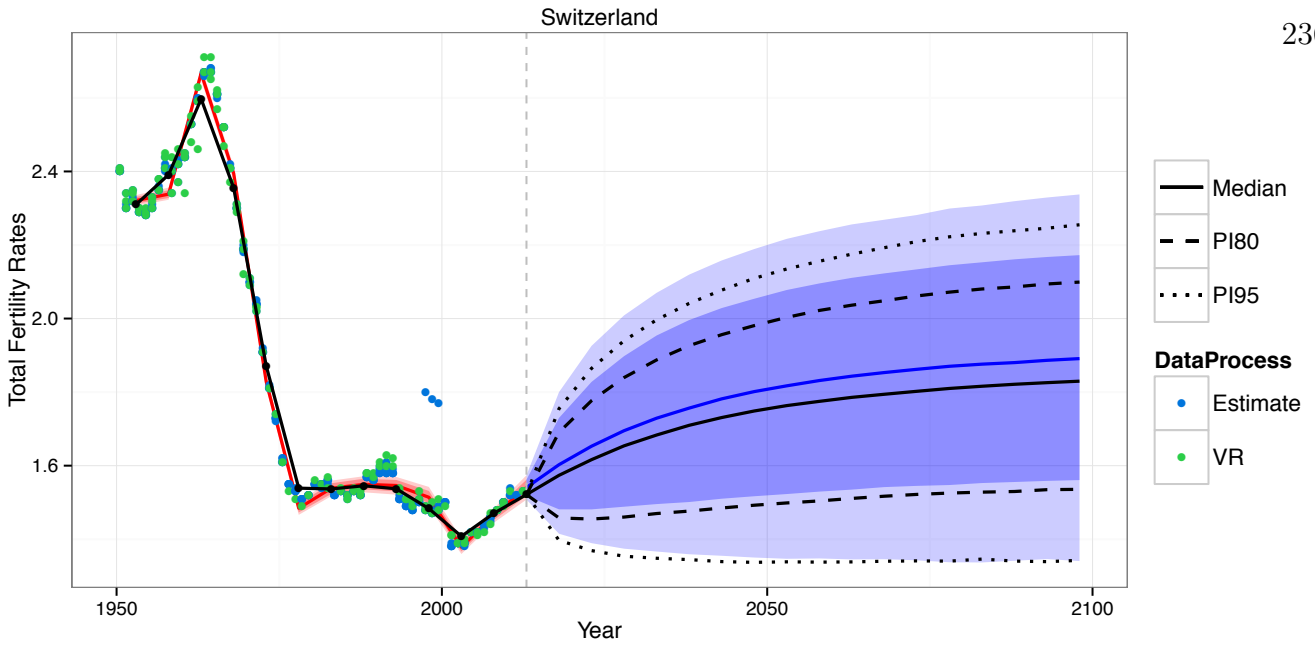

State of Palestine
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

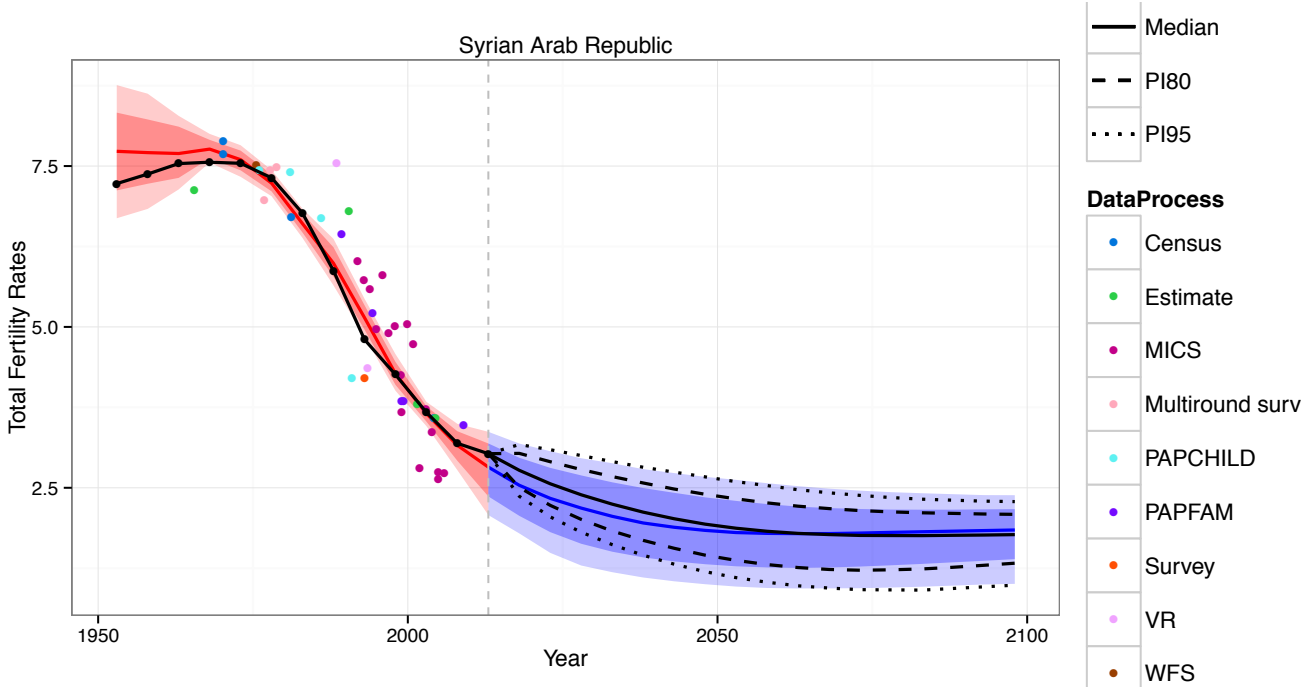

Sudan
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
MICS/PAPFAM
PAPCHILD
Survey
WFS

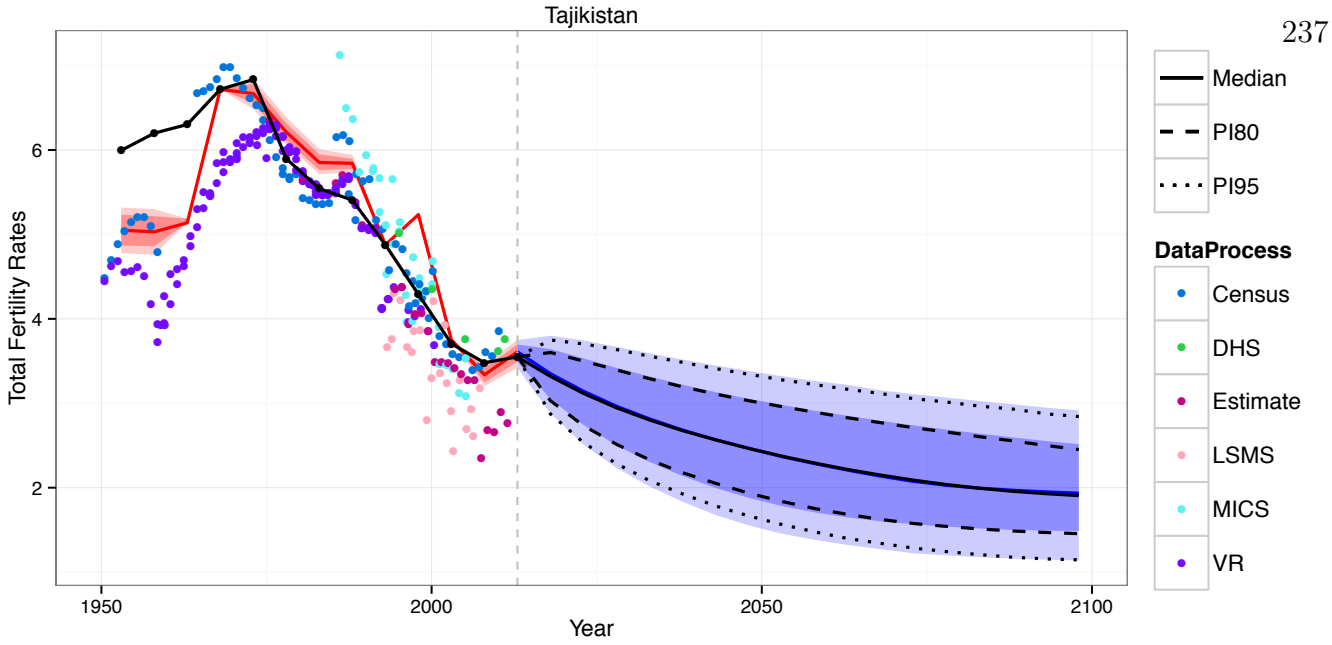

Suriname
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

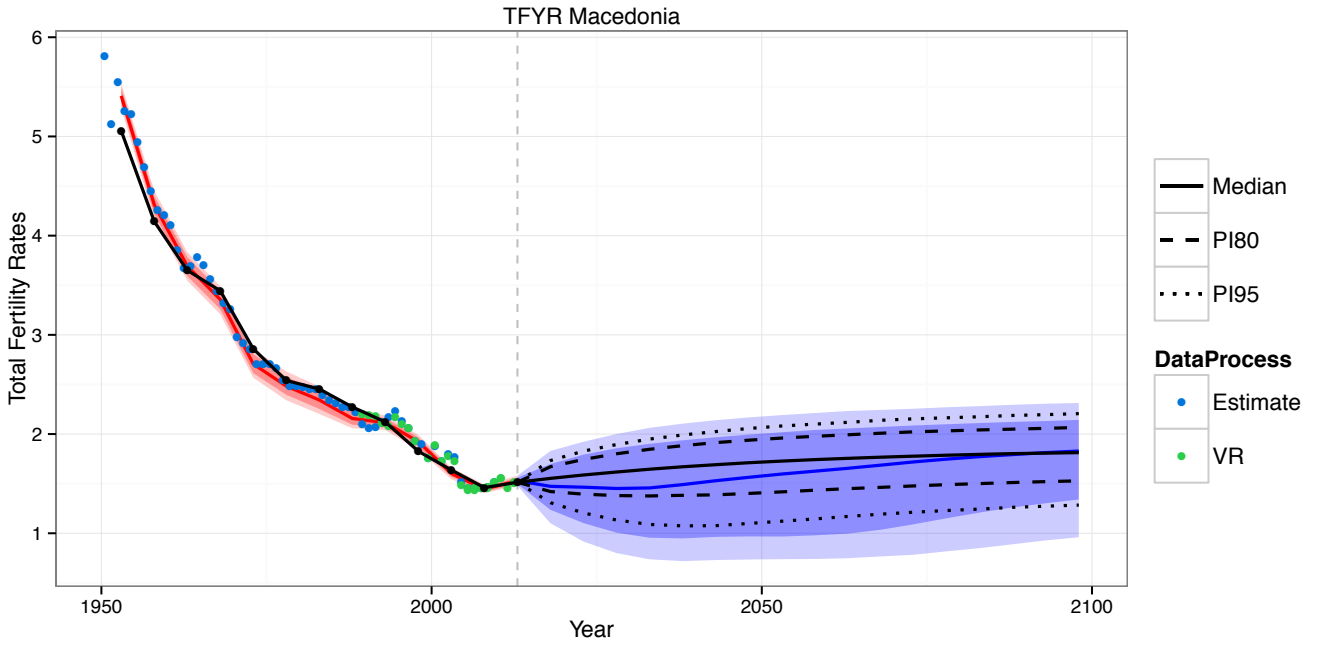

Swaziland
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
MICS
RHS
Survey

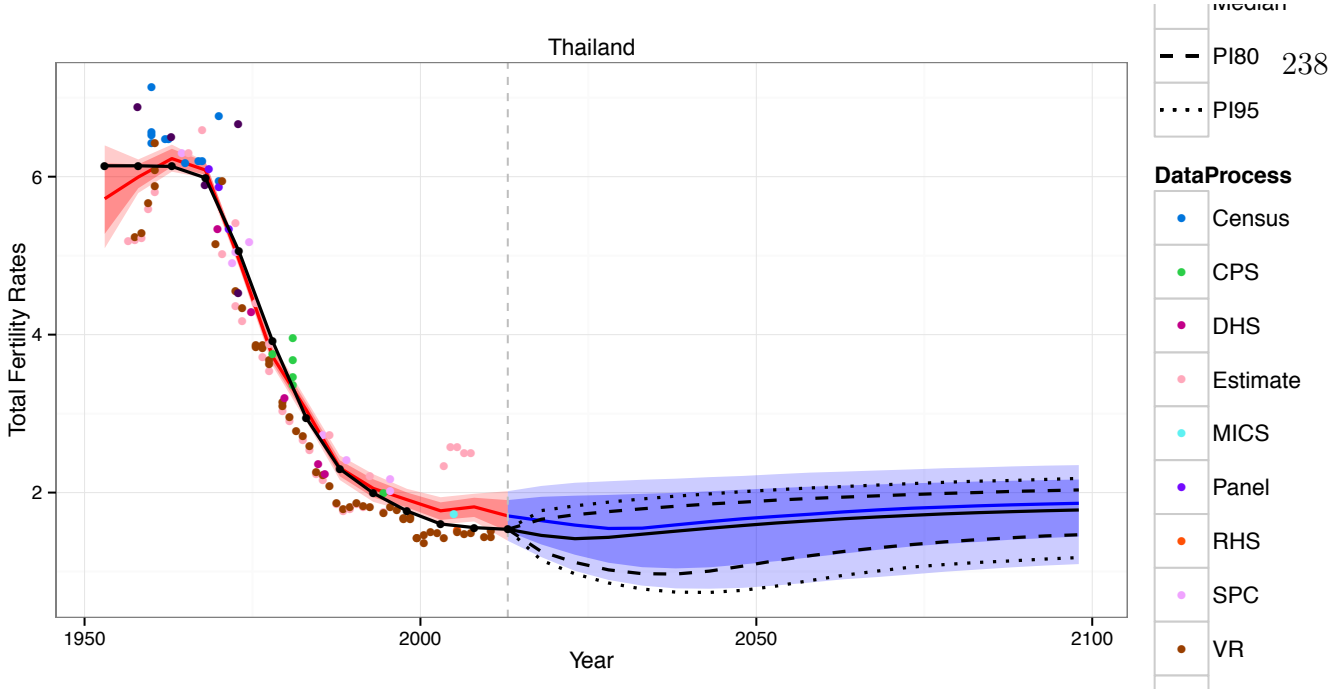

Sweden
Total Fertility Rates
2.50
2.25
2.00
1.75
1.50
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

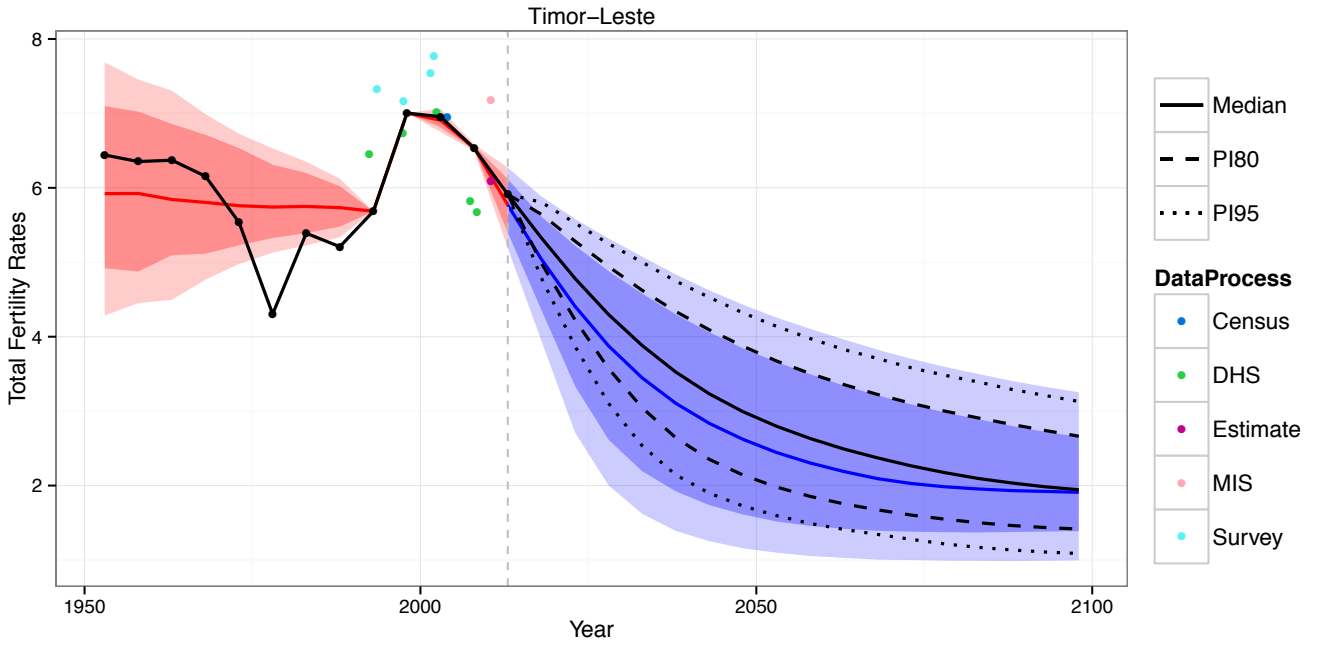

Switzerland
Total Fertility Rates
2.4
2.0
1.6
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

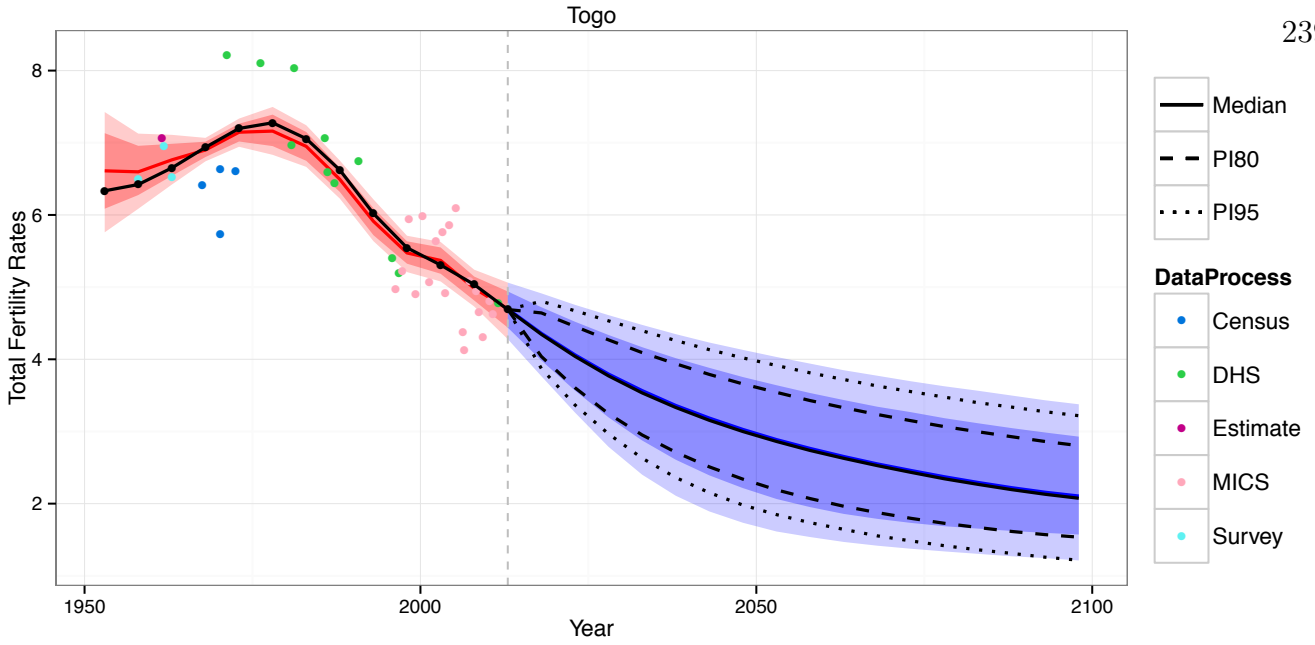

Syrian Arab Republic
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
MICS
Multiround surv
PAPCHILD
PAPFAM
Survey
VR
WFS

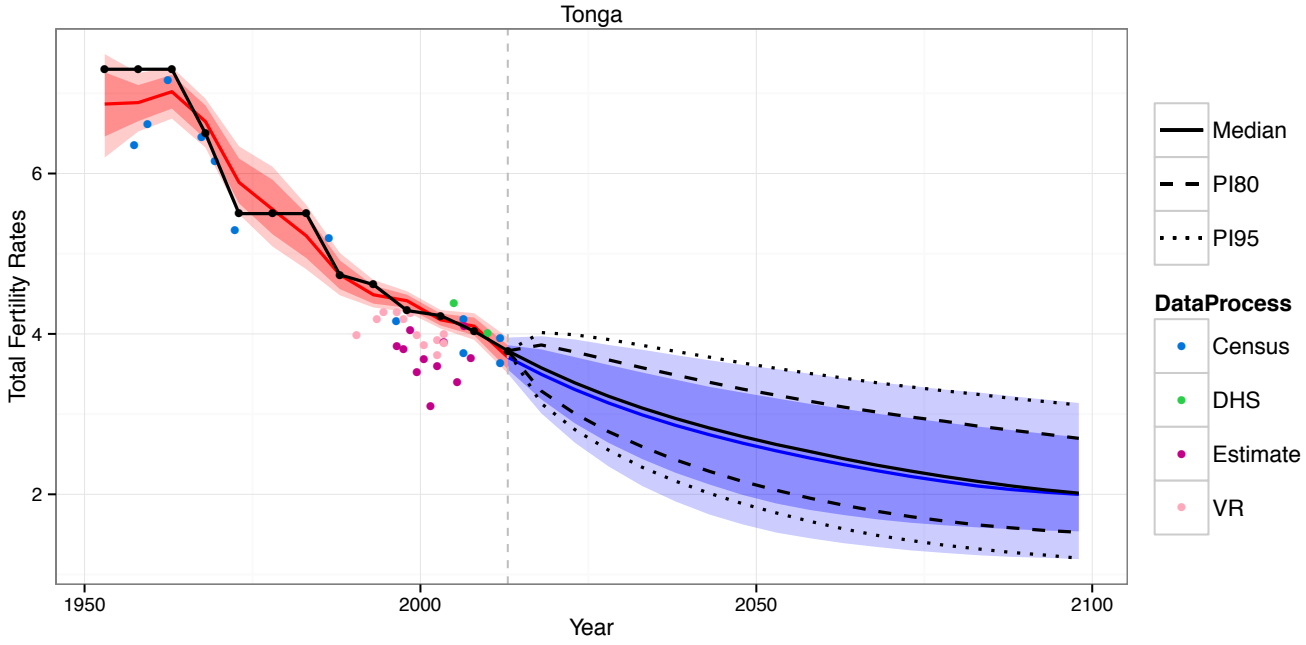

Tajikistan
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
LSMS
MICS
VR

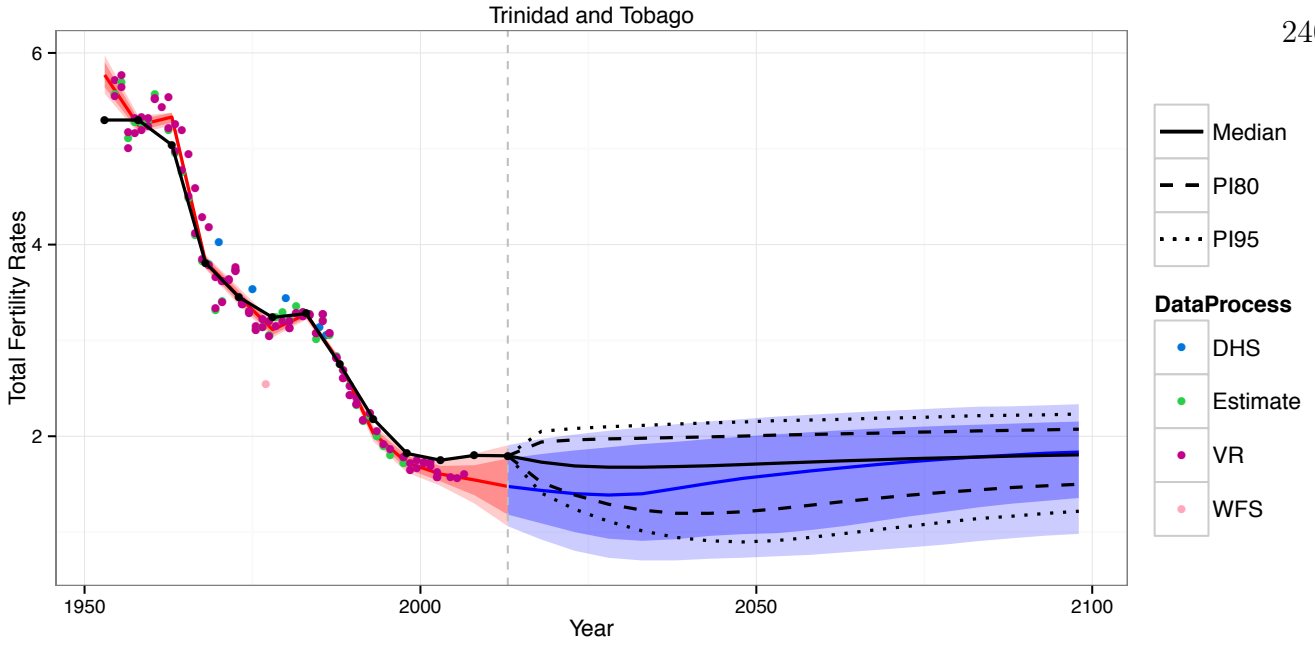

TFYR Macedonia
Total Fertility Rates
6
5
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

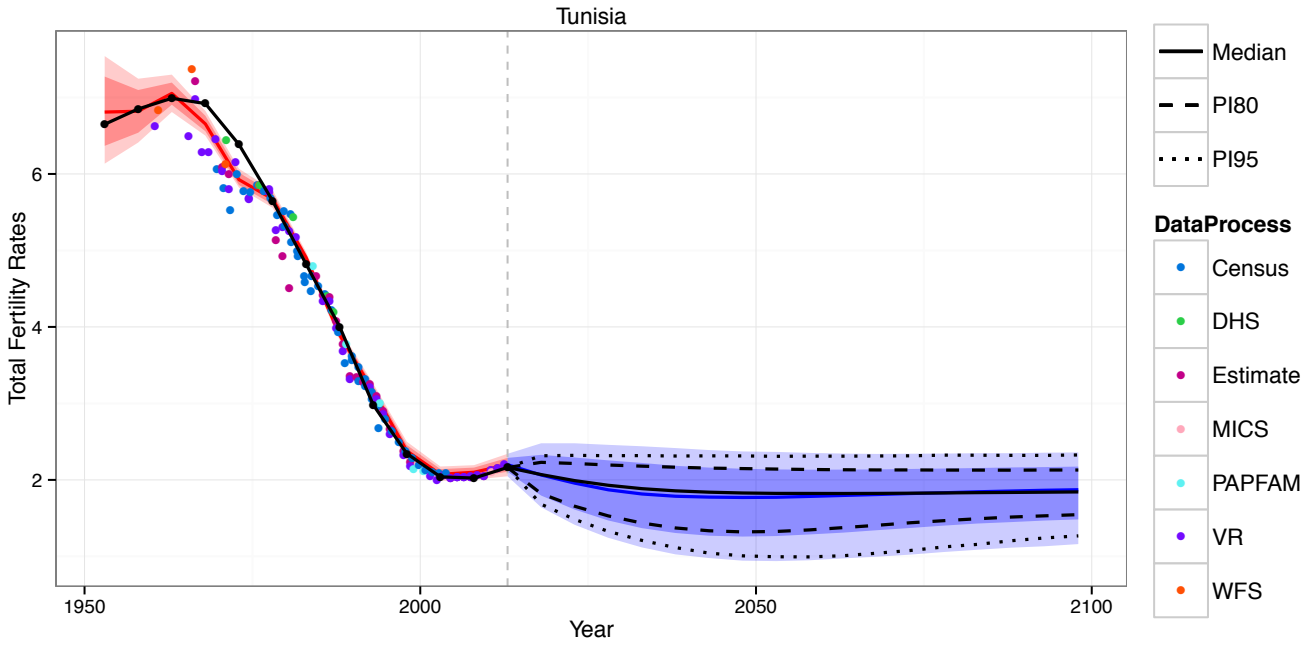

Thailand
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
PI80
PI95
DataProcess
Census
CPS
DHS
Estimate
MICS
Panel
RHS
SPC
VR

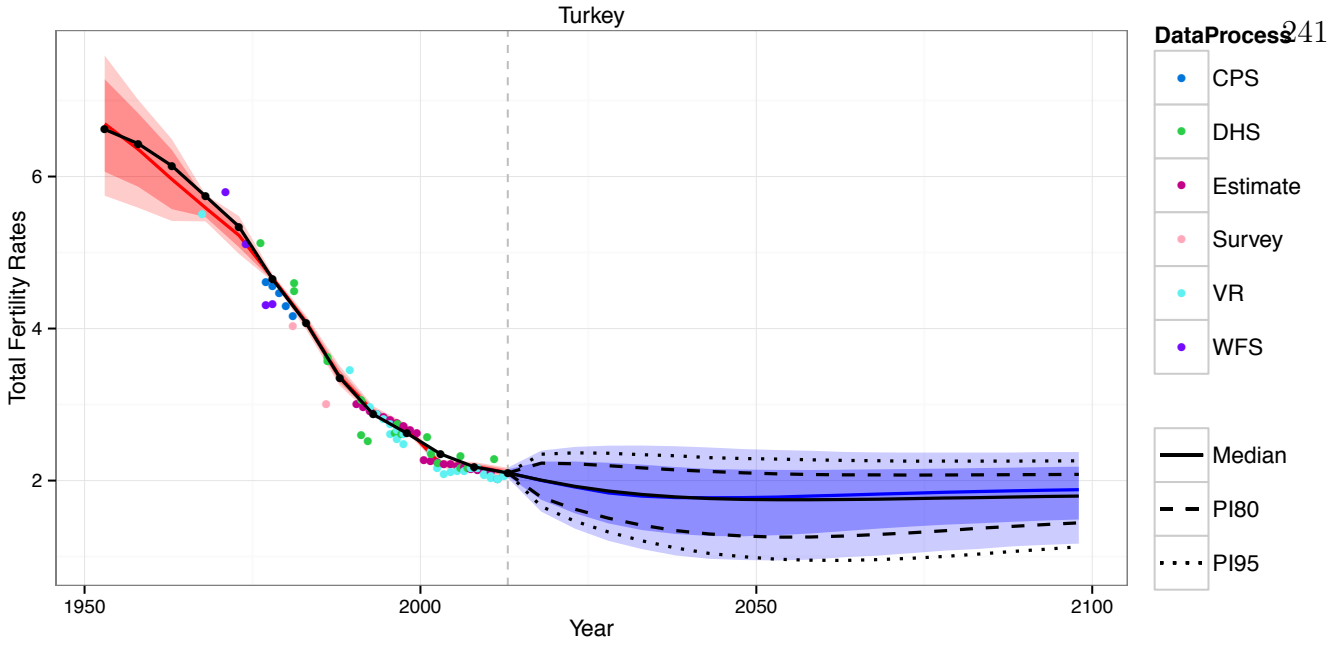

Timor−Leste
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MIS
Survey

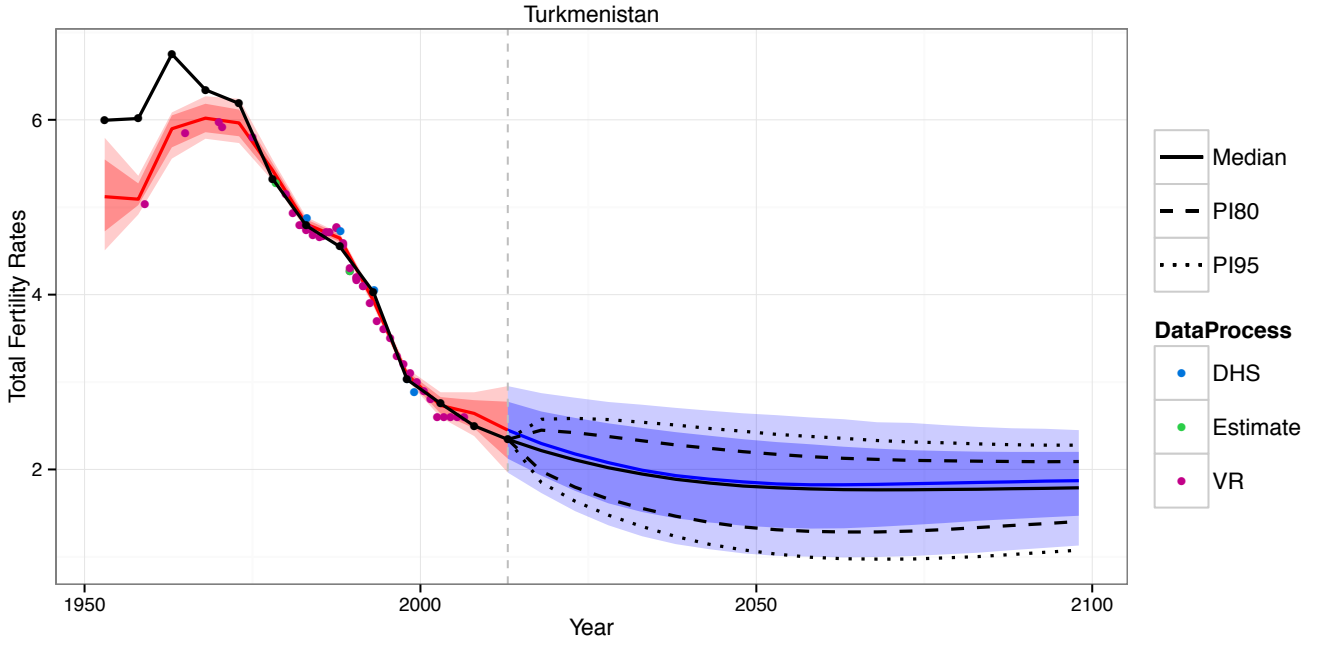

Togo
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS
Survey

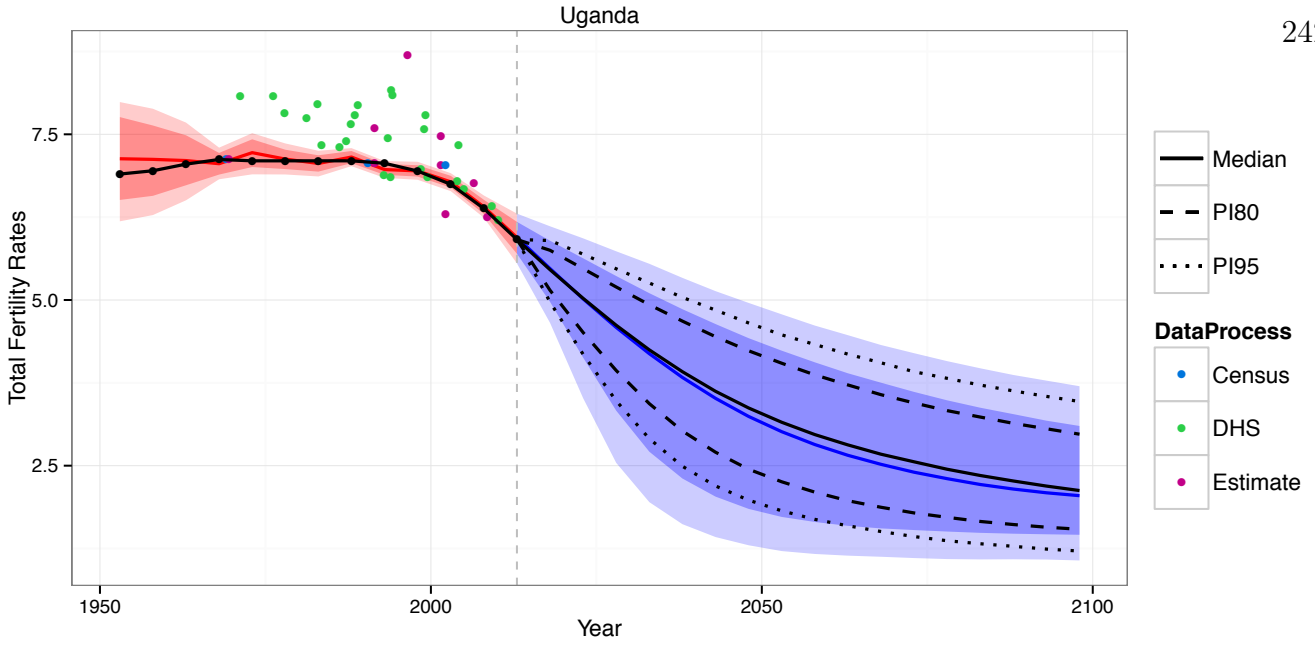

Tonga
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
VR

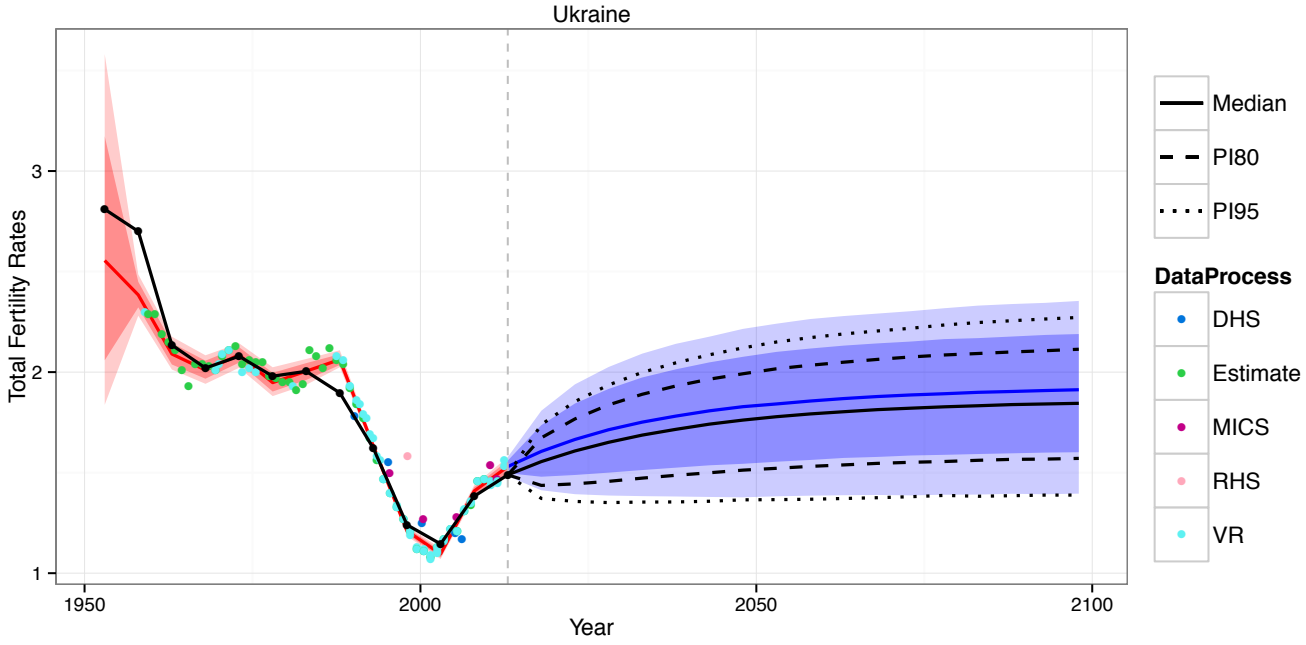

Trinidad and Tobago
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
VR
WFS

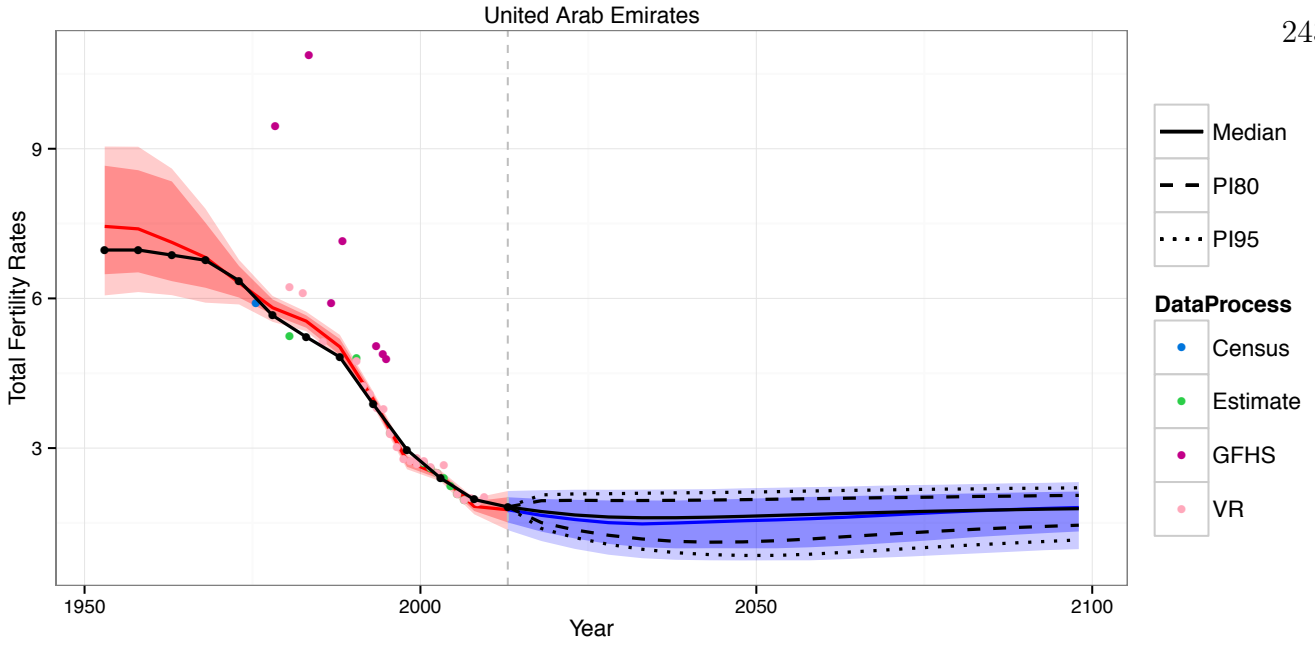

Tunisia
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS
PAPFAM
VR
WFS

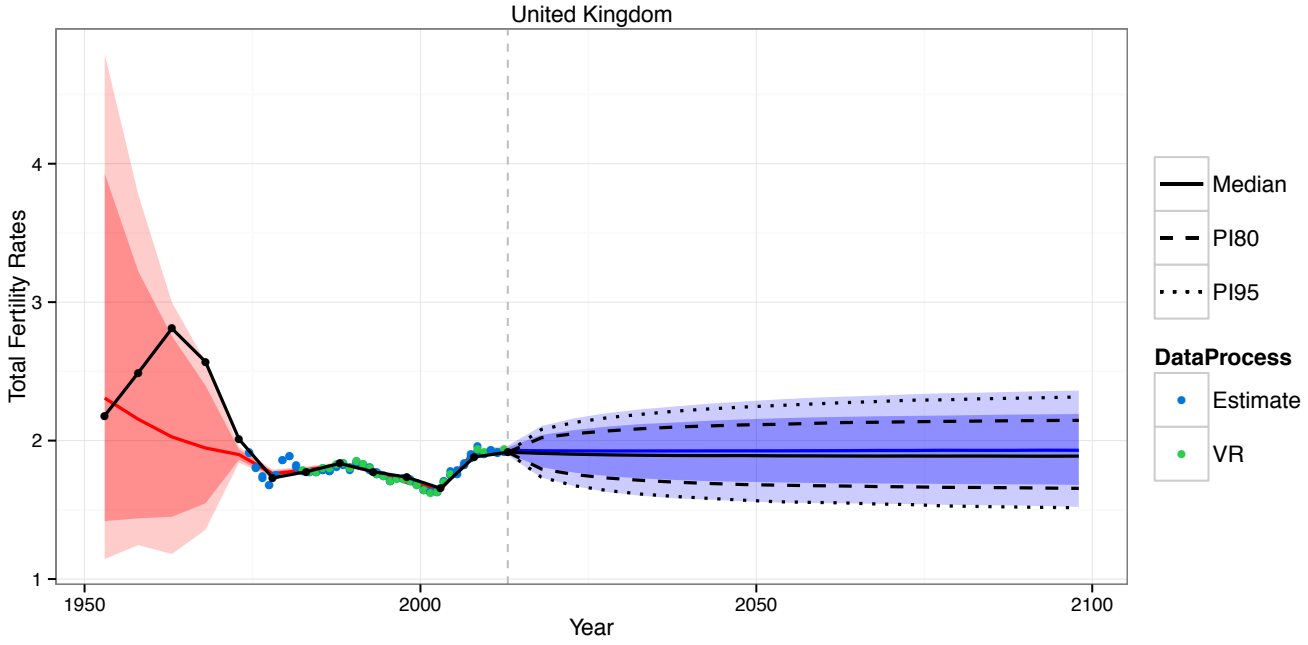

Turkey
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
DataProcess
CPS
DHS
Estimate
Survey
VR
WFS
Median
PI80
PI95

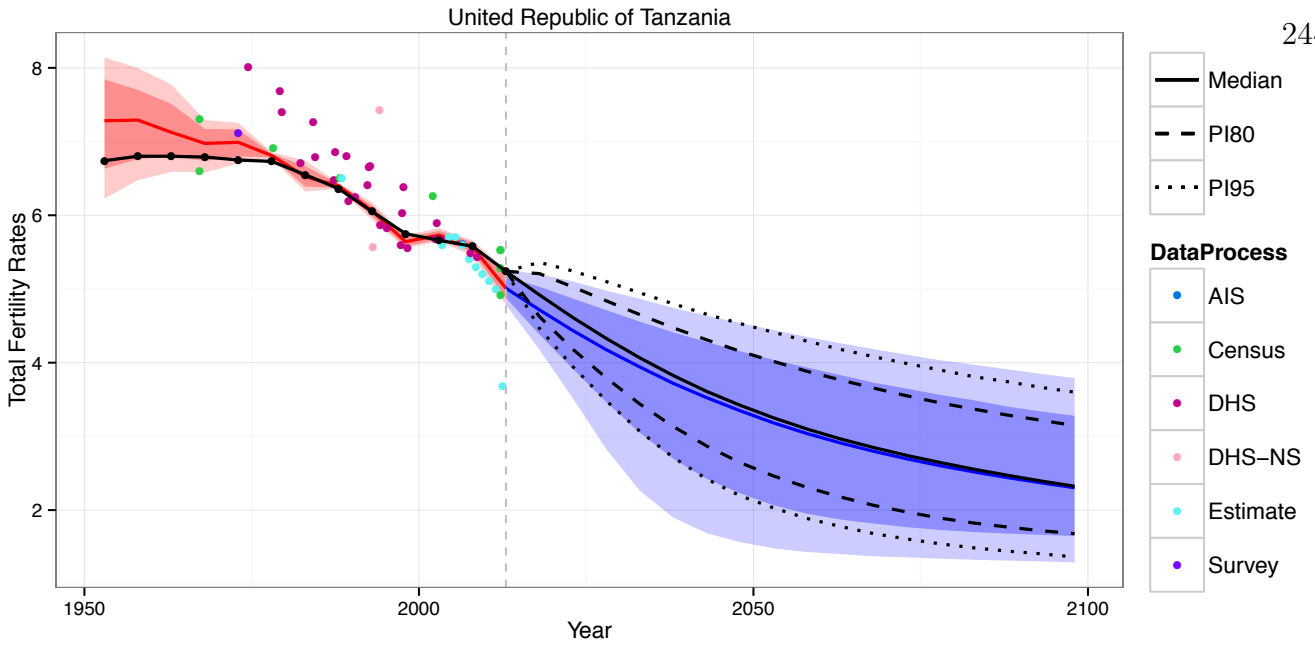

Turkmenistan
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
VR

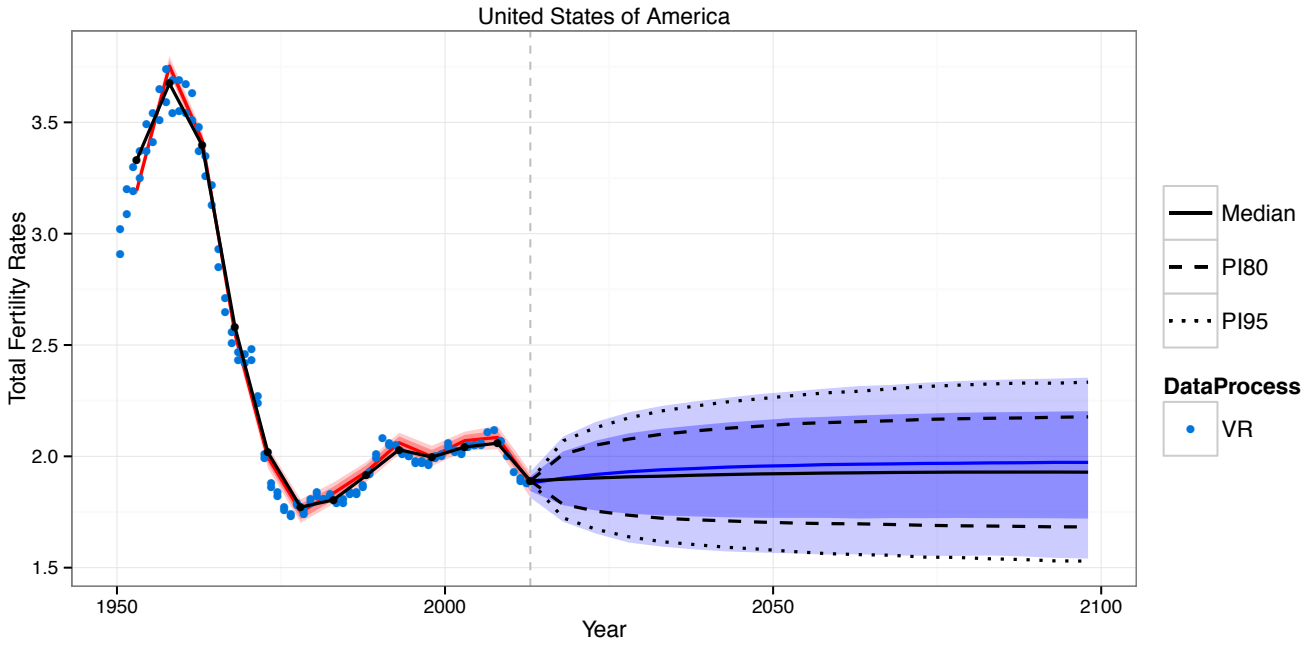

Uganda
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate

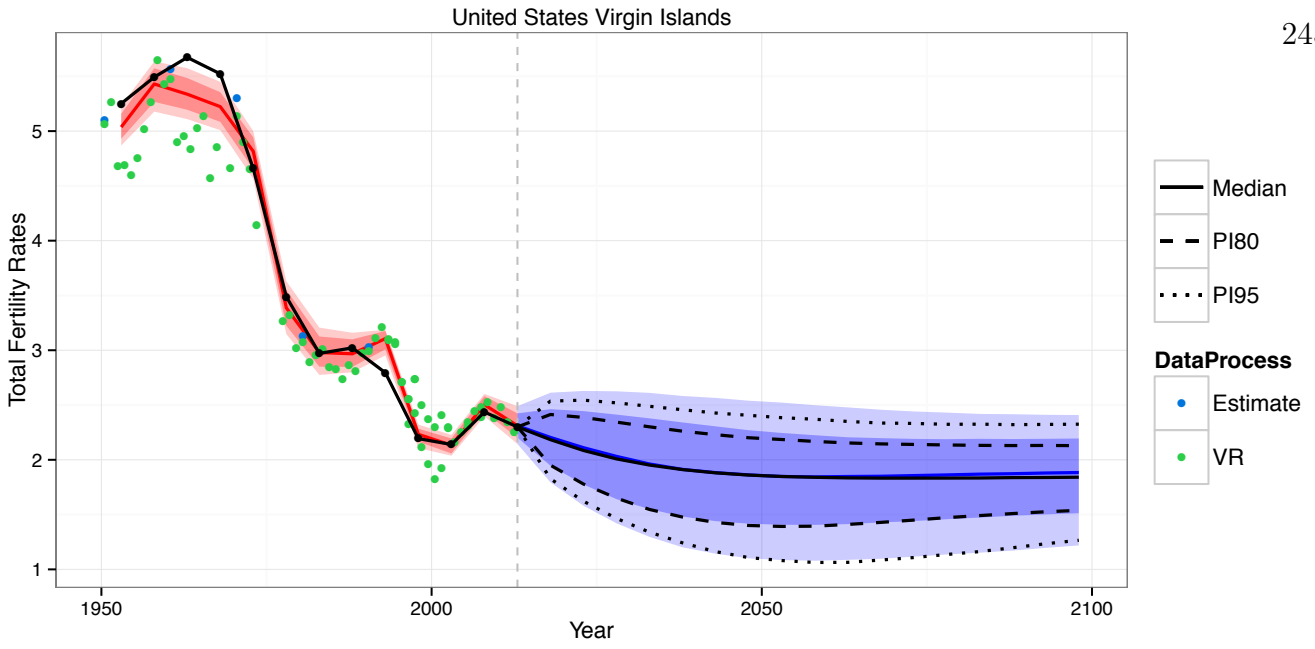

Ukraine
Total Fertility Rates
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
Estimate
MICS
RHS
VR

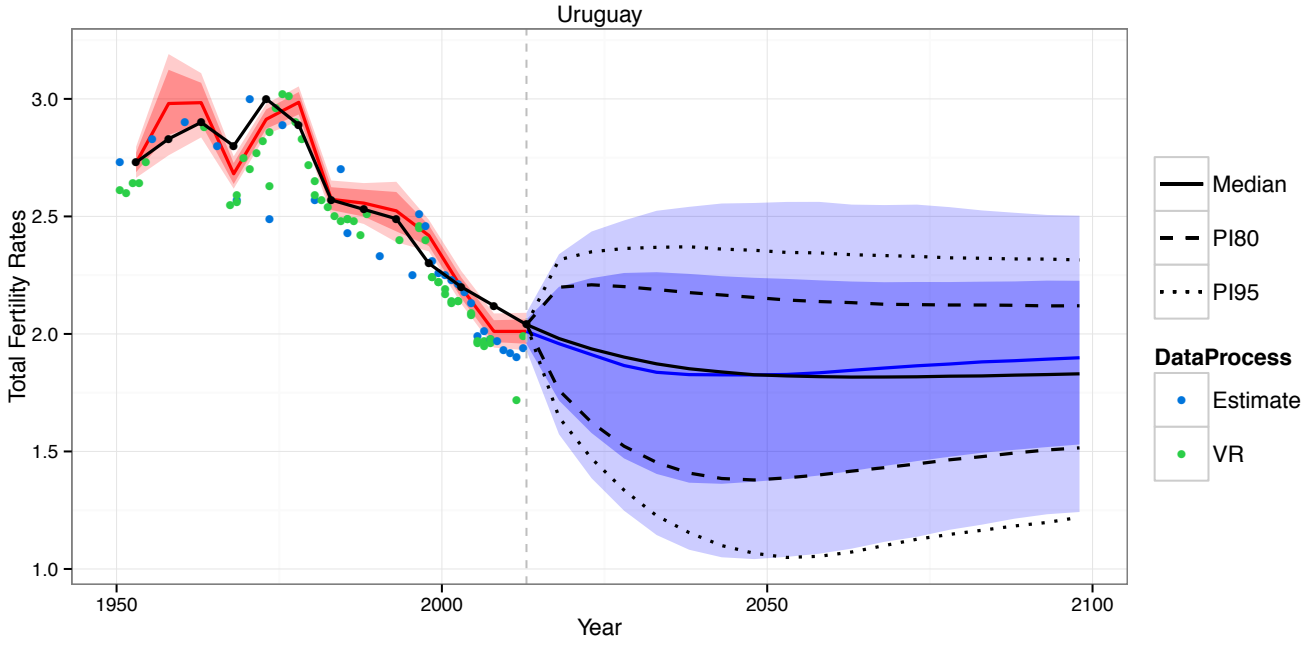

United Arab Emirates
Total Fertility Rates
9
6
3
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
Estimate
GFHS
VR

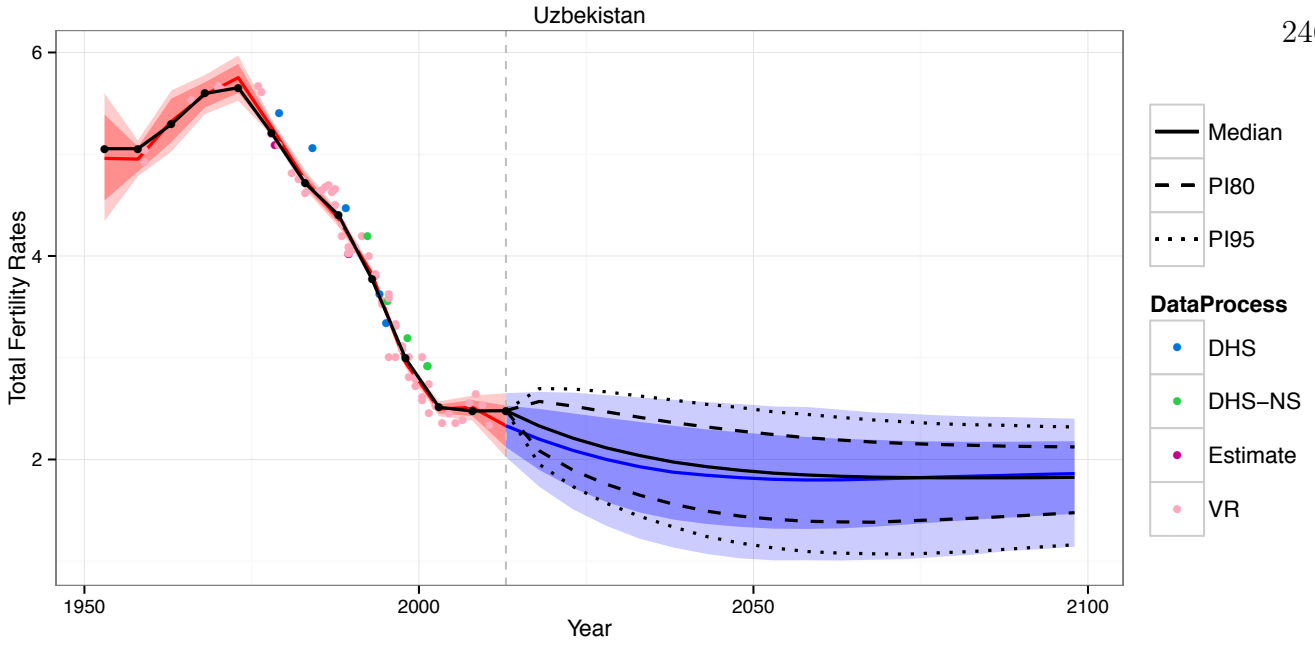

United Kingdom
Total Fertility Rates
4
3
2
1
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR

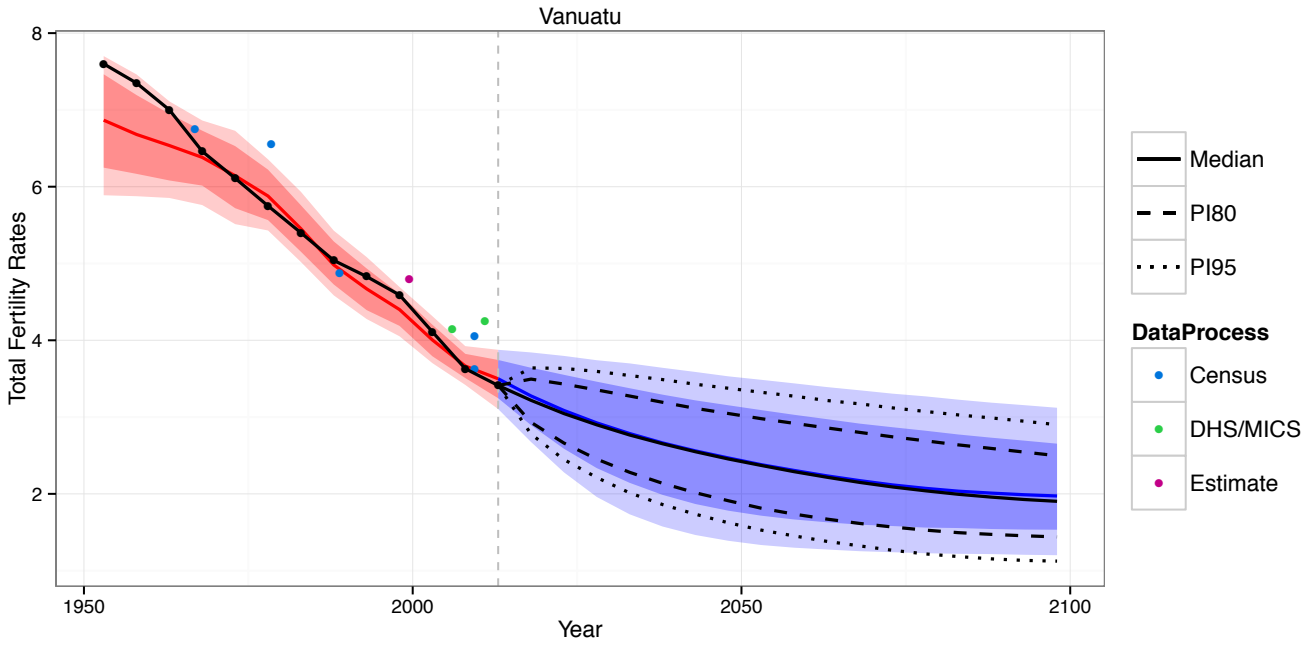

United Republic of Tanzania
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
AIS
Census
DHS
DHS−NS
Estimate
Survey

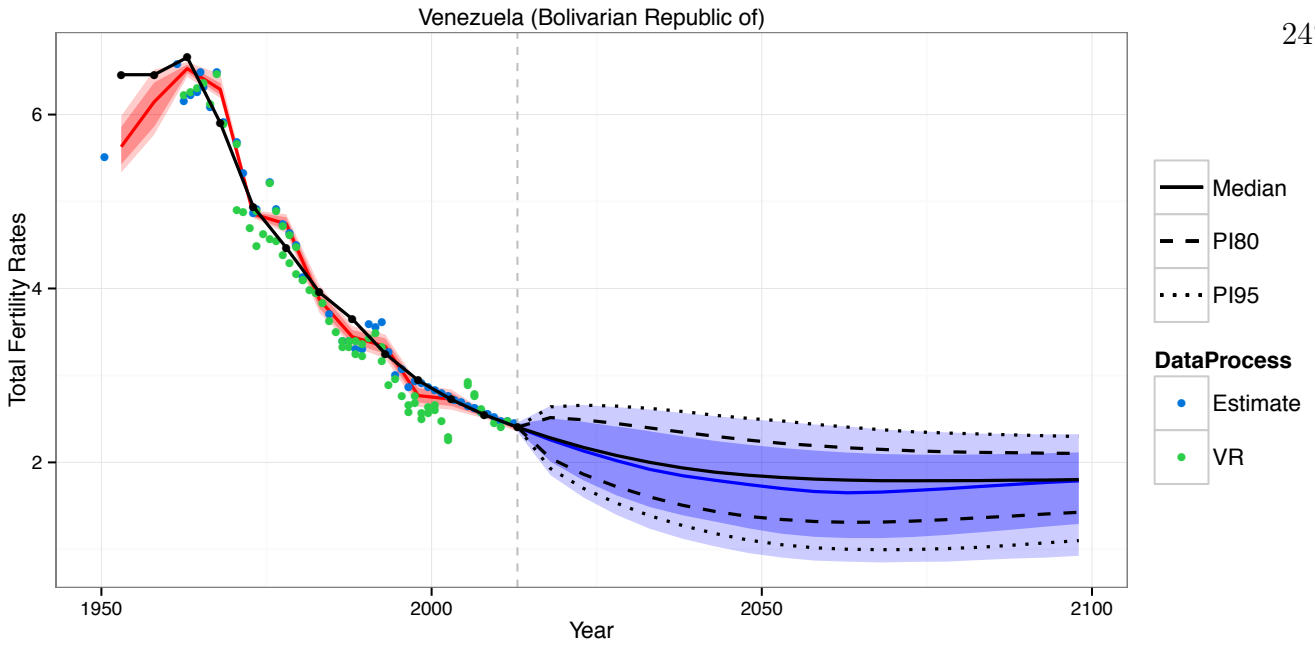

United States of America
Total Fertility Rates
3.5
3.0
2.5
2.0
1.5
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
VR

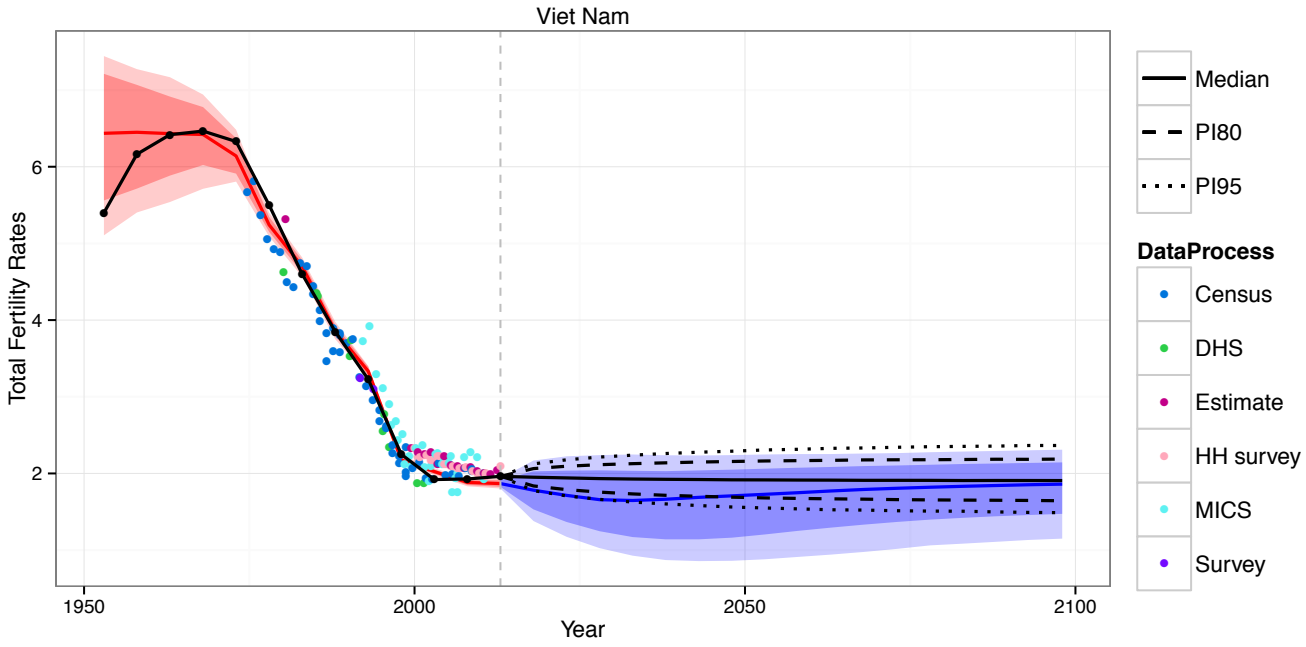

United States Virgin Islands
Total Fertility Rates
Year
1950
2000
2050
2100
1
2
3
4
5
Median
PI80
PI95
DataProcess
Estimate
VR

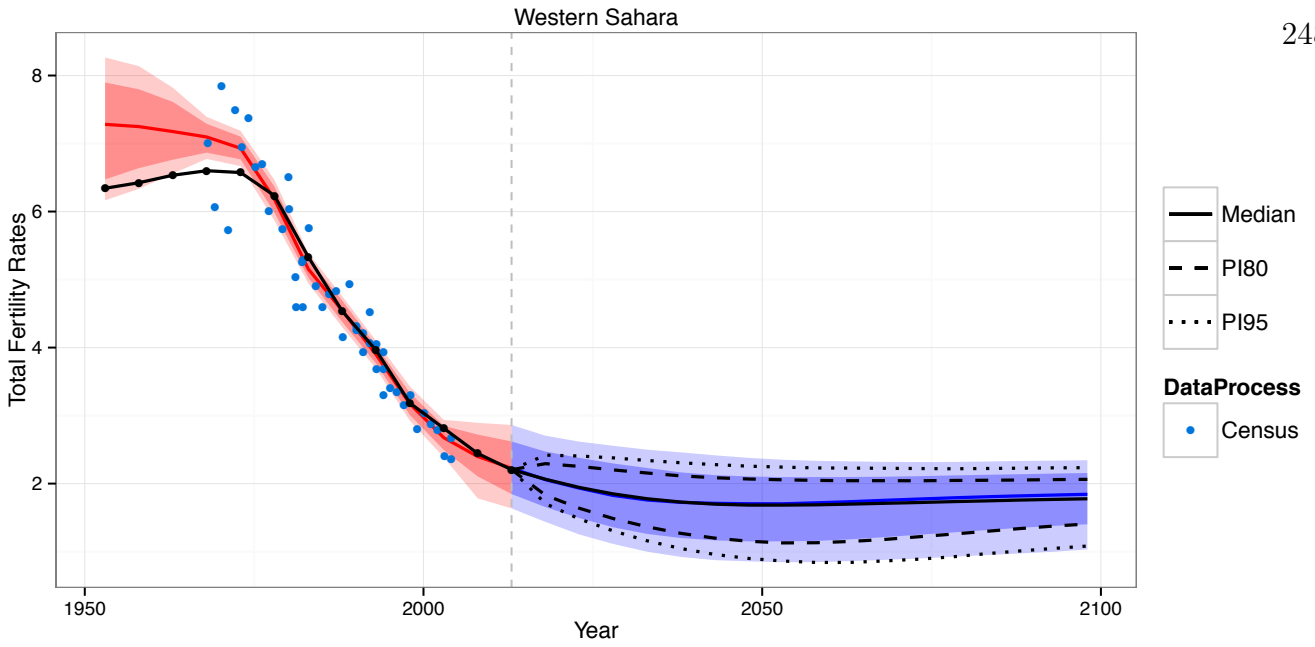

Uruguay
Total Fertility Rates
Year
1950
2000
2050
2100
1.0
1.5
2.0
2.5
3.0
Median
PI80
PI95
DataProcess
Estimate
VR

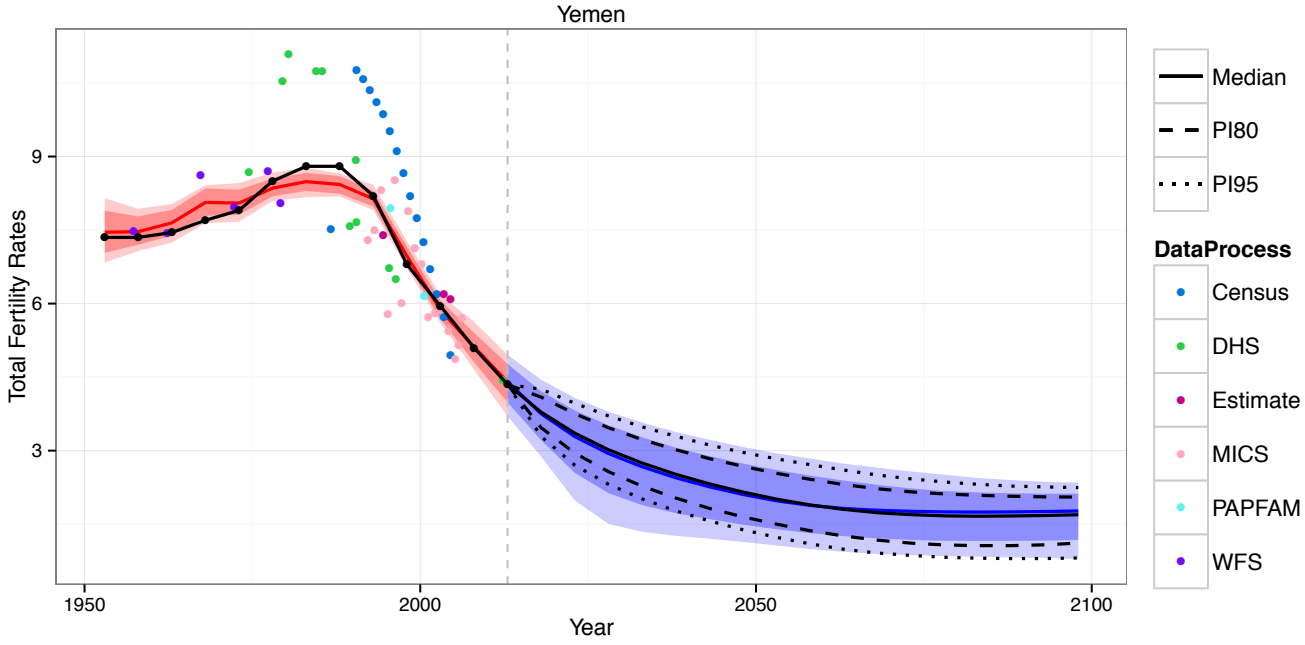

Uzbekistan
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
DHS
DHS−NS
Estimate
VR

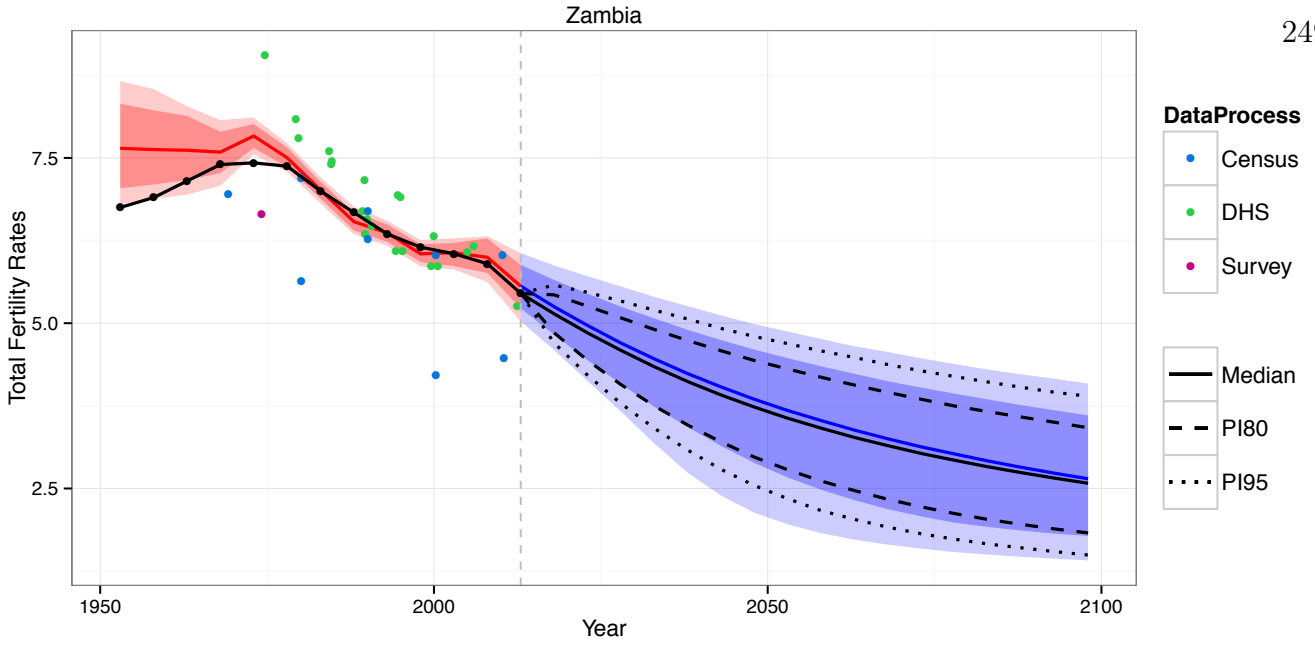

Vanuatu
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS/MICS
Estimate

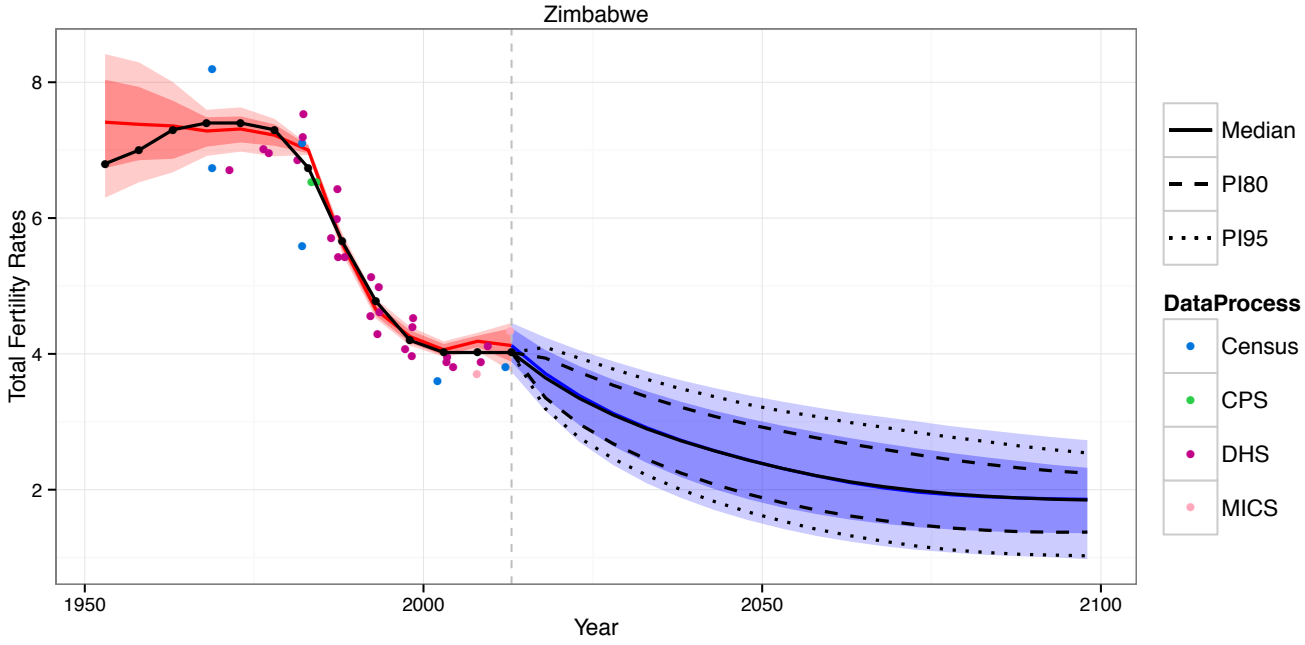

Venezuela (Bolivarian Republic of)
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Estimate
VR


Viet Nam
Total Fertility Rates
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
HH survey
MICS
Survey


Western Sahara
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census


Yemen
Total Fertility Rates
9
6
3
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
DHS
Estimate
MICS
PAPFAM
WFS


Zambia
Total Fertility Rates
7.5
5.0
2.5
1950
2000
2050
2100
Year
DataProcess
Census
DHS
Survey
Median
PI80
PI95


Zimbabwe
Total Fertility Rates
8
6
4
2
1950
2000
2050
2100
Year
Median
PI80
PI95
DataProcess
Census
CPS
DHS
MICS

Appendix B

APPENDICES TO CHAPTER 3

Model Extension

Here we provide a full description of the Bayesian hierarchical model, which was summarized in the main text for annual model. Level 1 is used if `uncertainty=TRUE`:

$$
\begin{aligned}
\text{Level 1: } & y_{c,t,s}|f_{c,t} \sim \mathcal{N}(f_{c,t}+\delta_{c,s}, \rho^2_{c,s})\,, \\
& \mathbb{E}[\delta_{c,s}] = \boldsymbol{x}_{c,s}\boldsymbol{\beta}\,, \\
& \mathbb{E}[\rho_{c,s}] = \boldsymbol{x}_{c,s}\boldsymbol{\gamma}\,; \\
\text{Level 2: } & \text{Phase I: } f_{c,t} = f_{c,t-1} + \varepsilon_{c,t}\,, \\
& \text{Phase II: } f_{c,t} = f_{c,t-1} - d_{c,t-1}\,, \\
& \texttt{ar.phase2=FALSE: } \; d_{c,t} = g_{c,t} + \varepsilon_{c,t}\,, \\
& \texttt{ar.phase2=TRUE: } \; d_{c,t} - g_{c,t} = \phi(d_{c,t-1} - g_{c,t-1}) + \varepsilon_{c,t}\,, \\
& \text{Phase III: } f_{c,t} = \mu_c + \rho_c(f_{c,t-1} - \mu_c) + \varepsilon_{c,t}\,, \\
& \boldsymbol{\theta}_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c) \\
& \varepsilon_{c,t} \sim \mathcal{N}(0, \sigma^2_{c,t})\,, \\
& g(f_{c,t}|\boldsymbol{\theta}_c) = -\frac{d_c}{1+\exp\left(-\frac{2\ln(9)}{\Delta_{c1}}\left(f_{c,t} - \sum_i \Delta_{ci} + 0.5\Delta_{c1}\right)\right)} \\
& \qquad\qquad + \frac{d_c}{1+\exp\left(-\frac{2\ln(9)}{\Delta_{c3}}\left(f_{c,t} - \Delta_{c4} - 0.5\Delta_{c3}\right)\right)}
\end{aligned}
$$

The country-specific variance, $\sigma_{c,t}$, varies according to the phase and the current fertility

level, as follows:

$$\sigma_{c,t} = c_{1975}(t)\left(\sigma_0 + (f_{c,t} - S)(-aI_{f_{c,t}>S} + bI_{f_{c,t}<S})\right) \text{ for } t \text{ is in Phase II.}$$
$$c_{1975}(t) = cI_{t\leq 1975} + I_{t>1975}\,.$$

The country-level parameters, $\{U_c, \rho_c, \mu_c, \gamma_{ci}, \Delta_{c4}, d_c\}$, are specified as follows:

$$\text{Level 3: } U_c \begin{cases} = f_{c,\tau} & \tau_c \leq 1950 \\ \sim U(\min\{5.5, \max_t\{f_{c,t}\}\}, 8.8) & \tau_c < 1950 \end{cases}$$
$$\phi_c = \log\left(\frac{d_c - 0.05}{0.5 - d_c}\right),$$
$$\phi_c \sim \mathcal{N}(\chi, \psi^2)\,,$$
$$\Delta'_{c4} = \log\left(\frac{\Delta_{c4} - 1}{2.5 - \Delta_{c4}}\right),$$
$$\Delta'_{c4} \sim \mathcal{N}(\Delta_4, \delta_4^2)\,,$$
$$p_{ci} = \frac{\Delta_{ci}}{U_c - \Delta_{c4}} \text{ for } i = 1, 2, 3\,,$$
$$p_{ci} = \frac{\exp(\gamma_{ci})}{\sum_j \exp(\gamma_{cj})}\,,$$
$$\gamma_{ci} \sim \mathcal{N}(\alpha_i, \delta_i^2)\,,$$
$$\mu_c \sim \mathcal{N}(\bar{\mu}, \sigma_\mu^2)\,,$$
$$\rho_c \sim \mathcal{N}(\bar{\rho}, \sigma_\rho^2)\,;$$

where τ_c is the starting year of phase II for country c.

The hyperparameters are $\{s_\tau, \sigma_0, a, b, S, c, \sigma_\epsilon, \chi, \psi, \Delta_4, \delta_4, \boldsymbol{\alpha}, \boldsymbol{\delta}, \bar{\mu}, \sigma_\mu, \bar{\rho}, \sigma_\rho\}$. Some of these refer to Level 2 and some to Level 3. The prior distribution of these hyperparameters is as follows (ϕ is used if `ar.phase2=TRUE`):

$$\text{Level 4: } 1/s_\tau^2 \sim \text{Gamma}(1, 0.4^2)\,,$$

$$
\begin{aligned}
\sigma_0 &\sim U[0.002, 0.6]\,, \\
a &\sim U[0, 0.2]\,, \\
b &\sim U[0, 0.2]\,, \\
S &\sim U[3.5, 6.5]\,, \\
c &\sim U[0.8, 2]\,, \\
\sigma_\epsilon &\sim U[0, 0.5]\,, \\
\chi &\sim \mathcal{N}(-1.5, 0.6^2)\,, \\
1/\psi^2 &\sim \text{Gamma}(1, 0.6^2)\,, \\
\Delta_4 &\sim \mathcal{N}(0.3, 1)\,, \\
1/\delta_i^2 &\sim \text{Gamma}(1, 1) \text{ for } i = 1, 2, 3, 4\,, \\
\alpha_1 &\sim \mathcal{N}(-1, 1)\,, \\
\alpha_2 &\sim \mathcal{N}(0.5, 1)\,, \\
\alpha_3 &\sim \mathcal{N}(1.5, 1)\,, \\
\bar{\mu} &\sim U[0, 2.1]\,, \\
\sigma_\mu &\sim U[0, 0.318]\,, \\
\bar{\rho} &\sim U[0, 1]\,, \\
\sigma_\rho &\sim U[0, 0.289]\,, \\
\phi &\sim U[0, 1]\,.
\end{aligned}
$$

Appendix C

APPENDICES TO CHAPTER 4

Table C.1: Extra percentage to achieve different objectives in Paris Agreement for top emitters. Russia is not included since currently the emission is lower than their promises in their NDCs.

Even 2°C	Likely 2°C	Even 1.5°C	country
38%	125%	203%	The United States
49%	151%	229%	Japan
25%	79%	120%	Germany
57%	160%	215%	Canada
136%	487%	875%	South Korea
90%	165%	170%	Brazil
17%	58%	97%	The United Kingdom
7%	24%	41%	China
55%	147%	191%	India