Statistical Methods for the Analysis and Prediction of Hierarchical Time Series Data with Applications to Demography

Daphne Liu

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

Doctor of Philosophy

University of Washington

2023

Reading Committee:
Adrian E. Raftery, Chair
Adrian Dobra
Elena A. Erosheva

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Statistical Methods for the Analysis and Prediction of Hierarchical Time Series Data with Applications to Demography

Daphne Liu

Chair of the Supervisory Committee:
Adrian E. Raftery
Statistics

This dissertation develops new methods for the analysis and prediction of hierarchical time series data with a focus on applications to demography.

The first two projects aim to estimate and project the potential effect that increases in education and access to family planning have on fertility decline in high-fertility countries. We first propose a new framework inspired by Granger causality for identifying the potential accelerating effect of education and family planning on fertility decline. We identify the mechanisms by which increases in education and access to family planning could lead to declines in fertility beyond what we would already expect the decline to look like based on past trends in fertility. We estimate the direct and indirect effects of education and family planning on fertility decline and explore how these effects differ within sub-Saharan Africa compared to other regions of the world. We build upon this work in the second project to propose a new method for conditional probabilistic projections of fertility given specific policy intervention outcomes for education and access to family planning. We develop a conditional Bayesian hierarchical model that creates conditional probabilistic projections of Total Fertility Rate (TFR) given probabilistic projections of women's educational attainment, contraceptive prevalence of modern contraceptive methods, and GDP per capita. The

conditional projection model enables the creation of projections corresponding to different policy intervention scenarios targeting women's educational attainment and contraceptive prevalence. We illustrate the conditional projection model by creating fertility and population projections for a range of policy intervention scenarios corresponding to meeting the United Nations Sustainable Development Goals for universal secondary education and universal access to family planning by 2030.

In the third project, we are motivated by the problem of missing data in a secondary school enrollment data set with two nonlinearly related measures of enrollment rates that have differing amounts of missing data. We propose a new method for multiple imputation of hierarchical nonlinear time series data that uses a sequential decomposition of the joint distribution and incorporates smoothing splines to account for nonlinear relationships between variables. Using a simulation study and an application to the school enrollment data, we show that the proposed method leads to substantial improvements in performance for estimation of parameters in uncongenial analysis models and for prediction of individual missing values compared to commonly used methods for multiple imputation of hierarchical time series data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

BIC: Bayesian Information Criterion.

CP: Contraceptive Prevalence.

FCS: Fully Conditional Specification.

FMI: Fraction of Missing Information.

GDP: Gross Domestic Product.

GER: Gross Enrollment Ratio.

GLS: Generalized Least Squares.

IS: Interval Score.

NER: Net Enrollment Rate.

MAE: Mean Absolute Error.

MAR: Missing At Random.

MCAR: Missing Completely At Random.

MCMC: Markov Chain Monte Carlo.

MICE: Multiple Imputation by Chained Equations.

MINTS: Multiple Imputation of hierarchical Nonlinear Time Series data.

MNAR: Missing Not At Random.

RMSE: Root Mean Squared Error.

SDG: Sustainable Development Goals.

SSA: Sub-Saharan Africa.

TFR: Total Fertility Rate.

UN: the United Nations.

WPP: World Population Prospects.

ACKNOWLEDGMENTS

I would like to thank my advisor, Adrian Raftery, for his years of mentorship and guidance. Adrian has been a constant source of encouragement since my very first quarter at the University of Washington, and I will forever be grateful for his support. It has been a privilege to learn and grow as a researcher under his tutelage.

Thank you to the members of my committee, Adrian Dobra, Elena Erosheva, and Darryl Holman, who have supported me through the dissertation process. An additional thank you to Elena for serving as my academic advisor within the Department of Statistics and for her valuable advice throughout graduate school. Thank you to the staff of the Department of Statistics, particularly Kristine Chan, Tracy Pham, Ellen Reynolds, and Asa Sourdiffe, who have helped me solve innumerable problems over the years.

Thank you to all current and former members of the Applied, Bayesian, and Computational Statistics Working Group for their helpful feedback on earlier versions of this work, without which this dissertation would not be possible. Thank you especially to Nick Irons, Yicheng Li, Peiran Liu, Hana Ševčíková, Nathan Welch, and Crystal Yu for their insight and suggestions during our PPgp group meetings.

Finally, thank you to my family and friends for your never-ending support through all the ups and downs of graduate school.

DEDICATION

to my family

Chapter 1

INTRODUCTION

1.1 Motivation

The United Nations Sustainable Development Goals (SDGs), a set of goals related to global development that were established in 2015, include targets that aim to ensure universal primary and secondary schooling and universal access to sexual and reproductive health-care services, including family planning, by 2030 (United Nations, 2015). Increasing education and access to family planning are both known to contribute to fertility decline (Hirschman, 1994). Achievement of these SDG targets is thus likely to have an impact on future fertility and population size for countries that are currently undergoing the fertility transition where fertility falls from high to low. Policymakers in high-fertility countries may be interested in assessing the potential for expanding education and access to family planning to bring about a more rapid decline in fertility in their countries, as high fertility and rapid population growth can lead to adverse economic, environmental, health, governmental, and political consequences (Bongaarts, 2013). Quantifying the potential impact of policy interventions targeting education and family planning on future fertility and population size, particularly in the context of meeting the SDG targets, could therefore be of interest for policymakers to help plan for the future infrastructure needs of their constituencies and to guide decisions regarding resource allocation.

In Chapter 2 of this dissertation, we develop methods to identify the mechanisms by which education and family planning can accelerate fertility decline and estimate their rela-

tive direct and indirect effects on fertility decline. This chapter is closely based on Liu and Raftery (2020a). In Chapter 3, we develop a Bayesian hierarchical model for probabilistic projections of fertility conditional on policy intervention outcomes for women's educational attainment and contraceptive prevalence. Using the conditional projection model, we create probabilistic projections of fertility and population under different policy intervention scenarios corresponding to meeting the SDG targets for education and family planning. This chapter is closely based on Liu and Raftery (in press).

While relatively complete expert-constructed data sets are available for some measures of education and family planning, estimates for other measures are only available sporadically across countries and times. This is a common occurrence in demography, where country-level estimates are obtained from censuses and surveys that may occur at different points in time in different countries, resulting in data sets of hierarchical time series with missing values for some country-time combinations. In the absence of an expert-driven harmonization or modeling effort to reconstruct estimates of the past, the amount of missing data can be substantial. Any analysis that seeks to make comparisons across countries and times thus needs to account for missing data.

One example where this occurs is for school enrollment rates, which are an important metric for measuring progress towards achievement of the SDG target for universal primary and secondary schooling. Two closely related measures of school enrollment are the Net Enrollment Rate (NER) and the Gross Enrollment Ratio (GER). Annual, internationally comparable estimates of both enrollment measures are available from the UNESCO Institute for Statistics. These estimates are based on survey and administrative data that have differing availability across countries and years, leading to a large number of country-years where enrollment data is missing. NER tends to have more missing values than GER, as measurement of NER requires knowledge of the age distribution for all children enrolled in school. Measurement of GER is comparatively easy as it can be calculated using only the

number of children who are enrolled in school. For analyses that are more interested in NER than GER, the greater availability of estimates of GER motivates the desire to impute missing values of NER using the strong, nonlinear relationship between NER and GER. However, existing methods for multiple imputation of hierarchical time series data can perform poorly when variables in the imputation model have a nonlinear relationship.

Chapter 4 of this dissertation develops a method for multiple imputation of hierarchical nonlinear time series data motivated by the school enrollment data, where we have an auxiliary variable that has a nonlinear relationship with the variable of interest. The proposed multiple imputation method uses a Bayesian hierarchical model to account for the hierarchical structure of the data, includes autoregressive components to account for the time series nature of the data, and incorporates smoothing splines to account for the nonlinear relationship between variables.

1.2 Background

This section summarizes background information on the relationship between education, family planning, and fertility decline, the creation of intervention-based projections of fertility and population size, and multiple imputation methods for hierarchical time series arising from survey data. Further details are available in Chapters 2–4.

1.2.1 Education and family planning as determinants of fertility decline

Education and family planning are the two main factors identified in the demographic literature that can be influenced by policy and may contribute to fertility decline (Axinn and Barber, 2001; Hirschman, 1994). Increased education is thought to contribute to fertility decline through two main mechanisms. First, education may increase the opportunity cost for women of having children. This first mechanism is measured by the educational attainment of women. More educated women tend to have higher status and access to opportunities,

thus increasing the opportunity costs of childbearing (Easterlin and Crimmins, 1985). Increased parental schooling may also change the value placed on large family sizes and spread information about family planning (Axinn and Barber, 2001). Second, education may accelerate fertility decline by increasing the cost of raising children. This second mechanism is measured by the enrollment rates of children. Children's schooling reduces their capacity for work and increases the cost of childrearing (Caldwell, 1982; Caldwell et al., 1985). The increased cost of raising children and the "quantity-quality tradeoff" may both play a role when making childbearing decisions (Axinn and Barber, 2001; Easterlin and Crimmins, 1985).

Access to family planning can also be influenced by policy and can contribute to fertility decline. Studies have consistently found a strong, negative association between contraceptive prevalence and fertility (Bongaarts, 2010; Tsui, 2001). Although education and other factors may change fertility preferences, family planning is required to translate those changed preferences into changes in fertility. Contraceptive use, one of the most important proximate determinants of fertility (Bongaarts, 1987), can therefore provide a means by which individuals can attain their desired childbearing.

Education and family planning may also have a combined impact on fertility decline. There is a well-documented positive association between education and family planning (Ainsworth et al., 1996; Bongaarts, 2010; Kirk and Pillet, 1998), where more educated women have a higher demand for and greater use of family planning, though these differentials in contraceptive use between education groups have been found to be smaller when the overall contraceptive prevalence in a country was higher Martín (1995). There may be an indirect effect of women's education on fertility decline through family planning, as increased education may impact fertility by increasing knowledge of family planning (Cochrane, 1979) or by changing attitudes towards its acceptability (Cleland and Wilson, 1987).

This dissertation aims to answer questions about the nature of the potential accelerating

effect of education and family planning on fertility decline in the high-fertility context. In Chapter 2, we identify the mechanisms by which education and family planning can accelerate fertility decline. We investigate the relative effects of education and family planning on fertility decline, with a focus estimating the direct and indirect effects of both quantities on fertility decline. We also explore the effects that education and family planning may have on fertility decline in sub-Saharan Africa compared to other world regions, as there is evidence to suggest the demographic transition in sub-Saharan Africa may be different from those seen in Latin America and the Caribbean and in Asia (Bongaarts and Casterline, 2013; Bongaarts et al., 2017; Grant, 2015). In Chapter 3, we build upon the findings of Chapter 2 and specify the assumptions needed for a causal interpretation of the estimated accelerating effects of education and family planning on fertility decline.

1.2.2 Intervention-based projections of fertility and population

The United Nations (UN) has produced estimates and projections of world population by country since 1951 and remains the premier producer of global demographic projections, with projections from the UN used by policymakers around the world. Governments and agencies in countries that do not have robust vital registration systems of their own often rely on the UN estimates and projections to inform planning and policy decisions. The UN does not currently produce projections for policy intervention scenarios, although projection variants (low, medium, and high) based on different underlying demographic assumptions are available. The low and high projection variants for fertility correspond to assuming the Total Fertility Rate (TFR) will be, respectively, half a child below or half a child above the medium variant TFR. While these variants can provide some guidance to policymakers, they have the drawback of being deterministic. The variants also have the drawback of being in terms of the TFR, rather than in terms of variables that can be more directly influenced by policy.

The UN projection model for TFR does not explicitly incorporate the effect of covariates; instead, the model implicitly captures the effect of covariates on fertility by modeling future TFR based on historical trends in TFR. However, to better provide guidance to policymakers, it may be important to incorporate these covariates explicitly into fertility projection models (Lutz et al., 2014).

Two other producers of global demographic projections, the Wittgenstein Centre for Demography and Global Human Capital and the Institute for Health Metrics and Evaluation (IHME), do produce scenario-based projections of fertility and population (Abel et al., 2016; Lutz et al., 2014; Vollset et al., 2020). The projections from the Wittgenstein Centre and IHME include scenarios based on policy interventions corresponding to attaining the SDGs in 2030. However, the projections from the Wittgenstein Centre and IHME either do not fully incorporate uncertainty or do not fully incorporate demographic knowledge. The Wittgenstein Centre produces projections corresponding to the Shared Socioeconomic Pathways (SSPs) used by the Intergovernmental Panel on Climate Change. While the different SSP scenarios lead to a range of possible population projections, the individual projections for each scenario are deterministic. Unlike the projections from the Wittgenstein Centre, the projections from IHME do come with associated measures of uncertainty. However, the IHME projection methodology has been criticized in the demographic community for questionable model assumptions and demographically implausible projection results (Alkema, 2020; Gietel-Basten and Sobotka, 2020, 2021). As there are substantial differences between the projections produced by the UN and the reference scenario projections produced by both the Wittgenstein Centre and IHME, the policy intervention scenario projections from the other sources cannot be directly compared with the UN projections. Demographic projections based on policy intervention scenarios using the UN methodology are thus of interest.

This dissertation aims to address this gap in Chapter 3, where we develop a conditional probabilistic projection model for TFR that extends the probabilistic fertility projec-

tion model used by the UN. The conditional TFR projection model explicitly accounts for women's educational attainment, contraceptive prevalence of modern methods, and GDP per capita and allows for the creation of probabilistic projections of TFR that are conditional on policy-based intervention scenarios related to education and family planning. We illustrate the conditional projection method by creating projections of fertility and population size given policy intervention scenarios corresponding to meeting the SDG targets for universal secondary education and universal access to family planning and compare our results with the results from the Wittgenstein Centre and IHME.

1.2.3 Multiple imputation for hierarchical time series arising from survey data

Multiple imputation is a widely used approach for handling missing data that was first developed by Rubin (1978, 1987) as a way to create imputed values for missing responses in public-use releases of data sets from sample surveys. In multiple imputation, $M > 1$ imputed values for missing observations are sampled from the posterior predictive distribution of the missing data given the observed data. Unlike in single imputation, the imputed values from multiple imputation incorporate both sampling variability and uncertainty about the imputation model. The M imputed values are used to create M completed data sets, each of which consists of the observed data and one set of imputed values for the missing observations. The completed data sets are analyzed separately using complete data methods and the results of the analyses are combined into one final, pooled result using combining rules from Rubin (1987) that account for both within-imputation variation and between-imputation variation.

For hierarchical data, multiple imputation approaches that do not explicitly account for the hierarchical structure of the data have been found to lead to biased results in downstream analyses (Enders et al., 2016; Lüdtke et al., 2017; Taljaard et al., 2008). Many approaches specifically for multiple imputation of hierarchical time series data have been developed (e.g. Enders et al., 2020; Grund et al., 2021; He et al., 2011; Liu et al., 2000; Speidel et al., 2018;

among others), with two of the most widely used approaches for social science data being Amelia II and multilevel extensions of Multiple Imputation by Chained Equations (MICE). Amelia, originally developed by King et al. (2001) and extended as Amelia II by Honaker and King (2010), is a multiple imputation method designed specifically for hierarchical time series data. Amelia is based on the joint modeling approach to multiple imputation, where imputed values are sampled from a joint distribution for all variables with missing data. Amelia assumes the complete data follow a multivariate normal distribution and estimates the imputation model using a combination of bootstrapping and an EM algorithm. MICE is a multiple imputation method developed by van Buuren and Groothuis-Oudshoorn (2011) that uses the fully conditional specification (FCS) approach to multiple imputation. Rather than explicitly specifying a joint imputation model, the FCS algorithm iteratively samples from univariate conditional imputation models for all variables with missing data until convergence is reached. Several methods that account for hierarchical data structures have been implemented within the MICE framework, including the method developed by Schafer and Yucel (2002). Using the Schafer and Yucel method within the MICE framework specifies univariate conditional distributions for each variable with missing data, where the conditional distributions are modeled as linear mixed effects models with homogeneous within-group variances. The commonly used Amelia and MICE methods assume that variables in the imputation model have a linear relationship. In settings where variables with missing data have a strong nonlinear relationship and transformation to approximate linearity is not possible, these methods are misspecified and can result in poor imputations.

An appealing feature of multiple imputation is the ability to use the same imputed data set to conduct many different analyses (Rubin, 1987; Schafer, 1997a). This means that public-use releases of survey data can be published with imputed values that can then be used in analyses by external researchers. However, the external researchers may not have access to full information about how the imputed values were created, for example if the

predictive variables used in the imputation model are not publicly available. This can lead to uncongeniality in the sense of Meng (1994) if the imputation model used to impute the data and the analysis models used by the external researchers make different assumptions about the data. More formally, an imputation model and an analysis model are congenial if there exists a Bayesian model such that (i) the posterior mean and variance from the Bayesian model for parameters of interest are asymptotically the same as the mean and variance estimates from the analysis model and (ii) the posterior predictive distribution of the missing data given the observed data derived from the Bayesian model is identical to the imputation model. Many of the theoretical properties of multiply imputed estimates, such as the consistency of estimates using Rubin's combining rules, rely on congeniality (Meng, 1994; Rubin, 1996; Xie and Meng, 2017). In practice, uncongeniality of the imputation and analysis models is a regular occurrence, especially when imputation and analysis are conducted in independent phases. The ability of a multiple imputation method to perform well for uncongenial analyses is thus of interest for practitioners.

We propose a new method for multiple imputation of hierarchical time series data in Chapter 4 of this dissertation that can accommodate nonlinear relationships between variables with missing data. We focus on the bivariate setting where one auxiliary variable is used to impute the variable of interest and the variable of interest has a larger amount of missing data than the auxiliary variable. This setting is motivated by a school enrollment rate data set that includes estimates of both NER and GER, where NER is the variable of interest for a substantive analysis model but suffers from a greater amount of missing data than GER. We compare the proposed multiple imputation method with the Amelia and MICE methods using a simulation study and an application to the school enrollment data and find the proposed method can lead to better performance for estimation of uncongenial analysis models when the variables with missing data have a nonlinear relationship.

1.3 Key Contributions

This dissertation develops methods for the analysis and prediction of hierarchical time series data with applications to demography. The main statistical contributions of this dissertation are summarized in this section.

1.3.1 Identification of the mechanisms through which education and family planning can accelerate fertility decline

In Chapter 2, we introduce the concept of an "accelerating effect" on fertility decline inspired by the philosophy of Granger causality (Granger, 1969), where measures of education and family planning are said to have an accelerating effect on fertility decline if they are associated with declines in fertility beyond what we would already expect the decline to look like based on past trends in fertility. We use this framework to identify the mechanisms by which education and family planning can accelerate fertility decline. We estimate the direct and indirect accelerating effects of these mechanisms on fertility decline and assess if the effect sizes are different in sub-Saharan Africa compared to other world regions.

1.3.2 Development of a conditional Bayesian hierarchical model for probabilistic projections of TFR given policy intervention scenarios

In Chapter 3, we develop a Bayesian hierarchical model for conditional probabilistic projections of TFR that extends the probabilistic fertility projection model used by the UN, which was originally developed by Alkema et al. (2011). The conditional TFR projection model explicitly incorporates women's educational attainment, contraceptive prevalence of modern methods, and GDP per capita as covariates and enables the creation of projections of TFR that are conditional on hypothetical policy intervention scenarios related to education and family planning. Through an out-of-sample validation exercise, we show the proposed model is able to create conditional TFR projections without compromising on the predictive

accuracy of the projection model used by the UN. We additionally build upon the accelerating effect framework developed in Chapter 2 to identify the assumptions needed for a causal interpretation of policy intervention scenario projections created using the conditional TFR projection model.

1.3.3 Development of a Bayesian method for multiple imputation of hierarchical nonlinear time series data

In Chapter 4, we develop a Bayesian method for multiple imputation of hierarchical time series data that can accommodate nonlinear relationships between variables with missing data. The proposed multiple imputation method uses a combination of smoothing splines, country-specific intercepts, and time series methods to create imputed values in the bivariate setting where one auxiliary variable is used to impute the variable of interest and the variable of interest has a larger amount of missing data than the auxiliary variable. Using a simulation study and an application to a school enrollment rate data set, we show the proposed multiple imputation method can lead to substantial improvements for estimation of parameters in uncongenial analysis models and for prediction of individual missing values compared to previously existing methods for multiple imputation of hierarchical time series data.

1.4 Outline of the Dissertation

This dissertation is organized as follows. Chapter 2 identifies the precise measures of education and family planning that have a potential accelerating effect on fertility decline. The direct and indirect effects of these measures on fertility decline are estimated and the potentially differential effect of these measures within sub-Saharan Africa compared to other world regions is investigated. Chapter 3 develops a Bayesian hierarchical model for conditional probabilistic projections of TFR given policy intervention outcomes for women's educational attainment and contraceptive prevalence. The conditional projection method is

illustrated using policy intervention scenarios based on meeting the UN SDGs targets related to education and family planning. Chapter 4 describes the multiple imputation method for hierarchical nonlinear time series data with an application to a school enrollment rate data set and compares the proposed method with several commonly used multiple imputation methods through simulation studies. Finally, Chapter 5 summarizes the methods proposed in this dissertation and offers ideas for future research.

Chapter 2

HOW DO EDUCATION AND FAMILY PLANNING ACCELERATE FERTILITY DECLINE?

In this chapter, we develop a new framework inspired by Granger causality for estimating the potential accelerating effect of education and family planning on fertility decline in the high-fertility setting. We use this framework to identify the precise nature of the accelerating effect: Does the effect of education operate through increasing educational attainment of women or educational enrollment of children? At which educational level is the effect strongest? Does the effect of family planning operate through increasing contraceptive prevalence or reducing unmet need? Is education or family planning more important? We find that women's educational attainment of lower secondary education is key to accelerating fertility decline and also find an accelerating effect of contraceptive prevalence for modern methods. We find the impact of contraceptive prevalence to be substantially larger than that of education. While the accelerating effects still exist in sub-Saharan Africa, we find the effect sizes are smaller in sub-Saharan Africa than in other world regions.

This chapter is closely based on the article "How Do Education and Family Planning Accelerate Fertility Decline?" published in *Population and Development Review* (Liu and Raftery, 2020a).

2.1 Introduction

The United Nations projects that world population will increase from its present 7.7 billion to 10.9 billion people in 2100, with more than half of this increase in sub-Saharan Africa (SSA), mostly in high-fertility countries (United Nations, 2019c). Much of the rest of the

increase will be in countries in Asia and Latin America with above-replacement fertility. It is widely thought that these countries would benefit from a slower population increase brought about by a more rapid decrease in fertility, as high fertility and rapid population growth are likely to have adverse economic, environmental, health, governmental, and political consequences (Bongaarts, 2013). Declining fertility can also yield a demographic dividend by reducing the dependence ratio, increasing women's participation in the paid labor force, and allowing increased investments in human and physical capital (Lee and Mason, 2006; Mason and Lee, 2006). This raises the question of how the fertility decline could be accelerated in high-fertility countries. There is widespread agreement in the literature that there are two main factors that can be influenced by policy and may help accelerate fertility decline: education and family planning (Hirschman, 1994).

Increased education is thought to accelerate fertility decline through two main mechanisms (Axinn and Barber, 2001; Hirschman, 1994). The first is by increasing the opportunity cost for women, measured by their educational attainment, of having children. Demand or structural theories of fertility decline, such as Easterlin and Crimmins (1985), argue that educated women have higher status and access to opportunities, thus increasing the opportunity costs of childbearing. Ideational theories argue that increased parental schooling changes the value placed on large family sizes and spreads information about family planning, increases consumption aspirations, and spreads Western family values (Axinn and Barber, 2001). Many studies on the relationship between education and fertility, particularly on the potentially causal nature of this relationship, focus on the mechanism of women's educational attainment (e.g. Behrman, 2015; Bongaarts et al., 2017; Cygan-Rehm and Maeder, 2013; Martín, 1995; Osili and Long, 2008; Raftery et al., 1995). Women's educational attainment is the mechanism proposed by Lutz et al. (2014) in their argument for the importance of including education in population projections.

Education may also accelerate fertility decline by increasing the cost of raising children,

which can be measured via the enrollment rates of children. The intergenerational wealth flows theory (Caldwell, 1982; Caldwell et al., 1985) argues for the importance of children's education, stating that children's schooling reduces their capacity for work and increases the cost of childrearing. Microeconomic theories based on the "quantity-quality tradeoff" have also emphasized the role of children's enrollment in shaping parents' future childbearing decisions (Axinn and Barber, 2001). Easterlin and Crimmins (1985) also acknowledge the role the cost of raising children may play in parents' future childbearing decisions. Studies that evaluate the impact of children's enrollment on fertility decline include Raftery et al. (1995), Subbarao and Raney (1995), and Masih and Masih (2000).

Family planning is the second quantity that can be influenced by policy and may accelerate fertility decline. Although education and other factors may change fertility preferences, family planning is needed to translate those changed preferences into changes in fertility. As one of the most important proximate determinants of fertility (Bongaarts, 1987), contraceptive use can provide a means by which individuals can attain their desired childbearing. Studies have consistently found a strong negative association between contraceptive prevalence and fertility (Tsui, 2001).

While the potential mechanisms by which increased education or family planning may affect fertility decline are well known, there is less consensus on the relative impact the different mechanisms may have. There is a well-documented positive association between education and family planning (Ainsworth et al., 1996; Bongaarts, 2010; Kirk and Pillet, 1998), where more educated women have a higher demand for and greater use of family planning. Using DHS data, Martín (1995) found that differentials in contraceptive use between education groups tended to be smaller when the overall contraceptive prevalence in a country was higher. Masih and Masih (2000) found that both the female secondary gross enrollment ratio and contraceptive use as measured by female sterilization had a significant impact on fertility in India in 1965–1991, where the combined effect of education and family

planning explained, in a Granger-causal sense, a substantial portion of the variability in the total fertility rate (TFR). However, Masih and Masih found that only education was exogenous. Through simulations and using data from the 1993 Indonesia Family Life Survey, Angeles et al. (2005) found a larger effect of family planning programs on reducing fertility than improvements in school quality.

An additional consideration when evaluating the relative effects of education and family planning on fertility decline is that contraception is a proximate determinant of fertility while education is not. Women's education has been hypothesized to have an indirect effect on fertility decline through family planning, as increased education may impact fertility by increasing knowledge of family planning (Cochrane, 1979) or by changing attitudes towards its acceptability (Cleland and Wilson, 1987). Evidence of this indirect effect has been seen in Indonesia, where Gertler and Molyneaux (1994) found much of the effect of increased women's educational attainment on fertility decline in the 1980s was an indirect effect through contraceptive use.

Figure 2.1 shows trends in the TFR, the percentage of women who have attained lower secondary education or higher, and the contraceptive prevalence of modern contraceptive methods for Kenya and Nigeria from 1975–1980 to 2010–2015. The decline in TFR from 1970 onwards has been faster in Kenya than in Nigeria. Correspondingly, we see faster growth in women's educational attainment and contraceptive prevalence in Kenya than in Nigeria. Notably, educational attainment and contraceptive prevalence have followed similar growth trajectories in Kenya while educational attainment has grown at a faster rate than contraceptive prevalence in Nigeria.

We are interested in whether the effect of education operates primarily through increased educational attainment of women or through increased educational enrollment of children. Identifying at which educational level the impact is strongest is also of interest, especially from a policy standpoint. We also aim to evaluate whether the impact of family planning

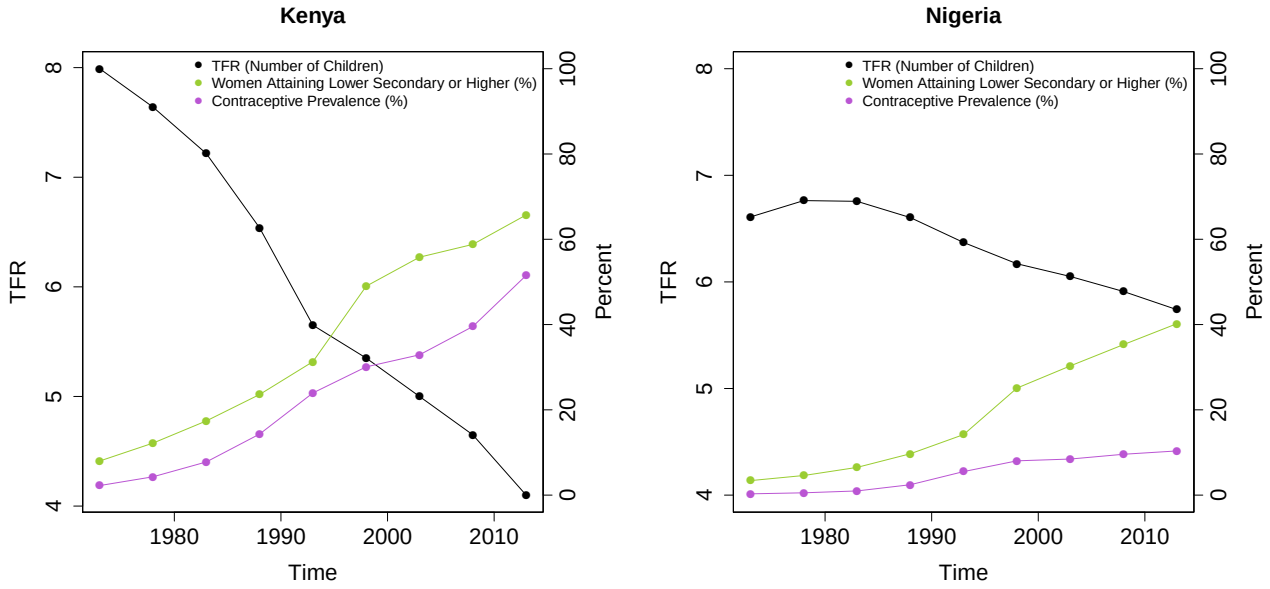


Figure 2.1: TFR (black), percentage of women who have attained at least lower secondary education or higher (green), and contraceptive prevalence (purple) for Kenya from 1970–1975 to 2010–2015 and Nigeria from 1975–1980 to 2010–2015.

on fertility decline operates by reducing unmet need for family planning or by increasing contraceptive prevalence. This distinction, while subtle, is vital for crafting effective family planning policies.

Finally, we are interested in quantifying the relative impacts of education and family planning on fertility decline. We will evaluate if increasing education or family planning accelerates fertility within a high-fertility context; that is, whether increases in education or family planning correspond to declines in TFR faster than what we would already expect given historical trends. If so, we aim to identify which education mechanism, level of education, and measure of family planning has the strongest effects.

We also explore the accelerating effect of covariates on fertility decline within SSA compared to other regions of the world. The SSA fertility transition has been slower than historical fertility transitions in Asia and Latin America, and fertility decline has even stalled in many parts of SSA (Bongaarts and Casterline, 2013). Countries in SSA may be experiencing different relationships between education, family planning, and fertility compared to other historically high-fertility regions. There is evidence of differences in ideal family size, which may diminish the effect of family planning in SSA (Bongaarts and Casterline, 2013; Bongaarts et al., 1984). There also appear to be differences in school quality, which may diminish the effect of education in SSA (Grant, 2015). We do not explicitly use measures of ideal family size or school quality, but these hypotheses for the SSA difference motivate our work.

2.2 Data

We use estimates of TFR from the United Nations *World Population Prospects* (WPP) 2019 Revision (United Nations, 2019c), which is available for 201 countries by five-year time periods. As we are interested in estimating relationships spanning all current and historical high-fertility transitions, we need estimates of TFR that are comparable across countries

and time periods. As recommended by Bongaarts (2017), we use estimates of TFR from the United Nations. The estimates of TFR from WPP are based on vital registers, censuses, and surveys such as the DHS and the multi-indicator cluster surveys. These individual data sources are available on an uneven basis across countries and across time and often have known biases or data quality issues that need to be adjusted for. For example, DHS fertility estimates suffer from inconsistent data quality across countries, often due to misreporting or omission of recent births (Schoumaker, 2014). The WPP estimates account for these adjustments and allow us to use information drawing from multiple data sources.

We denote the TFR for country c in five-year time period t by $f_{c,t}$. Decrements in TFR are constructed as a measure of fertility decline, with the TFR decrement from five-year time period $(t-1)$ to five-year time period t defined as $\Delta f_{c,t} = f_{c,t-1} - f_{c,t}$. This assigns larger positive values to the TFR decrement when fertility is declining faster. As we are mainly interested in changes to the rate of fertility decline, we focus on modeling the TFR decrement. Our outcome variable is thus on the timescale of differences in five-year time periods. Correspondingly, we expect any covariates we add to the model to be on the scale of changes over time.

We construct these changes so they are positive when the covariate is "improving." For example, if an education covariate X is increasing over time on average, we define the corresponding change as $\Delta X_{c,t} = X_{c,t} - X_{c,t-1}$. This ensures that $\Delta X_{c,t}$ is positive when education is increasing in country c. Changes over time in education, family planning, urbanization, and GDP variables were defined analogously to $\Delta X_{c,t}$. The change over time in child mortality $({}_5q_0)$ was defined to be in the same direction as the TFR decrement. For country c and five-year time periods $(t-1)$ and t, the decrement in child mortality was constructed as $({}_5q_0)_{c,t-1} - ({}_5q_0)_{c,t}$.

We identify a "high-fertility transition" subset of our data to serve as the main focus of our analyses. For each country, we are primarily interested in data corresponding to time

periods where the country was in Phase II of the fertility transition as defined by Alkema et al. (2011) and had TFR greater than 2.5. This results in a subset of 666 country-time period pairs with observations from 121 countries. The earliest time period we have data for is 1970–1975.

We consider two measures of education: women's educational attainment and children's enrollment. Educational attainment data for women aged 20–39 were obtained from the Wittgenstein Centre (2018). The Wittgenstein Centre provides a harmonized dataset of the educational attainment distribution using six levels of attainment: no education, incomplete primary, primary, lower secondary, upper secondary, and postsecondary. These attainment levels are constructed to be comparable across countries and times and are based on the International Standard Classification of Education. We focus on cumulative levels of attainment such as the proportion of women who completed primary education or higher. Data on children's enrollment are obtained from the World Bank World Development Indicators (World Bank, 2019). Net Enrollment Rates (NER) for primary and secondary education for both sexes combined are available from 1970 onwards. Missing NER values are imputed using a combination of a piecewise LOESS curve based on the gross enrollment ratio, also from the World Bank, and linear interpolation.

We consider the median estimates of contraceptive prevalence and unmet need for family planning from the United Nations *Estimates and Projections of Family Planning Indicators 2019* (United Nations, 2019a), based on the methodology of Alkema et al. (2013). All indicators are available for married or in-union women aged 15–49 years beginning from 1970. Contraceptive prevalence and unmet need are reported as percentages of the total number of married or in-union women. We convert these percentages to proportions between 0 and 1 for analyses.

Finally, we consider several control variables. Estimates of child mortality (${}_5q_0$) are obtained from WPP 2019, where we exclude mortality data from the time periods corresponding

to the genocides in Cambodia and Rwanda. We also consider measures of GDP per capita growth (as percent growth) from the World Bank and the percentage of population residing in urban areas from the UN World Urbanization Prospects 2018 (United Nations, 2018). GDP and urbanization measures are converted to proportions between 0 and 1 for analyses.

Examples of trends in TFR, women's educational attainment, children's enrollment, and family planning indicators for Nigeria and Kenya can be seen in Figures 2.2 and 2.3, respectively. We see that Nigeria has experienced a slow but steady decrease in TFR over time. Increases in women's educational attainment have mostly occurred from the 1990s onwards, though increased enrollment rates are seen early on. Nigeria has experienced relatively small improvements in family planning indicators for modern contraceptive methods since 1970. Kenya is an example of a sub-Saharan African country that has experienced a more rapid fertility decline than Nigeria. There have also been larger increases in women's educational attainment, particularly of lower secondary education, and larger increases in access to modern methods of family planning in Kenya than in Nigeria. However, the increase in enrollment rates in secondary education has been notably delayed in Kenya.

2.3 Methodology

2.3.1 Modeling framework

Our methodology draws inspiration from Granger causality to answer questions about how covariates may affect the acceleration of fertility decline. Granger causality is based on the assumption that the cause must temporally precede the effect. A covariate X is said to "Granger-cause" the outcome Y if X can provide additional information for forecasting Y that is not already captured in past values of Y (Granger, 1969). Following this logic, to investigate if education or family planning covariates have an accelerating effect on fertility decline beyond what we would already expect the decline to look like based on past trends, we need to include a measure of the "expected fertility decline" in our model. This measure

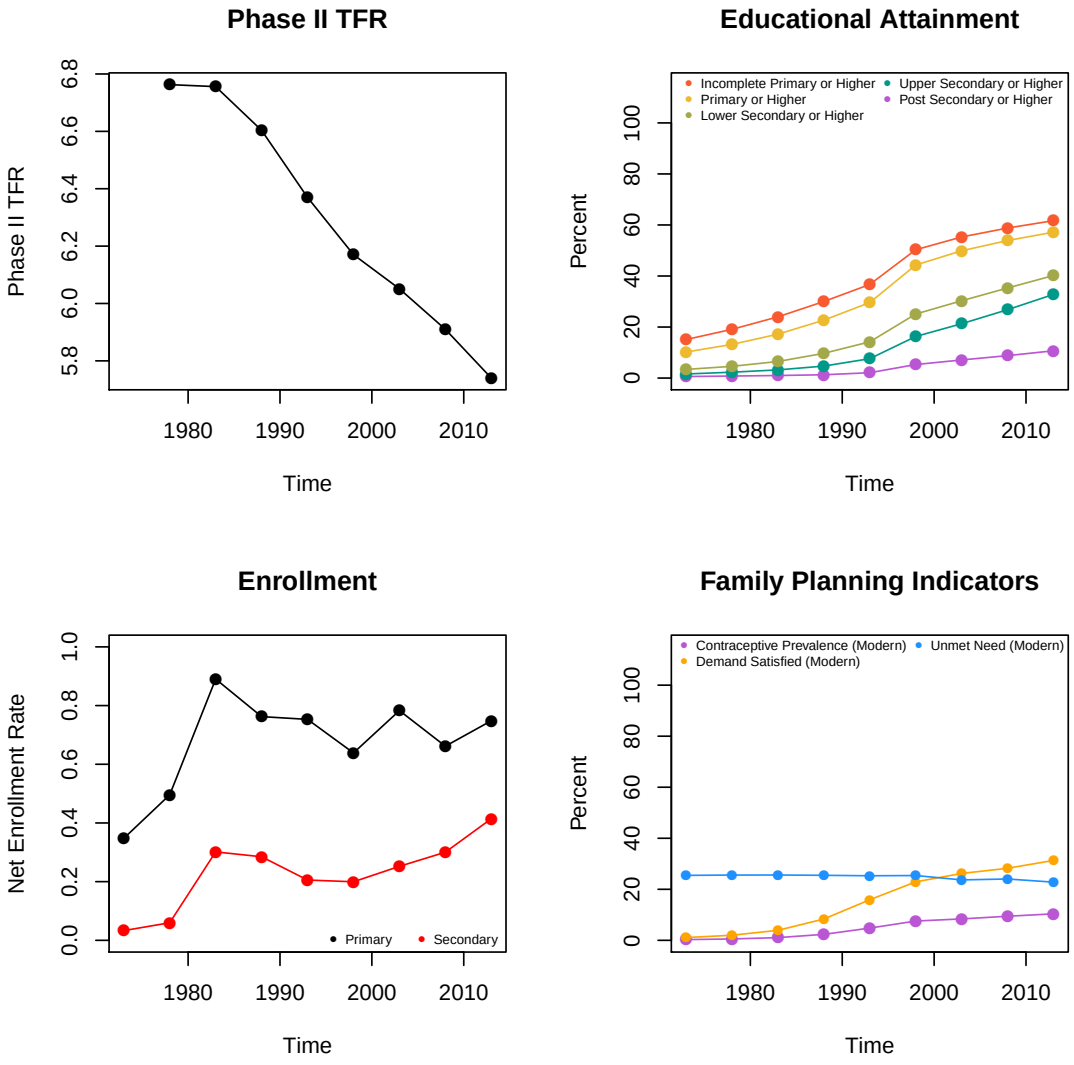


Figure 2.2: Trends in Phase II TFR, cumulative educational attainment, NER, and family planning indicators for modern methods in Nigeria from 1975–1980 to 2010–2015

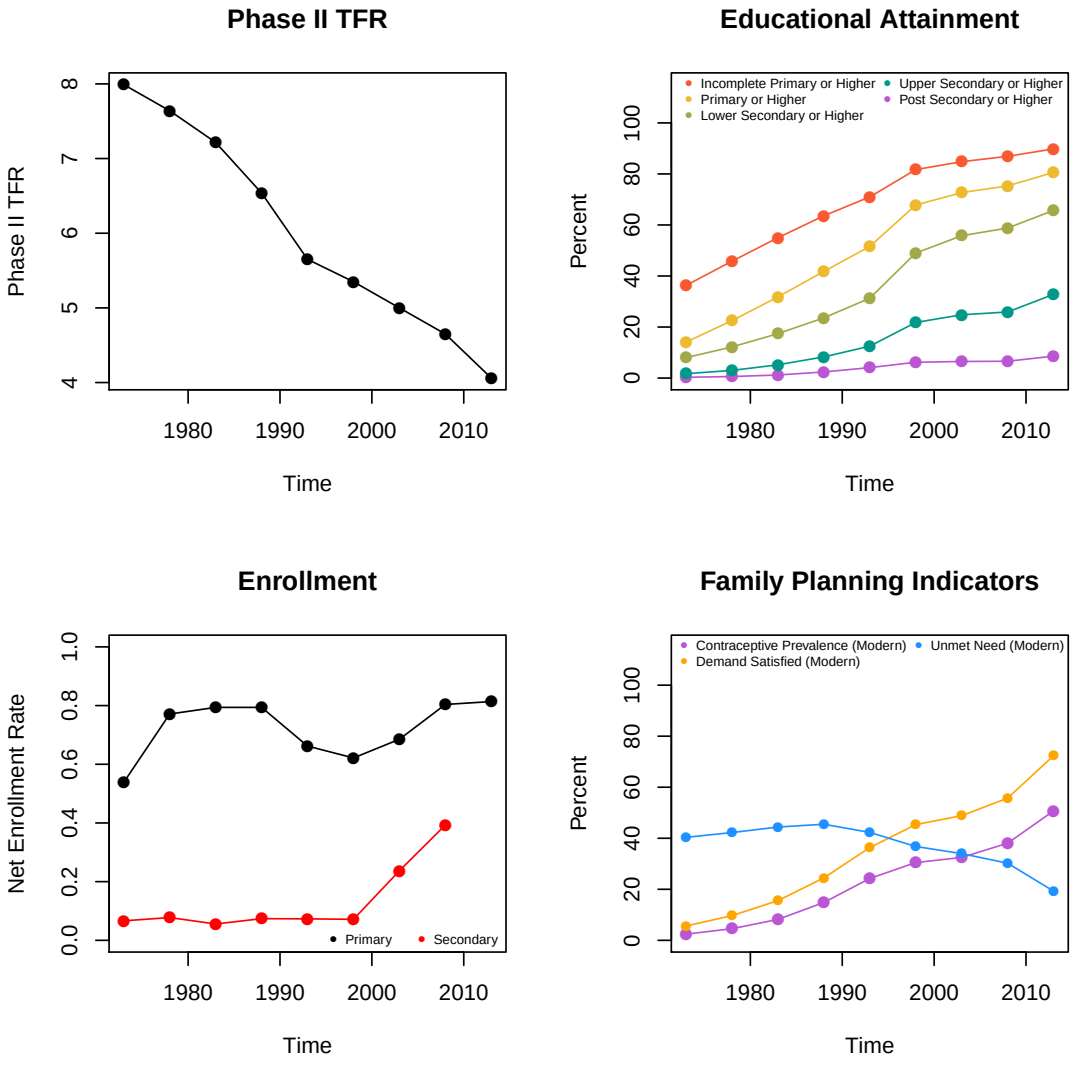


Figure 2.3: Trends in Phase II TFR, cumulative educational attainment, NER, and family planning indicators for modern methods in Kenya from 1975–1980 to 2010–2015

of expected decline should be based on past TFR trends and be both country- and time-specific. We draw from the Bayesian hierarchical model that is the basis of the model for probabilistic fertility projections currently used by the UN (United Nations, 2019d; Alkema et al., 2011; Fosdick and Raftery, 2014; Raftery et al., 2014).

In the Bayesian hierarchical model, the expected TFR decrement from five-year time period $(t-1)$ to five-year time period t is modeled as a double logistic function. For country c and time period t, the double logistic function is defined as

$$g(\theta_c, f_{c,t}) = \frac{-d_c}{1+\exp\big(-\frac{2\ln(9)}{\triangle_{c,1}}(f_{c,t}-\sum_{i=1}^{4}\triangle_{c,i}+0.5\triangle_{c,1})\big)} + \frac{d_c}{1+\exp\big(-\frac{2\ln(9)}{\triangle_{c,3}}(f_{c,t}-\triangle_{c,4}-0.5\triangle_{c,3})\big)}.$$

In the double logistic function, $\theta_c = (\triangle_{c,1}, \triangle_{c,2}, \triangle_{c,3}, \triangle_{c,4}, d_c)$ is a vector of country-specific parameters. For country c, the parameter d_c represents the maximum possible five-year TFR decrement. The parameters $\triangle_{c,i}$ for $i = 1, ..., 4$ describe the range of TFR values in which the pace of the fertility decline changes. Specifically, the start of the fertility transition occurs at TFR level $U_c = \sum_{i=1}^{4}\triangle_{c,i}$. At this TFR, the pace of the decline is around $0.1d_c$. From TFR levels U_c to $U_c - \triangle_{c,1}$, the pace of the fertility decline increases to at least $0.8d_c$. The pace of fertility decline is the highest for the TFR values denoted by $\triangle_{c,2}$, where it ranges from $0.8d_c$ to d_c. During the TFR range $\triangle_{c,3}$, the pace of fertility decline decreases and by TFR level $\triangle_{c,4}$ the pace of decline has decreased to $0.1d_c$.

The expected TFR decrement is incorporated into the Bayesian hierarchical model

$$\begin{aligned} f_{c,t} &= f_{c,t-1} - g(f_{c,t-1}|\theta_c) + \varepsilon_{c,t} \\ \varepsilon_{c,t} &\overset{iid}{\sim} N(0, \sigma(f_{c,t-1})^2)) \\ \theta_c &\sim h(\cdot, \phi) \\ \phi &\sim \pi(\cdot), \end{aligned}$$

where the country-specific parameter vector θ_c follows a world distribution $h(\cdot|\phi)$ with parameter ϕ. The prior distribution of ϕ is $\pi(\cdot)$. We used the median of the posterior distribution of this double logistic function from the Bayesian hierarchical model as our "expected fertility decline" covariate.

2.3.2 Modeling correlation

There is between-country correlation in TFR that must be accounted for in our model Fosdick and Raftery (2014). However, estimating correlation matrices directly can lead to noisy estimates since we have 121 countries and each country has at most eight observations. Thus, we cannot simply use the empirical correlation estimates.

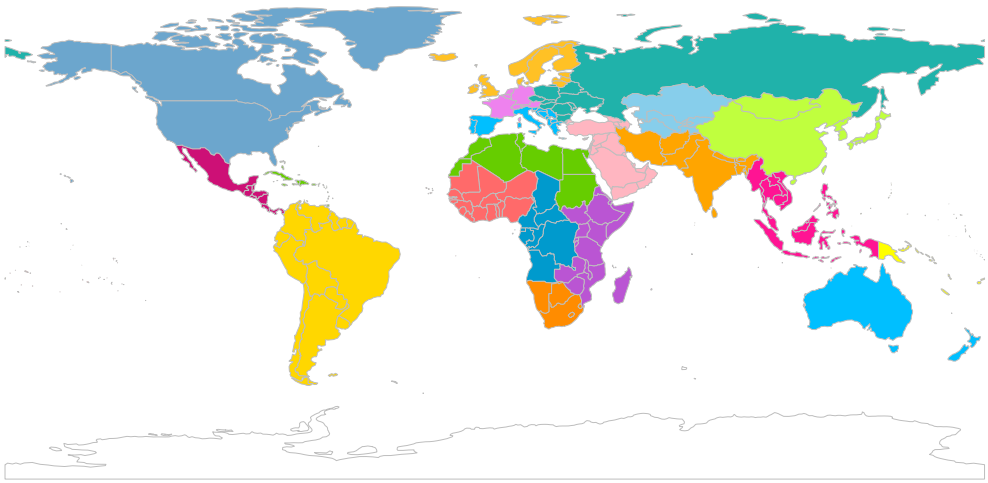


Figure 2.4: The 22 UN regions used for the GLS clustering scheme

We modeled the between-country correlation based on UN region membership. The UN regions are displayed in Figure 2.4. Each UN region is a set of countries that are relatively close geographically and homogeneous culturally. We expect there to be similar between-country correlation for all countries in the same UN region and at the same time point.

We used generalized least squares (GLS) to fit our models via maximum likelihood, as GLS allows us to introduce a between-country correlation structure by constructing clusters based on UN region × time point combinations from the 22 UN regions and eight time points. We assumed an exchangeable correlation structure within each UN region × time point cluster. This correlation structure implies that countries within the same UN region have the same amount of between-country correlation at a given time point. We also assumed homoscedastic errors and between-cluster independence.

Our model is specified as follows. Let $\boldsymbol{f}_{r,t}$ represent the vector of TFR values for all countries in UN region r at time t, let $\boldsymbol{g}_{r,t}$ represent the vector of the expected TFR decrement for all countries in region r at time t, and let $H_{r,t}$ represent the matrix of covariates for all countries in region r at time t. Our model can then be written as

$$\begin{aligned} \boldsymbol{\Delta f}_{r,t} &= \boldsymbol{f}_{r,t-1} - \boldsymbol{f}_{r,t} \\ &= \boldsymbol{\beta_0} + \boldsymbol{\beta_1 g}_{r,t} + \boldsymbol{\beta_h} H_{r,t} + \boldsymbol{\varepsilon}_{r,t}, \\ \boldsymbol{\varepsilon}_{r,t} &\sim N(0, \Sigma_{r,t}), \end{aligned} \tag{2.1}$$

where the (i, j)th term of $\Sigma_{r,t}$ represents the covariance between countries i and j from region r at time t. We model $\Sigma_{r,t}$ as

$$\Sigma_{r,t} = \sigma^2 R_{r,t} \quad \text{where } R_{r,t} = \begin{cases} 1 & \text{if } i = j \\ \rho & \text{if } i \neq j. \end{cases}$$

2.3.3 Model selection

We determined which measures of education and family planning to include in our model for TFR decrement using the Bayesian information criterion (BIC) as the model selection criterion (Raftery, 1995; Schwarz, 1978). We aimed to answer three main questions with the model selection. First, does education affect fertility decline primarily through increased educational attainment of women or through increased enrollment of children? Second, given

the selected education mechanism, which levels of education are the most important? Third, given the selected mechanism and levels of education, does family planning affect fertility decline primarily through reducing unmet need for family planning or though increasing contraceptive prevalence? Using a BIC-based model selection process, we identified which measures of education and family planning were favored by the data.

The model selection process selected one education variable, namely the change over time in women's completion of lower secondary education or higher. This suggests that it is women's educational attainment, rather than children's school enrollment, that best captures the accelerating effect of education on fertility decline. This also suggests that the levels of education most strongly associated with the accelerating effect are lower secondary education or higher.

The model selection process also selected one family planning variable, namely the change over time in prevalence of modern contraceptive methods. This suggests that contraceptive prevalence captures the driving mechanism behind the family planning effect better than unmet need for family planning.

We did not use the BIC to determine which control variables to include in our model. Instead, the control variables measuring child mortality, urbanization, and GDP were included as important background variables to consider, based on evidence in the literature. We also included an indicator variable for SSA:

$$\text{SSA}_{c,t} = \begin{cases} 1 & \text{if country } c \text{ is in SSA} \\ 0 & \text{if country } c \text{ is not in SSA.} \end{cases}$$

The selected variables and their abbreviated names can be found in Table 2.1. For the BIC-based model selection, all models were fitted using GLS, following Equation (2.1). All models included the SSA indicator, Expected TFR Decr, GDP Growth, GDP Growth Change, Urban Change, and Child Mortality Decr as covariates. All continuous variables were centered prior to model fitting.

Table 2.1: Abbreviated names and descriptions of BIC-selected measures of education and family planning and all control variables

Name	Description
Expected TFR Decr	Expected TFR decrement from Alkema et al. (2011)
LowSec+ Change	Change over time in proportion of women who have attained lower secondary education or higher
CP (Modern) Change	Change over time in contraceptive prevalence of modern methods
SSA	Indicator for whether a country is in sub-Saharan Africa
GDP Growth	Annual percentage growth rate of GDP per capita at market prices based on constant local currency
GDP Growth Change	Change over time of GDP Growth
Urban Change	Change over time of the percent of population residing in urban areas
Child Mortality Decr	Change over time of under-five mortality (${}_5q_0$)

Model selection for education

The education variables considered in the model selection were the change over time in NER and the cumulative levels of change over time in women's educational attainment. We only considered cumulative levels of women's educational attainment since changes in non-cumulative levels are difficult to interpret in terms of overall educational gains. For example, an increase in the proportion of women who have attained at most lower secondary education could correspond to more women moving from completing only primary to completing lower secondary education, which would indicate an overall improvement in women's education. However, this increase could also correspond to fewer women moving from lower secondary to upper secondary education, which would indicate an overall decrease in women's education. Using cumulative levels of change over time in attainment eliminates this interpretation problem, as an increase in the proportion of women who have attained lower secondary education or higher unambiguously indicates an overall improvement in women's education.

We first consider the selection of women's educational attainment over children's enrollment. Due to the limited availability of enrollment data, models including enrollment are

based on a reduced dataset of 550 country-time pairs with observations from 116 countries. We selected one education variable, the change over time in the proportion of women who have attained lower secondary education or higher ("LowSec+ Change"), from among the six levels of the change over time in cumulative levels of women's educational attainment (incomplete primary or higher, primary or higher, lower secondary or higher, upper secondary or higher, and postsecondary) and both levels of the change over time of NER (primary and secondary) using BIC as the model selection criterion. The first column of Table 2.2 summarizes the model including both women's educational attainment and children's enrollment variables. The different levels of women's educational attainment are abbreviated analogously to LowSec+ for lower secondary or higher. The change over time in NER is abbreviated to NER Change.

In the model including NER Change, we found that the only significant education variable was LowSec+ Change. Neither of the variables measuring children's enrollment was significant. Since we constructed the "change over time" education variables to be positive when education is increasing, we expected to find positive coefficient estimates for the education variables. However, we found that several of the coefficient estimates, including the coefficient estimates for both enrollment variables, were negative.

From these results, we have answered our first question of interest and found that women's educational attainment was selected over children's enrollment. We have also answered our second question, as the only significant levels of attainment corresponded to the levels lower secondary or higher. However, due to the limited availability of data on children's enrollment, the model including NER Change was fit using only 550 country-time pairs. Given that the selected education mechanism was attainment, we confirmed that the selected levels of attainment were truly lower secondary or higher once we considered all 666 country-time pairs. The second column of Table 2.2 summarizes the model with all levels of women's educational attainment but not including children's enrollment. We once again found LowSec+ Change

was the only significant education variable, supporting our choice of LowSec+ Change as the selected education variable.

Although we selected from the cumulative parameterization of the change over time in women's educational attainment, we additionally verified the selection of LowSec+ Change by checking the estimated effects for the non-cumulative parameterization. Details of this verification can be found in the Appendix.

Table 2.2: Education variable selection: summaries of the model with all education variables and the model with only attainment variables, where both models include all control variables and are fit by GLS with TFR decrement as the dependent variable

	Model with attainment and enrollment		Model with attainment only	
	Estimate	t-value	Estimate	t-value
(Intercept)	0.33	17.4***	0.30	16.3***
Expected TFR Decr	0.91	18.4***	0.91	19.2***
IncPri+ Change	−0.66	−1.3	−0.73	−1.5
Pri+ Change	0.79	1.3	1.00	1.8
LowSec+ Change	2.11	3.2**	2.38	3.9***
UppSec+ Change	−0.78	−1.0	−1.04	−1.5
PostSec+ Change	0.63	0.5	−0.49	−0.5
NER Change (Pri)	−0.16	−1.1		
NER Change (Sec)	−0.16	−1.6		
GDP Growth	−0.38	−1.4	−0.45	−1.7*
GDP Growth Change	0.55	2.9**	0.46	2.6*
Urban Change	−0.31	−0.6	−0.47	−1.0
Child Mortality Decr	0.07	0.1	0.70	0.9
SSA	−0.08	−2.5*	−0.09	−2.6**
Within-cluster correlation	0.23		0.30	
R^2	0.52		0.49	
BIC	-52.30		-86.45	
Country-time pairs	550		666	

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

Model selection for family planning

Finally, we consider the third model selection question: Given that LowSec+ Change is the selected education variable, does family planning affect fertility decline primarily by reducing unmet need for family planning or by increasing contraceptive prevalence? We considered only contraceptive prevalence and unmet need in the selection of family planning indicators despite the availability of a third indicator, demand for family planning satisfied, from the UN. Estimates of contraceptive prevalence and unmet need are both available as percentages of the total number of married or in-union women. Estimates of demand for family planning satisfied are available as a percentage of the total number of married or in-union women who are using any method of contraception or are having an unmet need for family planning (United Nations, 2019b). This difference in denominator is not ideal for direct comparisons of demand for family planning satisfied with the other family planning indicators. All three UN family planning indicators suffer from the "exposure to risk of pregnancy" limitation as argued by Bongaarts (2017), as the family planning indicators are measured among married or in-union women, while the TFR measures births among all women. Demand for family planning satisfied suffers an additional limitation in this regard since it further restricts the group of women considered. For these reasons, we considered only contraceptive prevalence and unmet need as our potential family planning indicators.

The UN provides estimates of contraceptive prevalence for all methods, modern methods, and traditional methods, and estimates of unmet need for all methods and modern methods. These different estimates are highly correlated. Using BIC, we selected the change over time in contraceptive prevalence of modern methods, denoted CP (Modern) Change, from among the five family planning indicators. The estimates of contraceptive prevalence from the UN do not consider contraceptive effectiveness, which Bongaarts (2017) identifies as a key limitation in analyzing the relationship between contraceptive prevalence and TFR. The selection of contraceptive prevalence of modern contraceptive methods partially addresses

the issue of differential contraceptive effectiveness, as the least effective methods (traditional methods) are omitted.

We compared our selected family planning indicator of CP (Modern) Change with the change over time in unmet need for family planning, which is the measure of family planning discussed by Bongaarts & Casterline (2013). For comparison purposes, we considered only the change over time in unmet need for modern methods, here called Unmet Need (Modern) Change. Table 2.3 summarizes the model including both CP (Modern) Change and Unmet Need (Modern) Change. We found that the effect of Unmet Need (Modern) Change was not significant and that its effect size was smaller than that of CP (Modern) Change, supporting our selection of family planning indicator.

Table 2.3: Family planning variable selection: summary of model with contraceptive prevalence, unmet need for family planning, the BIC-selected education variable, and all control variables, fit by GLS with TFR decrement as the dependent variable

	Estimate	t-value
(Intercept)	0.31	19.4***
Expected TFR Decr	0.82	18.1***
LowSec+ Change	1.53	4.7***
CP (Modern) Change	2.74	7.4***
Unmet Need (Modern) Change	0.15	0.3
GDP Growth	−0.58	−2.4*
GDP Growth Change	0.51	3.0**
Urban Change	−0.35	−0.8
Child Mortality Decr	−0.17	−0.2
SSA	−0.07	−2.4*
Within-cluster correlation	0.25	
R^2	0.55	
BIC	-180.43	
Country-time pairs	666	

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

2.4 Results

We first fit the model in Equation (2.1) with main effects only via GLS for the BIC-selected education and family planning variables and all control variables. Next, to identify the potentially differential effect of the covariates on fertility decline within SSA compared to the rest of the world, we considered interaction terms between the SSA indicator and the BIC-selected education and family planning variables and all control variables. We did not consider an interaction between SSA and Expected TFR Decr because the expected decrement is already country-specific by construction. The model with all interactions with SSA is summarized in the first column of Table 2.4, and the model with main effects only is summarized in the second column of Table 2.4. All continuous variables were centered prior to fitting these models.

In both models, we found a significant positive relationship between TFR decrement and LowSec+ Change, where larger increases in the proportion of women who have attained lower secondary education or higher were associated with larger decrements in TFR and thus faster fertility decline. In other words, we found an accelerating effect of women's educational attainment of lower secondary or higher education on fertility decline. Similarly, we found an accelerating effect of contraceptive prevalence of modern methods on fertility decline.

We found separate significant effects of women's educational attainment and contraceptive prevalence even after accounting for the expected TFR decrement and control variables. This follows our expectations from the literature, where generally it has been found there are significant independent effects of family planning and socioeconomic conditions like that of education on fertility (Hirschman, 1994).

Although we found that education and family planning are both important for accelerating fertility decline beyond what we already expect based on past trends, the magnitudes of the coefficient estimates indicate that faster increases in contraceptive prevalence were associated with larger gains in the rate of fertility decline than faster increases in educational

Table 2.4: Final models with BIC-selected education and family planning covariates, all control variables, and with and without interactions with the SSA indicator, fit by GLS with TFR decrement as the dependent variable

	Including interactions with SSA		Main effects only	
	Estimate	t-value	Estimate	t-value
(Intercept)	0.31	19.6***	0.31	19.5***
Expected TFR Decr	0.81	18.3***	0.82	19.1***
LowSec+ Change	1.79	4.6***	1.52	4.8***
CP (Modern) Change	3.38	9.3***	2.67	9.8***
GDP Growth	−1.20	−3.3**	−0.58	−2.4*
GDP Growth Change	0.77	3.7***	0.51	3.0**
Urban Change	−1.46	−2.5*	−0.35	−0.8
Child Mortality Decr	0.44	0.4	−0.14	−0.2
SSA	−0.06	−2.1*	−0.07	−2.4*
SSA:LowSec+ Change	−0.57	−0.8		
SSA:CP (Modern) Change	−1.55	−2.8**		
SSA:GDP Growth	1.08	2.1*		
SSA:GDP Growth Change	−0.61	−1.7		
SSA:Urban Change	1.81	2.1*		
SSA:Child Mortality Decr	−1.40	−0.9		
Within-cluster correlation	0.23		0.25	
R^2	0.57		0.55	
BIC	-168.49		-186.85	
Country-time pairs	666		666	

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

attainment. In the model with interactions, the effect of a change in CP (Modern) Change on TFR decrement was slightly less than twice what an equivalent change in LowSec+ Change would have on TFR decrement. Note that the regression coefficients for LowSec+ Change and CP (Modern) Change can be compared because they are both on the same scale.

We found the larger effect size for contraceptive prevalence compared to educational attainment still held when we considered the composite effects of the observed values and coefficient estimates of CP (Modern) Change and LowSec+ Change. In our dataset, there has

been more rapid observed change in educational attainment than in contraceptive prevalence. The median observed value of LowSec+ Change was 0.045 while the median observed value of CP (Modern) Change was 0.034. As both variables are on the scale of proportions, these values are comparable. We considered the composite median observed effect of LowSec+ Change as the median observed value of LowSec+ Change multiplied by the coefficient estimate for LowSec+ Change. Using the coefficient estimates from the main effects model, we found the composite median observed effect of LowSec+ Change was $0.045 \times 1.52 = 0.068$. Analogously, we found the composite median observed effect of CP (Modern) Change was $0.0341 \times 2.67 = 0.091$. Even with a smaller observed value, CP (Modern) Change still had a larger composite median effect than LowSec+ Change. The same trends of larger median observed values of LowSec+ Change but larger composite median observed effect of CP (Modern) Change were found within SSA and within non-SSA.

We did not find Child Mortality Decr to be significant despite the well-documented marginal relationship between child mortality and fertility. The sign of the coefficient estimate for Child Mortality Decr was also opposite to what we would expect. This was unsurprising once we looked at the marginal relationship between Child Mortality Decr and TFR decrement. Although the non-decrement versions of child mortality and TFR were highly correlated at 0.77, the decrement versions only had a correlation of 0.10. However, when we considered the main effects model without the expected TFR decrement term (model 14a in in Table 2.8), we found Child Mortality Decr was significant. We also found the correlation between Expected TFR Decr and Child Mortality Decr was 0.27. These findings suggest that past trends in fertility decline may account for the explanatory potential of child mortality decrement on TFR decrement, thus leading to an insignificant result in the model including Expected TFR Decr.

Like Child Mortality Decr, Urban Change was not significant in the main effects model but was significant in the model without the expected TFR decrement term, suggesting

the explanatory potential of the change over time in urbanization on fertility decline may be accounted for by the term representing past trends in fertility decline. As we did not find Urban Change to be significant in the model with main effects only, we believe the significance of Urban Change in the model with interactions is not of practical significance.

The control variables GDP Growth and GDP Growth Change must be interpreted together, as they measure aspects of the same quantity and have a significant positive correlation of 0.53. Due to the wide range of trends in GDP Growth and GDP Growth Change possible when a country is undergoing modernization, there is not a simple interpretation of the coefficients on GDP Growth and GDP Growth Change. The overall contribution of the GDP control variables to the predicted TFR decrement reflects both the growth rate and acceleration of GDP.

All of the interaction terms with the SSA indicator in the model with interactions implied a weaker relationship of the covariate on TFR decrement in SSA compared to non-SSA. This will be explored further in a later section.

2.4.1 Direct and indirect effects

We used path analysis to explore the structure of direct and indirect effects of our selected covariates on TFR decrement. Path analysis was developed by Wright (1921) as a way to decompose correlations between dependent and independent variables into direct and indirect effects. The results of a path analysis can be illustrated in a path diagram where unidirectional arrows are used to indicate the assumed causal relationships between covariates and bidirectional arrows are used to connect variables where there is no assumed causal relationship. Unidirectional arrows are labeled with standardized path coefficients, which are regression coefficients from regressions using standardized versions of the covariates, while bidirectional arrows are labeled with correlations.

A path diagram can illustrate the logical temporal ordering of the assumed causal pathway

underlying our analyses. This ordering was arranged in levels in terms of proximity to TFR decrement, where Level 4 is the most proximate to fertility decline. No ordering is assumed among variables at the same level. The ordering was as follows:
Level 1: Urban Change, GDP Growth, GDP Growth Change.
Level 2: Education: LowSec+ Change.
Level 3: Child Mortality Decr.
Level 4: Family planning: CP (Modern) Change.

We included the three control variables measuring a form of "modernization" together on the same level. Among these modernization variables, we assumed a causal relationship only among covariates for Urban Change in the path diagram. Despite modernization being a central part of demographic transition theory, there is uncertainty about the direct effects of modernization variables on fertility decline (Hirschman, 1994). There is greater support for an effect of urbanization on fertility decline in the literature (Bricker and Ibbitson, 2019; Garenne, 2008; White et al., 2008) than for measures of GDP. Thus, we did not make any assumption about the causal pathway between GDP and fertility decline. However, we still included the GDP variables as covariates in the regression for TFR decrement to ensure that the regression represented in the path diagram corresponded to the main effects model in Table 2.4. In the path diagram, we drew unidirectional arrows from Urban Change pointing towards covariates that were assumed to be more proximate to fertility decline. Connections between the GDP variables and all other covariates were assumed to be bidirectional arrows.

The path diagram is shown in Figure 2.5. All regressions in the path diagram include the SSA indicator and Expected TFR Decr as covariates, where the unidirectional arrows from SSA and Expected TFR Decr to all other variables in the path diagram represent assumptions on temporal ordering rather than causality. The direct effects of the GDP variables on TFR Decr are displayed in Figure 2.5, but the bidirectional arrows connecting GDP Growth and GDP Growth Change to SSA, Expected TFR Decr, Urban Change, LowSec+ Change, Child

Mortality Decr, and CP (Modern) Change are omitted for readability. Error terms for Urban Change, LowSec+ Change, Child Mortality Decr, CP (Modern) Change, and TFR Decr are omitted from Figure 2.5 for readability. Arrows corresponding to effects with $P > 0.05$ are also omitted from Figure 2.5 for readability. Path coefficients for the omitted bidirectional arrows, error terms, and arrows corresponding to effects with $P > 0.05$ are reported in the Appendix. The path coefficients displayed in Figure 2.5 are the standardized regression coefficients for models fit using GLS with the UN region $\times$ time point clustering scheme.

Figure 2.5 provides a visual representation of the relative strengths of the direct effects in the main effects only model summarized in Table 2.4 and the indirect effects between the covariates. We found that Expected TFR Decr had the largest direct effect on TFR decrement, as expected. The second largest direct effect on TFR decrement corresponded to contraceptive prevalence as measured by CP (Modern) Change. This effect was about twice as large as the direct effect of women's education as measured by LowSec+ Change.

In a traditional path diagram where path coefficients are obtained using ordinary least squares (OLS), the indirect effect of variable X on variable Y can be computed by multiplying the standardized path coefficients along the indirect path from X to Y. We used GLS with the UN region $\times$ time point clustering scheme to obtain our standardized path coefficients, which resulted in slightly different coefficient estimates from the OLS estimates. However, we can still estimate the indirect effects by multiplying path coefficients from GLS regressions. Comparing direct and indirect effects of women's education is of particular interest. We found that the direct effect of LowSec+ Change on TFR decrement was 0.14. The estimated indirect effect of LowSec+ Change through CP (Modern) Change was $(0.16)(0.28) = 0.0448$ and the estimated indirect effect of LowSec+ Change through Child Mortality Decr was $(-0.13)(-0.01) = 0.0013$. Thus the direct effect of LowSec+ Change on TFR decrement is three times larger than its total indirect effect. Also, the indirect effect of LowSec+ Change is predominantly through CP (Modern) Change. This finding is in line with the literature,

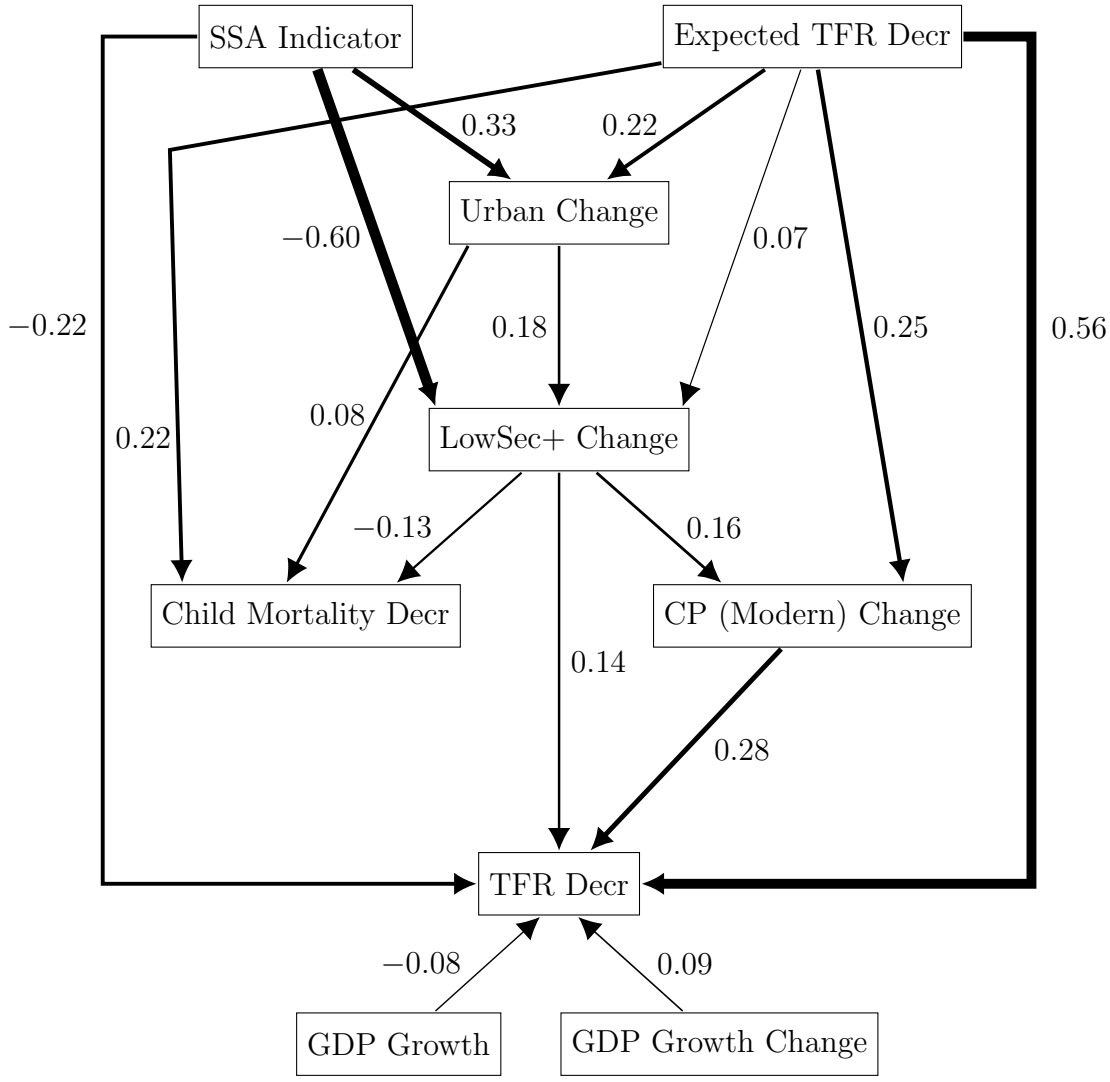


Figure 2.5: Path diagram with standardized path coefficients fit using GLS with error terms, bidirectional arrows, and arrows corresponding to effects with P $>$ 0.05 omitted for readability and line thicknesses proportional to path coefficient magnitudes

which suggests that one venue through which increased women's education impacts fertility decline is through increasing knowledge and acceptance of family planning. We also found that the direct effect of contraceptive prevalence (0.28) was greater than the sum of the

direct and indirect effects of education (0.1861).

2.4.2 SSA difference

We explore the effect of education and family planning on fertility decline within SSA by rewriting the model with SSA interactions in terms of one model for countries in SSA and one model for the rest of the world, displayed in Table 2.5.

Table 2.5: Comparison of coefficient estimates from the model with interactions in Table 2.4 for countries not in SSA and countries in SSA

	non-SSA	SSA
(Intercept)	0.31	0.25
Expected TFR Decr	0.81	0.81
LowSec+ Change	1.79	1.22
CP (Modern) Change	3.38	1.83
GDP Growth	-1.20	-0.11
GDP Growth Change	0.77	0.16
Urban Change	-1.46	0.35
Child Mortality Decr	0.44	-0.96

We found that the effects of LowSec+ Change and CP (Modern) Change on TFR decrement were weaker in SSA than in non-SSA countries, with a bigger decrease in effect size for CP (Modern) Change. According to our model, a faster increase in the proportion of women who have attained at least lower secondary education in SSA corresponds to a smaller decrease in TFR than what an increase in educational attainment of the same rate would have corresponded to in non-SSA. Similarly, a faster increase in contraceptive prevalence in SSA corresponds to a smaller decrease in TFR than what an increase in contraceptive prevalence of the same rate would have corresponded to in non-SSA. We also found a weaker effect of GDP Growth, GDP Growth Change, and Urban Change in SSA compared to countries not in SSA. The direction of the effect changed signs between SSA and non-SSA for both Urban

Change and Child Mortality Decr. The magnitude of the effect of Child Mortality Decr was larger in SSA than in non-SSA, however from Table 2.4 we found that neither the main effect for Child Mortality Decr nor its interaction with SSA was significant.

The smaller effect size of LowSec+ Change on TFR Decr in SSA indicates the accelerating effect that increased women's educational attainment has on fertility decline is diminished in SSA compared to other high-fertility regions. Martín (1995) also found the expected negative relationship between women's education and fertility to be weaker than anticipated in SSA compared to trends from historical high-fertility transitions in non-SSA regions. The weaker effect of women's education on fertility decline in SSA may be due to reductions in school quality or limited expansion of the labor market in SSA (Grant, 2015). These same factors that impact the strength of the association between women's education and fertility may result in a weaker accelerating effect of education as well.

The weaker effect of CP (Modern) Change we found for SSA compared to non-SSA refers only to the accelerating effect that increased contraceptive prevalence may have on fertility. Several studies on the effect of contraceptive prevalence on TFR in SSA have found a weaker effect than expected compared to global trends (Bongaarts, 1987; Tsui, 2001; Westoff and Bankole, 2001). Bongaarts (2017) outlined technical and methodological pitfalls that may bias analyses of the relationship between contraceptive prevalence and TFR in SSA, arguing that the average effect of contraceptive prevalence on TFR is actually the same between SSA and non-SSA after making data adjustments and controlling for regional fixed effects.

The major methodological pitfall discussed in Bongaarts (2017) is the confounding of cross-sectional effect estimates with between-country fertility differences that are constant over time. This is avoided in our analysis since we estimate the relationship between the change over time of contraceptive prevalence and the change over time of TFR rather than the relationship between contraceptive prevalence and TFR. The potential confounding bias is further reduced through the incorporation of country-specific effects in our model via the

country-specific expected TFR decrement term.

Our work still has some of the technical limitations discussed by Bongaarts (2017). For TFR, we alleviate data quality issues and the delayed impact of contraception on TFR estimates by using estimates of TFR for five-year periods from the UN. However, we were unable to implement all of Bongaarts' suggestions for adjusting estimates of contraceptive prevalence. We aggregated yearly estimates of contraceptive prevalence of modern methods into five-year periods to enable direct comparisons with the TFR for five-year periods. The overlap with postpartum infecundability is partially addressed by the use of five-year periods and the differences in contraceptive effectiveness of different contraceptive methods are partially addressed by focusing on modern methods only, as the least effective contraceptive methods are all classified as traditional methods by the UN.

Our chosen measure of contraceptive prevalence is limited by an incomplete exposure of risk to pregnancy and may experience confounding effects from variability in the age structures of women of reproductive age. While estimates of contraceptive prevalence for all women (whether or not they are married or in-union) are available from the UN, these estimates are only available starting from 1990. When we used contraceptive prevalence estimates for all women to fit the model with SSA interactions, we still found a significant difference in the effect of CP (Modern) Change in SSA compared to non-SSA, with a smaller effect size in SSA. Thus we chose to use estimates of contraceptive prevalence for married or in-union women to make use of the additional data covering years 1970-1990. Details of the model using contraceptive prevalence estimates for all women can be found in the Appendix. Also, the estimates of contraceptive prevalence are affected by the age distribution of women within each country while the estimates of TFR are not. This difference may result in a confounding effect of age structure on estimates of the relationship between CP (Modern) Change and TFR Decr.

The direct comparison in Table 2.5 is illuminating, but does not explain the difference in

the average rate of fertility decline between non-SSA and SSA. To explore potential explanations for this difference, we considered sequential models of TFR decrement that added covariates in one at a time. The order in which covariates were added was chosen to reflect the logical temporal ordering of the potentially causal relationships between the covariates and TFR decrement. We used the same order as was used for the path analysis. Note that the three control variables measuring "modernization" (Urban Change, GDP Growth, and GDP Growth Change) do not have an intrinsic temporal ordering. As we added covariates one at a time into the sequential models, we made an arbitrary selection of the order in which to add the three modernization control variables. We chose the following ordering: SSA indicator, Urban Change, GDP Growth, GDP Growth Change, LowSec+ Change, Child Mortality Decr, and finally CP (Modern) Change. After comparing the main effects, we also considered all interactions with the SSA indicator. These interactions were added in the same temporal order. All sequential models were fit using GLS with the UN region $\times$ time point clustering scheme, and all continuous variables were centered prior to fitting the models.

We considered these sequential models both with and without the expected TFR decrement term. Models without the expected TFR decrement term can show general trends without taking historical TFR trajectories into account and provide a descriptive account of relationships between the covariates and TFR decrement. The results of the sequential models without the expected TFR decrement term can be directly compared to existing work investigating how the SSA fertility decline differs from other historical fertility declines, such as Bongaarts and Casterline (2013). The sequential models with Expected TFR Decr are summarized in Tables 2.6 and 2.7. The sequential models without Expected TFR Decr are summarized in Tables 2.8 and 2.9.

For all the sequential models, the biggest change in the SSA coefficient estimate came from adding LowSec+ Change into the model. Thus we found that the difference in the average

rate of fertility decline between SSA and non-SSA could be partially explained by differences in trends in LowSec+ Change between the two geographic areas. This is illustrated in Figure 2.6a, where the median trends in LowSec+ Change for SSA and non-SSA are plotted over time. Although there are clearly differences in the median trends in LowSec+ Change, with larger increases in attainment in non-SSA, the difference appears to be narrowing over time.

We did not find much change in the SSA coefficient estimate when we added in CP (Modern) Change, although we did see an increase in R^2 in all models. In Figure 2.6b, we see that the median trends in CP (Modern) Change for SSA and non-SSA have some overlap. However, we did find a significant difference in the way changes to CP (Modern) Change impact TFR Decrement in SSA compared to non-SSA when we considered the sequential models with interactions both with and without the expected TFR decrement. We found that increases in CP (Modern) have a smaller effect on TFR Decrement in SSA when compared to non-SSA. This is in line with the findings of Bongaarts and Casterline (2013), which point to the high ideal family size in Africa as an obstacle to accelerated fertility decline even in the presence of low unmet need for contraception.

Table 2.6: Comparison of sequential models with Expected TFR Decr and with main effects only, fit via GLS with TFR decrement as the dependent variable

Sequential Models	Model 1	Model 2a	Model 3a	Model 4a	Model 5a	Model 6a	Model 7a
(Intercept)	0.33^{***}	0.33^{***}	0.33^{***}	0.33^{***}	0.31^{***}	0.30^{***}	0.31^{***}
Expected TFR Decr	0.94^{***}	0.94^{***}	0.94^{***}	0.94^{***}	0.93^{***}	0.92^{***}	0.82^{***}
SSA	-0.09^{**}	-0.09^{**}	-0.10^{**}	-0.10^{**}	-0.07^{*}	-0.07^{*}	-0.07^{*}
Urban Change		-0.01	0.01	0.20	-0.26	-0.29	-0.35
GDP Growth			-0.06	-0.34	-0.36	-0.40	-0.58^{*}
GDP Growth Change				0.38^{*}	0.43^{*}	0.44^{*}	0.51^{**}
LowSec+ Change					1.98^{***}	2.03^{***}	1.52^{***}
Child Mortality Decr						0.81	-0.14
CP (Modern) Change							2.67^{***}
R^2	0.45	0.45	0.45	0.45	0.48	0.48	0.55

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

Table 2.7: Comparison of sequential models with Expected TFR Decr and with interactions, fit via GLS with TFR decrement as the dependent variable

Sequential Models	Model 1	Model 2b	Model 3b	Model 4b	Model 5b	Model 6b	Model 7b
(Intercept)	0.33^{***}	0.32^{***}	0.32^{***}	0.32^{***}	0.30^{***}	0.30^{***}	0.31^{***}
Expected TFR Decr	0.94^{***}	0.95^{***}	0.95^{***}	0.94^{***}	0.93^{***}	0.89^{***}	0.81^{***}
SSA	-0.09^{**}	-0.10^{**}	-0.10^{**}	-0.10^{**}	-0.07^{*}	-0.06	-0.06^{*}
Urban Change		-0.45	-0.46	-0.24	-0.86	-1.09	-1.46^{*}
GDP Growth			-0.31	-0.97^{*}	-0.91^{*}	-1.02^{**}	-1.20^{**}
GDP Growth Change				0.68^{**}	0.68^{**}	0.73^{**}	0.77^{***}
LowSec+ Change					2.02^{***}	2.09^{***}	1.79^{***}
Child Mortality Decr						3.01^{*}	0.44
CP (Modern) Change							3.38^{***}
SSA:Urban Change		0.95	0.91	0.67	1.06	1.26	1.81^{*}
SSA:GDP Growth			0.39	1.07^{*}	0.90	1.07^{*}	1.08^{*}
SSA:GDP Growth Change				-0.71	-0.56	-0.65	-0.61
SSA:LowSec+ Change					-0.15	-0.24	-0.57
SSA:Child Mortality Decr						-3.59^{*}	-1.40
SSA:CP (Modern) Change							-1.55^{**}
R^2	0.45	0.45	0.45	0.46	0.49	0.49	0.57

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

Table 2.8: Comparison of sequential models without the expected TFR decrement term and with main effects only, fit via GLS with TFR Decr as the dependent variable.

Sequential Models	Model 8	Model 9a	Model 10a	Model 11a	Model 12a	Model 13a	Model 14a
(Intercept)	0.42^{***}	0.42^{***}	0.42^{***}	0.42^{***}	0.39^{***}	0.38^{***}	0.38^{***}
SSA	-0.11^{*}	-0.12^{**}	-0.12^{**}	-0.13^{**}	-0.09^{*}	-0.10^{*}	-0.09^{**}
Urban Change		2.00^{***}	2.14^{***}	2.34^{***}	1.72^{**}	1.44^{*}	1.10^{*}
GDP Growth			-0.52	-0.82^{*}	-0.82^{*}	-1.03^{**}	-1.21^{***}
GDP Growth Change				0.40	0.46	0.53^{*}	0.61^{**}
LowSec+ Change					2.45^{***}	2.71^{***}	1.87^{***}
Child Mortality Decr						4.94^{***}	3.03^{**}
CP (Modern) Change							3.87^{***}
R^2	0.04	0.05	0.06	0.06	0.12	0.15	0.29

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

We found similar results for the models with and without Expected TFR Decr, where the only major differences centered around Urban Change and Child Mortality Decr. Coefficient

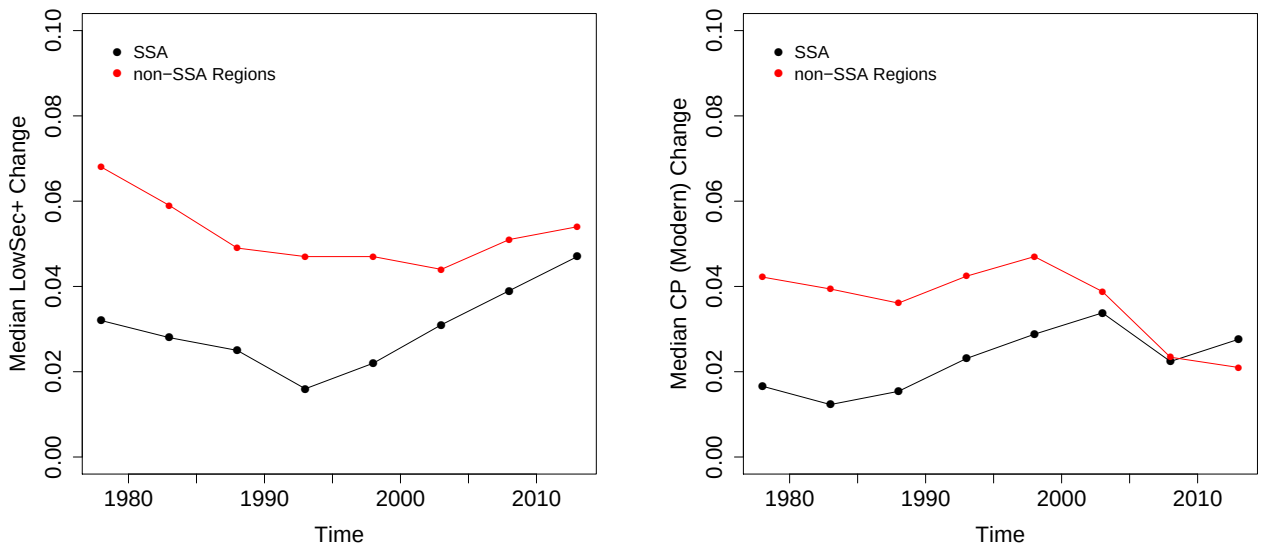
Table 2.9: Comparison of sequential models without Expected TFR Decr and with interactions, fit via GLS with TFR decrement as the dependent variable.

Sequential Models	Model 8	Model 9b	Model 10b	Model 11b	Model 12b	Model 13b	Model 14b
(Intercept)	0.42^{***}	0.42^{***}	0.42^{***}	0.42^{***}	0.40^{***}	0.37^{***}	0.37^{***}
SSA	-0.11^{*}	-0.11^{**}	-0.11^{*}	-0.10^{*}	-0.07	-0.04	-0.05
Urban Change		2.85^{***}	2.77^{***}	3.02^{***}	2.39^{**}	0.99	0.30
GDP Growth			-1.08^{*}	-1.97^{***}	-1.89^{***}	-2.15^{***}	-2.22^{***}
GDP Growth Change				0.92^{**}	0.92^{**}	1.10^{***}	1.08^{***}
LowSec+ Change					2.01^{***}	2.30^{***}	1.86^{***}
Child Mortality Decr						11.44^{***}	7.30^{***}
CP (Modern) Change							4.29^{***}
SSA:Urban Change		-1.92	-1.92	-2.54^{*}	-2.25	-0.90	0.05
SSA:GDP Growth			1.21^{*}	2.43^{***}	2.17^{**}	2.46^{***}	2.26^{***}
SSA:GDP Growth Change				-1.45^{**}	-1.21^{*}	-1.38^{**}	-1.23^{**}
SSA:LowSec+ Change					0.85	0.51	-0.09
SSA:Child Mortality Decr						-11.99^{***}	-8.22^{***}
SSA:CP (Modern) Change							-1.59^{*}
R^2	0.04	0.06	0.07	0.08	0.13	0.21	0.33

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

estimates for LowSec+ Change and CP (Modern) Change were consistent across all models with significant positive effects and a larger effect size for CP (Modern) Change compared to LowSec+ Change. The effect sizes for both LowSec+ Change and CP (Modern) Change were smaller in the models with Expected TFR Decr compared to the models without Expected TFR Decr, indicating that some of the positive association of increased educational attainment and increased contraceptive prevalence with fertility decline can be accounted for by past trends in fertility decline. However, we still observed significant, positive effects for LowSec+ Change and CP (Modern) Change in the models including Expected TFR Decr and observed similar relative effect sizes of LowSec+ Change and CP (Modern) Change in the models with and without Expected TFR Decr. Similarly, both GDP Growth and GDP Growth Change were significant in the models with and without Expected TFR Decr, however the effect sizes were smaller in magnitude in the models with Expected TFR Decr.

The differences in the coefficient estimates for Child Mortality Decr and Urban Change



(a) LowSec+ Change

(b) CP (Modern) Change

Figure 2.6: Comparison of median trends in LowSec+ Change and CP (Modern) Change for SSA (black) and non-SSA (red) from 1975–1980 to 2010–2015

between the models with and without the expected TFR decrement were larger. In the sequential models without the expected TFR decrement term, we found a significant effect of Child Mortality Decr and coefficient estimates that reflected the expected positive relationship between Child Mortality Decr and TFR decrement. However, once Expected TFR Decr was included in the models, the child mortality term was no longer significant and in some cases had the wrong sign on the coefficient estimates. We observed similar differences for Urban Change. While the models including the expected TFR decrement term allow us to examine the additional impact covariates may have on fertility decline beyond what we would already expect the decline to look like based on past trends, the models without the expected TFR decrement term only describe the overall associations between fertility decline and the covariates. Thus, the differences in significance for Child Mortality Decr and

Urban Change may indicate that once we account for past trends in fertility decline via the expected TFR decrement term, Child Mortality Decr and Urban Change do not provide any additional information about changes in TFR in a Granger causality context.

2.5 Discussion

Our analyses aimed to estimate the effects of education and family planning on fertility decline in a high-fertility context, with a focus on the accelerating effect of education and family planning on TFR decline. For education, we aimed to determine whether the effect of education operates through increased educational attainment of women or through increased educational enrollment of children. We were also interested in determining which educational level has the strongest impact on fertility decline. For family planning, we aimed to assess whether the effect of family planning operates by reducing unmet need or by increasing the prevalence of modern contraceptive methods. We also aimed to compare the effects of education and family planning to determine which contributes more to accelerating fertility decline.

We found significant accelerating effects of educational attainment and contraceptive prevalence. Specifically, we found that larger rates of increase in the proportion of women who have attained lower secondary education or higher corresponded to faster declines in TFR. We found a separate accelerating effect of increasing contraceptive prevalence of modern contraceptive methods. Contraceptive prevalence had the largest effect size of all covariates we considered, including education. We found that the effect size for contraceptive prevalence was larger than the effect size for education even when taking the smaller observed values for contraceptive prevalence compared to the observed values for women's educational attainment into account. Using path analysis, we found a small indirect effect of women's educational attainment on fertility decline through contraceptive prevalence, but the direct effect of education was three times larger than the total indirect effect of education.

The accelerating effects we found were dampened within SSA compared to the rest of the world. This dampening is partly explained by differences in the pace of change of women's educational attainment between SSA and non-SSA countries. However, the amount of the average difference in TFR decrement that can be explained by the educational differences is small, and the differences in trends in the rate of change of women's educational attainment between SSA and non-SSA are narrowing over time.

Our approach is inspired by Granger causality and so it attempts to estimate causal effects. However, this does not fully exclude the possibility of the results being affected by unobserved confounders, although the risk of this is smaller than with a traditional regression analysis. Thus, caution is needed when interpreting the estimated parameters.

Nevertheless, our findings do suggest several possible implications for policies aimed at accelerating fertility decline. First, it is women's educational attainment, not children's enrollment, that leads to accelerated rates of fertility decline. Of the different education levels, we found that lower secondary education had the most important accelerating effect. Primary education had a much smaller effect, and additional education beyond the lower secondary level (typically around ages 14–16) also had a smaller effect. Lower secondary education is generally considered the final stage of basic education, and this suggests that making completion of lower secondary education universal throughout the world would accelerate fertility decline. This is Target 4.1 of the Sustainable Development Goals.

We found that women's education and contraceptive prevalence both had significant effects, with contraceptive prevalence having a substantially larger effect size. Finally, policies leading to increases in education and family planning within currently high-fertility countries in SSA may have a lessened effect on fertility decline than has previously been seen from similar policies in other historically high-fertility regions. For education, this may partly reflect differences in educational quality (Grant, 2015), suggesting a focus on improving educational quality in SSA.

Chapter 3

BAYESIAN PROJECTIONS OF TOTAL FERTILITY RATE CONDITIONAL ON THE UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS

In this chapter, we propose a conditional Bayesian hierarchical model for projecting fertility given women's educational attainment and contraceptive prevalence. The conditional projection model enables the quantification of the potential accelerating effect of education and family planning policies on fertility decline in the high-fertility context for different policy intervention scenarios. To illustrate the effect policy changes could have on future fertility, we create probabilistic projections of fertility that condition on scenarios related to achieving the Sustainable Development Goals for universal secondary education and universal access to family planning by 2030.

This chapter is closely based on the article "Bayesian Projections of Total Fertility Rate Conditional on the United Nations Sustainable Development Goals" published in the *Annals of Applied Statistics* (Liu and Raftery, in press).

3.1 Introduction

World population in the next century will be driven by high-fertility countries. The United Nations projects that more than half of the projected increase in world population from 7.8 billion people in 2020 to 10.9 billion people in 2100 will occur in high-fertility countries, primarily in sub-Saharan Africa (United Nations, 2019c). Much of the rest of the population increase is projected to occur in countries with above-replacement fertility, mostly in Asia and Latin America.

Policymakers in these countries have an interest in slowing this population increase by accelerating fertility decline, as high fertility and rapid population growth may have adverse economic, environmental, health, governmental, and political consequences (Bongaarts, 2013). Reductions in fertility can also benefit the economy through what is known as the demographic dividend, where declining fertility can lead to accelerated economic growth by reducing the dependency ratio, increasing women's participation in the paid labor force, and allowing increased investments in human and physical capital (Lee and Mason, 2006; Mason and Lee, 2006).

There is widespread agreement in the demographic literature that increasing education and increasing family planning are the two main factors that can be influenced by policy and may help accelerate fertility decline (Hirschman, 1994). Education is thought to accelerate fertility decline by increasing the opportunity cost of having children for women and by increasing the cost of raising children (Axinn and Barber, 2001; Caldwell, 1982; Caldwell et al., 1985; Easterlin and Crimmins, 1985). These increased costs are evident in education differentials in fertility that have been observed across countries, with more highly educated women tending to have fewer children than less educated women (Bongaarts, 2003; Martín, 1995).

Family planning is also thought to accelerate fertility decline, as family planning is needed to translate changes in fertility desires into changes in realized fertility. Contraceptive prevalence, in particular, is a proximate determinant of fertility that provides a venue for individuals to achieve their desired childbearing (Bongaarts, 1987). Liu and Raftery (2020a) found significant accelerating effects of women's educational attainment and contraceptive prevalence on fertility decline in the high-fertility setting. That is, we found that faster increases in women's educational attainment and contraceptive prevalence, for example due to policy interventions, were associated with faster increases in the pace of fertility decline beyond what we would expect the fertility decline to look like assuming no policy intervention.

We found that the accelerating effect of education on fertility operates through increasing mother's education rather than through increasing children's enrollment, and that the attainment level with the largest effect size was lower secondary education or higher. The accelerating effect of family planning on fertility was found to operate primarily through increasing contraceptive prevalence of modern contraceptive methods rather than through decreasing unmet need for family planning.

Figure 3.1 shows the relationship between the median total fertility rate, the median proportion of women attaining lower secondary education or higher, and the median contraceptive prevalence rate of modern methods for different world regions, plotted as time series covering the five-year time periods 1970–1975 to 2015–2020. We can see a strong negative association between educational attainment and fertility and between contraceptive prevalence and fertility across regions.

There is also evidence to suggest that interventions related to education and family planning may have a smaller accelerating effect on fertility in sub-Saharan Africa (SSA) than elsewhere. Most of the world's currently high-fertility countries are located in SSA, and the fertility transition being experienced in SSA is slower than the historical fertility transitions observed in Asia and Latin America (Bongaarts and Casterline, 2013), where the fertility transition refers to the decline from high to low fertility that occurs as a country develops. The relationships between education, family planning, and fertility are different in SSA than in other regions, with higher fertility and lower contraceptive use for a given level of education compared to other regions (Bongaarts et al., 2017). Differences in ideal family size may diminish the effect of family planning policy interventions in SSA (Bongaarts and Casterline, 2013; Bongaarts et al., 1984), while differences in school quality may diminish the effect of education policy interventions (Grant, 2015). Liu and Raftery (2020a) found that the accelerating effect that increases in women's educational attainment and increases in contraceptive prevalence have on fertility decline were indeed smaller in SSA compared to

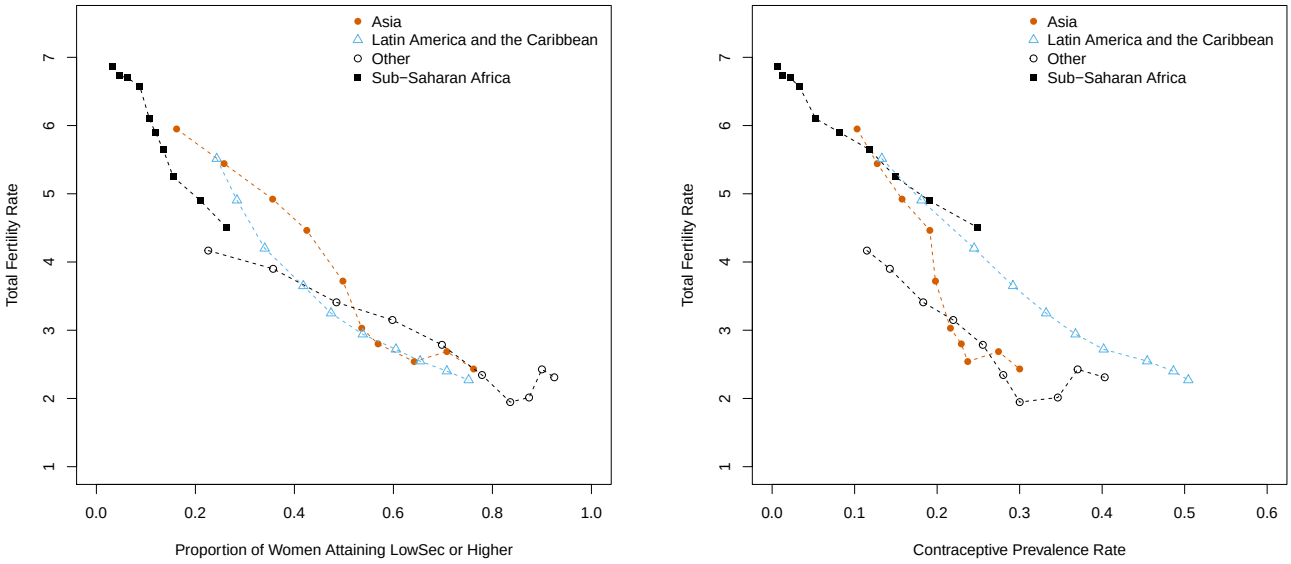


Figure 3.1: Relationship between the median total fertility rate, the median proportion of women attaining lower secondary (LowSec) education or higher, and the median contraceptive prevalence rate of modern methods plotted as time series covering five-year time periods from 1970–1975 to 2015–2020 for each region. Only countries used to estimate the second stage of the conditional TFR projection model and only time periods corresponding to when the country was in Phase II of the fertility transition were used to calculate the medians.

other regions.

This raises the question of how the accelerating effects of women's educational attainment and contraceptive prevalence could impact future fertility and population size in high-fertility countries, particularly in the context of meeting policy goals for education and family planning. Two of the United Nations Sustainable Development Goals (SDGs) refer directly to increasing educational attainment and increasing access to family planning, and thus are likely to have an effect on future population.

The SDGs are a set of goals related to global development that were identified by the

United Nations as a follow-up to the previous Millennium Development Goals. The SDGs were established in 2015 with a target date of completion in 2030. The SDG targets that relate directly to education and family planning are Targets 4.1 and 3.7. Target 4.1 relates to universal educational attainment goals, specifically, "By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes." Target 3.7 includes goals related to family planning, specifically, "By 2030, ensure universal access to sexual and reproductive health-care services, including for family planning, information and education, and the integration of reproductive health into national strategies and programmes" (United Nations, 2015).

The United Nations (UN) has produced estimates and projections of world population by country since 1951 and remains the premier producer of global demographic projections, with projections from the UN used by policymakers around the world. In particular, governments and agencies in countries that do not have robust vital registration systems of their own often rely on the UN estimates and projections to inform planning and policy decisions. However, the UN does not currently produce projections for policy-based scenarios, although projection variants (low, medium, and high) based on different underlying demographic assumptions are available. For fertility projections, the low and high projection variants correspond to assuming the total fertility rate (TFR) will be, respectively, half a child below or half a child above the medium variant TFR. While these variants can provide some guidance to policymakers, they have the drawback of being deterministic, with no statistical interpretation.

Two other producers of global demographic projections, the Wittgenstein Centre for Demography and Global Human Capital and the Institute for Health Metrics and Evaluation (IHME), do produce scenario-based projections of fertility and population that include scenarios corresponding to attaining the SDGs in 2030 (Abel et al., 2016; Vollset et al., 2020). These existing population projections based on policy scenarios either do not fully incor-

porate uncertainty or do not fully incorporate field-specific demographic knowledge. There are also substantial differences between the reference scenario projections produced by the UN and the reference scenario projections produced by both the Wittgenstein Centre and IHME due to differences in methodology and underlying assumptions, so the policy-based projections from other sources cannot be directly compared with the UN reference scenario projections. Demographic projections based on policy scenarios using the UN methodology are thus of interest.

The UN projection model for TFR currently does not explicitly incorporate the effect of covariates, whereas the Wittgenstein Centre emphasizes the effect of education on fertility and IHME incorporates the effects of both education and family planning in their fertility model. Instead, the UN projection model implicitly captures the effect of covariates on fertility by modeling future TFR based on historical trends in TFR. However, for policy-making it may be important to incorporate these covariates explicitly into fertility projection models (Lutz et al., 2014).

Here we develop a conditional probabilistic projection model for TFR that extends the probabilistic fertility projection model used by the UN. The conditional TFR projection model explicitly accounts for women's educational attainment, contraceptive prevalence of modern methods, and GDP per capita and allows for the creation of projections of TFR that are conditional on policy-based intervention scenarios related to education and family planning. Using a Bayesian framework, we address the question of what the quantitative effect of meeting SDG Targets 3.7 and 4.1 would be on future fertility and population.

This chapter is organized as follows. In Section 3.2, we describe the data and methods. Section 3.3 presents the projection results for TFR and population size using Nigeria as a case study. We also present regional aggregate results for sub-Saharan Africa. In Section 3.4, we describe the out-of-sample validation results for the conditional TFR projection model. In Section 3.5, we discuss and compare our results with related work. Finally, we summarize

the findings of this chapter in Section 3.6.

3.2 Methods

3.2.1 Data

Estimates of TFR were obtained from the United Nations *World Population Prospects* (WPP) 2019 Revision (United Nations, 2019c). TFR is a period measure of fertility that measures the expected number of children a woman would bear in her lifetime if she were to experience the period-specific fertility rates at each age and if she lived through the reproductive age range, here defined as ages 15–49. The estimates from WPP 2019 are available for 201 countries by five-year time periods, are comparable across countries and time periods, and are based on vital registers, censuses, and surveys such as the Demographic and Health Surveys (DHS) and the Multi-Indicator Cluster Surveys (MICS).

Estimates of educational attainment for women in the broad age group 20–39 were obtained from the Wittgenstein Centre Data Explorer Version 2.0 (Wittgenstein Centre, 2018; Lutz et al., 2018). The Wittgenstein Centre produces a harmonized data set of the educational attainment distribution that is comparable across countries and times. The educational attainment distribution uses six levels of attainment based on the International Standard Classification of Education: no education, incomplete primary, primary, lower secondary, upper secondary, and post secondary. We focus on cumulative attainment, specifically on the proportion of women attaining lower secondary education or higher, abbreviated as LowSec+. Liu and Raftery (2020a) found this to be the summary measure of education most closely associated with fertility decline.

Probabilistic projections of educational attainment were created following the Wittgenstein Centre methodology using the "wicedproj" package[1] in R, which was released alongside Abel et al. (2016). Figure 3.2 illustrates estimates and projections of women's attainment of

[1]Available at `https://github.com/bifouba/wicedproj`, downloaded on June 17, 2020

LowSec+ for Nigeria. Intervention-based probabilistic projections of educational attainment corresponding to achieving the SDGs were obtained using the methodology developed by Abel et al. (2016), with further details in the Appendix.

Estimates and projections of contraceptive prevalence of modern methods for all women aged 15–49 were obtained following the methodology of Kantorová et al. (2020). Kantorová et al. created probabilistic estimates and projections of family planning indicators using a Bayesian hierarchical model, which is implemented in the "FPEMglobal" package[2] in R. We used the median estimates of contraceptive prevalence of modern methods from a converged simulation of FPEMglobal as an input to our TFR projection model, where contraceptive prevalence is a proportion between 0 and 1 and is constructed for five-year time periods from 1970–1975 through 2015–2020. Probabilistic projections of contraceptive prevalence from 2020–2025 to 2095–2100 were similarly obtained using a converged simulation of FPEMglobal. Figure 3.2 illustrates estimates and projections of contraceptive prevalence for Nigeria. Intervention-based probabilistic projections of contraceptive prevalence corresponding to meeting the SDGs were created by modifying the non-intervention projections. Details of the SDG intervention projections of contraceptive prevalence can be found in Section 3.2.5, with further details in the Appendix.

Estimates of GDP per capita were obtained from the Maddison Project (Maddison Project, 2018) and projections of GDP were obtained using a Bayesian hierarchical model developed by Raftery et al. (2017). As we are not interested in interventions targeting GDP, we considered only non-intervention projections of GDP for our conditional TFR projections.

3.2.2 Model

We build upon the unconditional model for probabilistic fertility projections that is the basis for fertility projections produced by the UN (Alkema et al., 2011; Fosdick and Raftery,

[2]Available at `https://github.com/FPcounts/FPEMglobal`, version 1.1.0 downloaded on June 17, 2020

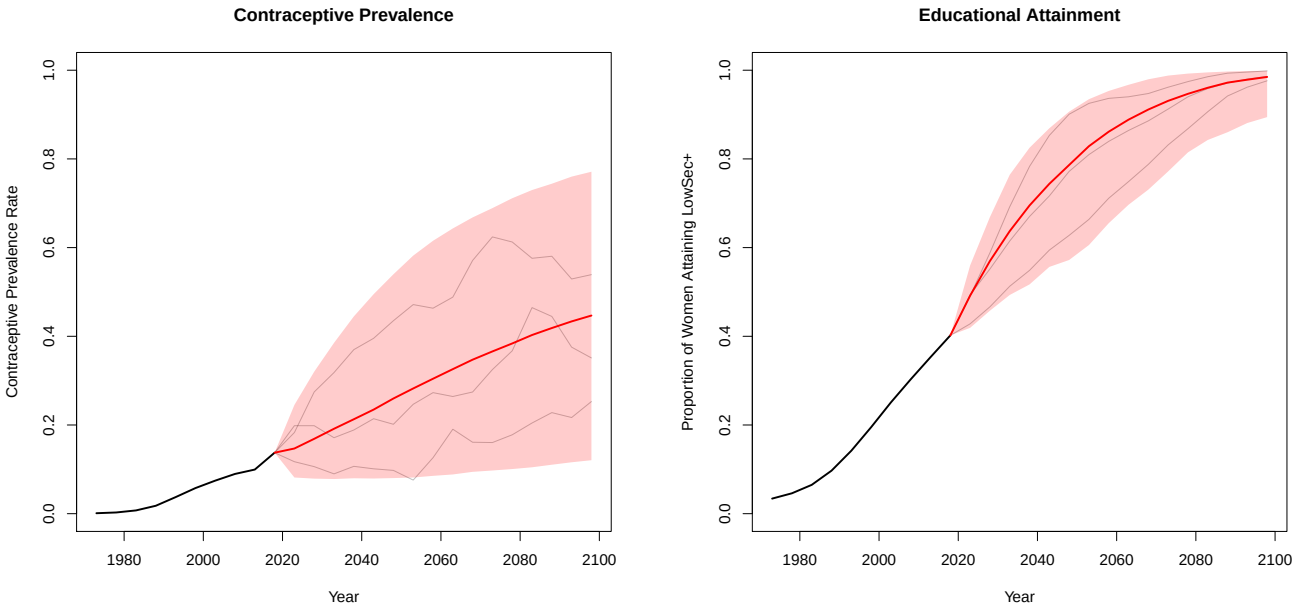


Figure 3.2: Estimates and non-intervention projections of contraceptive prevalence of modern methods for all women and proportion of women attaining lower secondary education or higher for Nigeria from 1970–1975 to 2095–2100. Estimates of the past are plotted in black, medians and 95% intervals for projections are plotted in red, and sample projection trajectories are plotted in grey.

2014; Raftery et al., 2014). The unconditional model is a Bayesian hierarchical model that has been implemented in the R package "bayesTFR" (Ševčíková et al., 2011).[3] The unconditional fertility projection model, referred to in this dissertation as the "bayesTFR" model, divides the fertility transition into three phases as defined by Alkema et al. (2011): Phase I is the high-fertility pre-transition phase, Phase II is the transition phase where fertility falls from high to low, and Phase III is the low-fertility post-transition phase. Since all or almost all countries have already begun the fertility transition, Phase I is not needed for TFR projections and thus is not modeled. The bayesTFR Phase II and Phase III models, as well

[3]bayesTFR version 6.4.0 was used

as our modifications to the Phase II model, are described in this section.

In the bayesTFR Phase II model, fertility decline is modeled as a random walk with drift, where the drift term represents the systematic decline. Let $f_{c,t}$ denote the TFR in country c and five-year time period t. Decrements in TFR are constructed as a measure of fertility decline, with the TFR decrement between five-year time periods t and $(t+1)$ defined as $\Delta f_{c,t+1} = f_{c,t+1} - f_{c,t}$. The unconditional Phase II model is written

$$\begin{aligned}
\Delta f_{c,t+1} &= f_{c,t+1} - f_{c,t} \\
&= -\, g(f_{c,t}|\boldsymbol{\theta}_c) + \varepsilon_{c,t}, \\
\boldsymbol{\theta}_c &\sim h(\cdot|\phi), \\
\phi &\sim \pi(\cdot),
\end{aligned}$$

where $g(f_{c,t}|\boldsymbol{\theta}_c)$ is a five-parameter double logistic function that represents the expected TFR decrement from five-year time period t to five-year time period $(t+1)$. The double logistic function is defined as

$$\begin{aligned}
g(f_{c,t}|\boldsymbol{\theta}_c) = &\frac{-d_c}{1+\exp\left(-2\frac{\ln(9)}{\Delta_{c1}}\left(f_{c,t} - \sum_i \Delta_{ci} + 0.5\Delta_{c1}\right)\right)} \\
&+ \frac{d_c}{1+\exp\left(-2\frac{\ln(9)}{\Delta_{c3}}\left(f_{c,t} - \Delta_{c4} - 0.5\Delta_{c3}\right)\right)}
\end{aligned}$$

and takes the current value of TFR ($f_{c,t}$) and a vector of country-specific parameters $\boldsymbol{\theta}_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$ as inputs. The country-specific parameter vector specifies the shape of each country's individual decline curve and follows a world distribution $h(\cdot|\phi)$ with parameter ϕ, where the prior distribution of ϕ is $\pi(\cdot)$. Further details of the decline function can be found in the Appendix.

The error term for each five-year time period t in the bayesTFR Phase II model is given

by

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

where τ_c is the start time period of Phase II, m_τ is the mean of the error in the start period, and s_τ is the standard deviation of the error in the start period. In time periods following the start of the fertility transition, the standard deviation depends on the current level of the TFR and is given by the function $\sigma(f_{c,t})$, with details given in the Appendix.

Accounting for between-country correlation in TFR is important when constructing aggregates, such as regional or world TFR. The bayesTFR projection model accounts for between-country correlation in projections using a pairwise likelihood method developed by Fosdick and Raftery (2014), which models correlation between countries i and j as a function of whether countries i and j are contiguous, whether they had a common colonizer after 1945, and whether they belong to the same UN region. In estimation of the bayesTFR model, the error terms are assumed to be independent.

We create a conditional TFR projection model by extending the unconditional Phase II model to include a covariate term and to account for between-country correlation in estimation. Covariates were constructed as changes over time on the same scale as the TFR decrements $\Delta f_{c,t+1}$. For example, the change over time from t to $(t+1)$ for covariate X is denoted by $\Delta X_{c,t+1}$. The covariates added are the change over time in the proportion of women aged 20–39 who have attained at least lower secondary education, denoted by ΔLowSec+, the change over time in the contraceptive prevalence of modern methods for all women of reproductive age, denoted by ΔCP, and the percent change in GDP per capita, denoted by ΔGDP. Each covariate is centered at its expected value assuming no policy intervention, with details of the centering found in the Appendix. The centering ensures the covariates reflect acceleration in the trends for women's educational attainment, contraceptive prevalence, or GDP per capita beyond what we would expect these trends to look like assuming

no policy intervention. For example, if women's educational attainment increases at the pace expected under no policy intervention, the covariate ΔLowSec+ will be close to 0. However, if women's educational attainment increases at a faster pace than expected under no policy intervention, the covariate ΔLowSec+ will be greater than 0. We also consider interaction terms between the covariates and an indicator function SSA_c for whether country c is in sub-Saharan Africa.

The conditional TFR projection model is specified as

$$\begin{aligned}\Delta f_{c,t+1} &= f_{c,t+1} - f_{c,t} \\ &= -g(f_{c,t}|\boldsymbol{\theta}_c) + \boldsymbol{\Delta X}_{c,t}\boldsymbol{\beta} + \varepsilon_{c,t},\end{aligned}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_E \\ \beta_F \\ \beta_G \\ \beta_{E,SSA} \\ \beta_{F,SSA} \\ \beta_{G,SSA} \end{bmatrix}, \quad \boldsymbol{\Delta X}_{c,t}^T = \begin{bmatrix} (\Delta\text{LowSec+})_{c,t} \\ (\Delta\text{CP})_{c,t} \\ (\Delta\text{GDP})_{c,t} \\ (\Delta\text{LowSec+})_{c,t} \times SSA_c \\ (\Delta\text{CP})_{c,t} \times SSA_c \\ (\Delta\text{GDP})_{c,t} \times SSA_c \end{bmatrix},$$

$$\beta_j \sim N\left(0, 0.25 \times \frac{Var(\Delta f)}{Var(\Delta X_j)}\right) \quad \text{for } j \in (E; F, G; E, SSA; F, SSA; G, SSA)\,.$$

The prior distributions of the coefficients β_j were chosen to be diffuse, where the prior variances are determined by the ratio of the sample variance of observed changes in f to the sample variance of observed changes in X_j.

We account for between-country correlation in the estimation of the conditional model using clusters based on UN region membership. Each UN region consists of countries that are both spatially contiguous and relatively culturally homogeneous, so we expect similar between-country correlation for all countries in the same UN region and at the same time period. Let $\tilde{\boldsymbol{\sigma}}$ denote the vector of values of $\sigma(f_{c,t})$ ordered by UN region and time period.

The error term for the conditional TFR projection model is specified as

$$
\begin{aligned}
\boldsymbol{\varepsilon} &\sim N(0, \Sigma), \\
\Sigma &= \text{diag}(\tilde{\boldsymbol{\sigma}}) \cdot R \cdot \text{diag}(\tilde{\boldsymbol{\sigma}}), \\
R[i,j] &= \begin{cases} 1 & \text{if } i = j \\ \rho^{[bc]} & \text{if } i, j \in \text{same UN region and same time period} \\ 0 & \text{otherwise} \end{cases}, \\
\rho^{[bc]} &\sim \text{Uniform}(0, 1),
\end{aligned}
$$

where the (i, j)th term of the correlation matrix R represents the correlation between country-time pair i and country-time pair j. If observations i and j are from the same time period and refer to countries within the same UN region, the between-country correlation is $\rho^{[bc]}$. Further details of the conditional TFR projection model can be found in the Appendix.

To create fertility projections, we use the same unconditional post-transition Phase III model as bayesTFR. The post-transition phase represents what happens to a country's TFR once it has completed the fertility transition in Phase II, where the end of Phase II is defined by the UN as the midpoint of the five-year time periods where two successive increases in TFR have been observed after TFR has fallen below two children per woman (United Nations, 2019d). In Phase III, fertility is assumed to converge towards and fluctuate around country-specific long-term TFR levels. It is modeled as a first-order autoregressive (or AR(1)) time series model written as

$$f_{c,t+1} \sim N(\mu_c + \rho_c(f_{c,t} - \mu_c), s^2),$$

where μ_c is the long-term mean for country c. The country-specific means μ_c are assumed to be drawn from a world distribution with mean μ, which itself has a prior distribution restricting it to be no greater than replacement-level fertility, i.e. $\mu \leq 2.1$. The country-specific autoregressive parameter ρ_c is restricted to $0 < \rho_c < 1$, and s is the standard devation

of the random errors. The Phase III model is estimated using a Bayesian hierarchical model, with further details available in Alkema et al. (2011) and Raftery et al. (2014).

3.2.3 Causal assumptions

The primary goal of creating the conditional TFR projection model is to create intervention-based projections of TFR for interventions corresponding to policy outcomes, which is an inherently causal goal. The directed acyclic graph (DAG) in Figure 3.3 illustrates the causal assumptions underlying our analysis that are necessary for the causal interpretation of intervention-based projections of TFR from the conditional projection model.

Each node in the DAG represents a time-varying variable, where single-bordered nodes represent continuous variables and double-bordered nodes represent nodes that are deterministic functions of their parents. TFR decline from time t to time $(t+1)$ is represented by the node labeled Δf_{t+1}. TFR at time t is represented by the node labeled f_t. Given its parents f_{t-1} and Δf_t, f_t can be calculated deterministically and thus is represented as a double-bordered node. The expected TFR decrement from time t to time $(t+1)$ is represented by the node labeled $g(f_t|\boldsymbol{\theta})$. Given the parameters of the double logistic curve $\boldsymbol{\theta}$, $g(f_t|\boldsymbol{\theta})$ is a deterministic function of its parent f_t and thus is represented as a double-bordered node.

The covariate nodes are labeled as ΔCP for the contraceptive prevalence term, ΔE for the educational attainment term, and ΔGDP for the GDP per capita term. Each covariate node represents the main effect of the covariate and its interaction with the SSA indicator. For example, the $\Delta \mathrm{E}_t$ node represents the contribution of both $(\Delta\text{LowSec+})_t$ and $(\Delta\text{LowSec+})_t \times SSA$.

Each dashed arrow in the DAG represents a deterministic relationship, while each solid arrow represents an assumed causal relationship in the direction indicated by the arrow. These causal relationships are informed by demographic background knowledge, primarily the proximate determinants framework for fertility developed by John Bongaarts (Bongaarts,

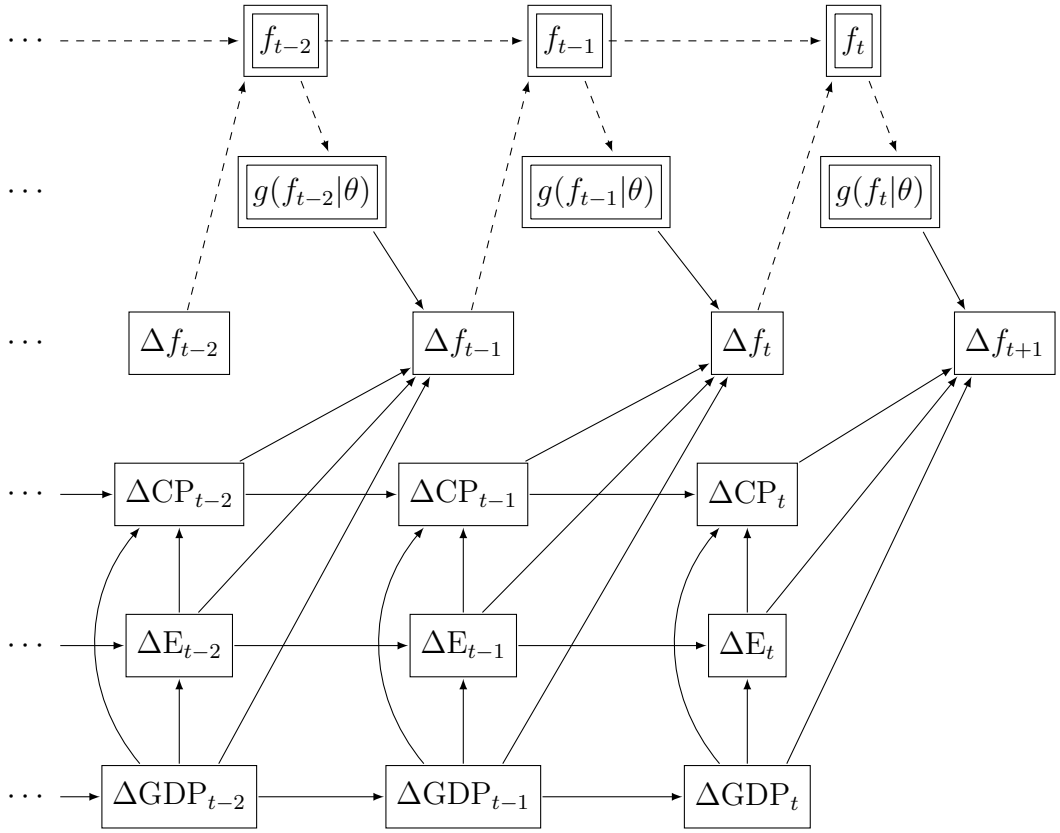


Figure 3.3: Directed acyclic graph (DAG) for TFR f, the double logistic expected TFR decrement term g, contraceptive prevalence CP, educational attainment E, and GDP per capita GDP. Single-bordered rectangles denote nodes representing continuous variables and double-bordered rectangles denote nodes that are deterministic functions of their parents. Deterministic relationships are indicated by dashed arrows while stochastic relationships are indicated by solid arrows.

1978, 2010; Bongaarts et al., 1984).

Child mortality and urbanization, two key variables associated with the fertility transition, are assumed to be mediated through variables included in the DAG. Liu and Raftery (2020a) found that the effect of child mortality on fertility decline was mediated through the double logistic expected TFR decrement. We found an insignificant effect of child mortality decrement on TFR decrement after controlling for the double logistic expected TFR decre-

ment term. We also found that inclusion of the child mortality decrement did not improve model fit.

Urbanization is assumed to be mediated through the GDP term. Liu and Raftery (2020a) found this to be the case, where the effect of urbanization was insignificant after controlling for GDP. There is considerable debate about causal relationships underlying modernization variables, such as GDP and urbanization, and fertility decline (de Silva and Tenreyro, 2017; Hirschman, 1994). The potential for reverse causality between GDP and fertility decline, where past values of fertility decline cause future values of GDP, is omitted from our DAG, though studies such as Herzer et al. (2012) suggest this causal pathway may exist. We note that the inclusion of reverse causal paths like $\Delta f_{t-1} \rightarrow \Delta\text{GDP}_t$ in the DAG does not change the assumptions needed for a causal interpretation of interventions on education and family planning.

We follow the logic of the back-door criterion introduced by Pearl (1993) to identify the adjustment set W needed to estimate the causal effect that interventions on $X = \{\Delta\text{E}_t, \Delta\text{CP}_t\}$ have on $Y = \{\Delta f_{c,t+1}\}$. We find that the set $W = \{g(f_t|\boldsymbol{\theta}), \Delta\text{GDP}_t\}$ satisfies the generalized back-door criterion developed by Maathuis and Colombo (2015). The set W does not contain descendants of X and, for every X_i in X, the set $W \cup X \backslash \{X_i\}$ blocks every back-door path from X_i to Y. Following Theorem 3.1 from Maathuis and Colombo (2015), W is then an appropriate adjustment set for estimating the causal effect of X on Y. In the conditional projection model, we include all elements of the adjustment set as covariates. Thus, under the assumptions underlying the DAG in Figure 3.3, our estimated effects can be interpreted as causal.

3.2.4 Estimation

The conditional projection model is estimated using a Markov chain Monte Carlo algorithm with Gibbs sampling, Metropolis-Hastings, and slice sampling steps. We estimated

the conditional TFR projection model in two stages. In the first stage, the country-specific parameters $\boldsymbol{\theta}_c$ for the double logistic expected TFR decrement function were estimated in the absence of the covariates. In the second stage, the $\boldsymbol{\beta}$ coefficients were estimated jointly, conditionally on the posterior distributions of the double logistic parameters from the first stage. Using two stages for estimation allows us to preserve the demographic interpretation of the double logistic parameters, while still explicitly accounting for the effect of covariates in the TFR projection model.

The first stage was estimated analogously to the UN projection model, using estimates of TFR for 201 countries spanning 1950–1955 through 2015–2020 from WPP 2019. This allowed us to use all the available data to estimate the double logistic expected TFR decrement term and ensured that our estimates of $\boldsymbol{\theta}_c$ were comparable to the estimates used by the UN.

In the second stage, we were primarily interested in the accelerating effect of education and family planning policy interventions on future TFR in the high-fertility setting. We thus restricted the subset of data used for estimation of the second stage to consider only current and historical "high-fertility" transitions, defined for each country as the time periods where the country had begun the fertility transition and had TFR greater than 2.5. Countries without available covariate data were excluded from analyses in the second stage. This resulted in a subset of 114 countries with 1007 observations, where the earliest time period for which we have data is 1970–1975. The top panel of Figure 3.4 shows the number of observations from each country included in the second stage of estimation.

Estimating the effect of education and family planning on TFR in a second stage conditional on the double logistic parameters also provides us with an interpretation of the coefficients that better lends itself to the intervention setting. For example, the covariate ΔLowSec+ measures the rate of increase in women's educational attainment beyond what would be expected if no policy intervention targeting education occurred. An accelerating effect of education on fertility decline occurs if larger values of ΔLowSec+ correspond to faster

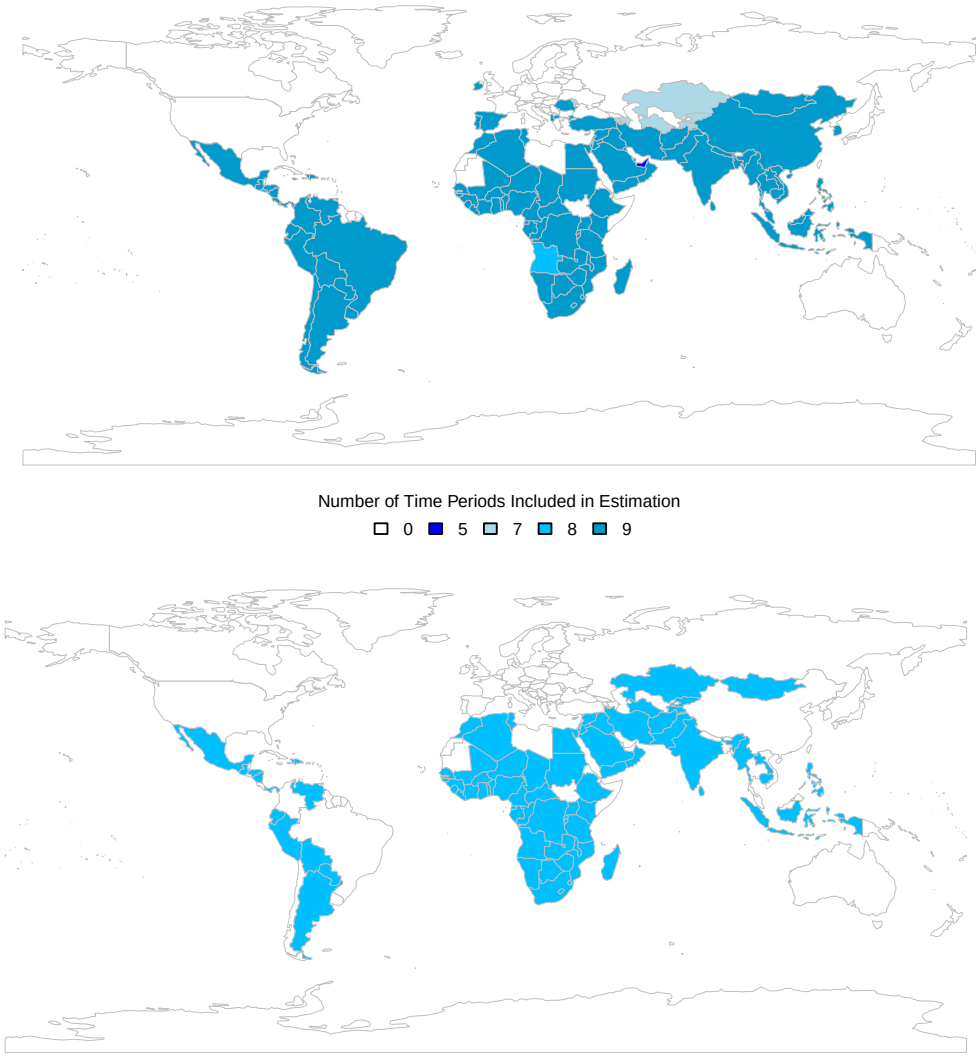


Figure 3.4: Top panel: Number of observations for each country used for estimation of the conditional TFR projection model, where the total number of observations is 1007; observations come from 114 countries and cover time periods 1970–1975 to 2015–2020. Bottom panel: TFR projections are created using the conditional TFR projection model for the 83 countries highlighted in blue.

declines in TFR than what we would expect TFR decline to be, assuming no intervention targeting education. By estimating the effect of the covariates in a second stage, β_E estimates

precisely this effect. The coefficient β_E can then be interpreted as the effect ΔLowSec+ has on TFR decline after we account for the pace of the fertility decline as estimated by the double logistic function, controlling for the other covariates.

3.2.5 Projection

Our focus is on creating intervention-based projections of TFR in the high-fertility setting. The relationships between education, family planning, and fertility decline estimated in the high-fertility context may not apply once fertility has declined to around replacement level. For example, there is evidence that the effect of educational attainment on fertility decline may be weaker in the low-fertility setting than in the high-fertility setting (Adserá, 2017a, Sobotka et al., 2017). Due to these differing relationships, we created intervention-based projections only for countries with TFR $\geq$ 2.1 in 2015–2020 that have available covariate data, which are highlighted in blue in the bottom panel of Figure 3.4. TFR projections for countries with current TFR less than 2.1 and countries without available covariate data were created using the bayesTFR Phase II model. Note that for our purposes, TFR projections for low-fertility countries and countries without available covariate data were primarily of interest to produce projections for the sub-Saharan Africa regional aggregate. Within sub-Saharan Africa, there is one country with current TFR $<$ 2.1 and eight countries without available covariate data. For all countries, the post-transition phase was projected using the bayesTFR Phase III model.

As an input to the conditional TFR projections, we require probabilistic projections for each of the covariates. Uncertainty about the future values of the covariates is propagated into the conditional projections of TFR by randomly drawing from the projected distributions of the covariates when constructing each trajectory of projected TFR. That is, for each projected trajectory of TFR for country c, we use one trajectory from the projected distribution of ΔLowSec+ for country c, one trajectory from the projected distribution of

ΔCP for country c, and one trajectory from the projected distribution of ΔGDP for country c. The resulting distribution of conditional TFR projections reflects both the uncertainty from the conditional TFR projection model parameters and uncertainty from the covariate projections.

We considered five scenarios for intervention-based projections of educational attainment and contraceptive prevalence as inputs to the conditional TFR projection model. These scenarios are summarized in Table 3.1. The covariate projections were centered at their expected values assuming no policy intervention, which ensures the intervention-based covariate projections reflect acceleration in trends beyond what we would expect if no intervention occurred. For example, in the non-intervention setting, women's educational attainment is projected to increase at the same pace that would occur assuming no intervention. Thus, the centered non-intervention projections of ΔLowSec+ are close to 0. In the SDG intervention scenarios, women's educational attainment is projected to increase at faster rates than would occur assuming no intervention. Correspondingly, the centered intervention-based projections of ΔLowSec+ are larger than 0. Further details of the methods used to construct the intervention-based covariate projections can be found in the Appendix.

First, we considered non-intervention projections of the covariates as our reference scenario. We note that TFR projections from the reference scenario are not expected to be identical to the projections produced by the UN for WPP 2019. However, the two sets of projections should be very similar, as the reference scenario was constructed to reflect the assumption of no additional policy intervention targeting education or family planning that is implicit in the WPP 2019 projections.

Next, we considered two scenarios corresponding to attaining SDG Targets 4.1 and 3.7 simultaneously in 2030. Both scenarios interpret achievement of Target 4.1 as attaining universal lower secondary education by 2030, following the implementation developed by Abel et al. (2016). Both scenarios also include the effect of achieving Target 4.1 on family

planning, where increased educational attainment is assumed to increase demand for family planning. Where the two scenarios differ is in their implementations of Target 3.7. The first of these scenarios, labeled "Both SDGs (0% Unmet)," interprets Target 3.7 as meaning that unmet need for family planning will decline to zero in 2030. This scenario can be thought of as an upper bound on the possible effect the SDG intervention could have on TFR projections.

The second scenario, labeled "Both SDGs (75% DS)," interprets Target 3.7 as meaning that at least 75% of the demand for family planning in 2030 will be satisfied using modern methods. Demand satisfied using modern methods is defined as the ratio of contraceptive prevalence of modern methods to total demand, where total demand is the sum of total contraceptive prevalence and unmet need for family planning. This scenario follows the benchmark for achievement of Target 3.7 proposed by Fabic et al. (2015) and provides a more realistic interpretation of the effect that attaining the target might have on projections of contraceptive prevalence. The SDG intervention projections of contraceptive prevalence for this scenario were constructed following the accelerated transition method developed by Cahill et al. (2020) with some modifications.

We also considered a more gradual policy intervention in which the SDGs are achieved in 2040 rather than 2030, called "Both SDGs 2040 (75% DS)." This scenario follows the same implementation as the Both SDGs (75% DS) scenario, with the modification that the SDG targets are assumed to be met in 2040 instead of 2030. For many high-fertility countries, achieving the SDG targets in 2030 is highly ambitious (Abel et al., 2016; Friedman et al., 2020), so considering a scenario where the same goals are met a decade later may be viewed as a more realistic policy intervention.

Finally, we considered an "Education SDG Only" scenario where only SDG Target 4.1 is met in 2030 with no additional policy intervention for family planning. For the Education SDG Only scenario, the projection model was re-estimated to include only education, GDP,

and their interactions with the SSA indicator as covariates. Education is a less proximate cause of fertility decline than family planning, as seen in Figure 3.3. Re-estimating the projection model without the family planning variable is necessary to ensure that the coefficient of education fully captures both the direct effect that education has on fertility and the indirect effect education has on fertility through the effect of education on family planning. By comparing results from the Education SDG Only scenario with results from the scenarios where both SDG targets are met, we were able to quantify the additional effect that meeting Target 3.7 would have on TFR projections.

For all intervention-based projection scenarios, we assumed that the policy efforts required to attain the SDGs in the target year (2030 or 2040) are sustained out to 2100.

TFR projections from the conditional TFR projection model are translated into population projections using the cohort-component method as implemented in the "bayesPop" R package (Ševčíková et al., 2016), which is based on the demographic balancing equation,

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

where the population size at time $(t+1)$ is equal to the population size at time t plus the number of births and number of immigrants occurring in time interval t to $(t+1)$ and minus the number of deaths and number of emigrants occuring in time interval t to $(t+1)$ (Preston et al., 2001). The bayesPop package uses the cohort-component method of population projection, an age- and sex-specific version of the demographic balancing equation, to create age- and sex-specific population projections. Population projections were created using probabilistic projections of mortality and migration as inputs. Projections of mortality were created using the "bayesLifeHIV" R package[4] (Godwin and Raftery, 2017) and projections of migration were created following the method of Azose and Raftery (2015).

[4]Available at `https://github.com/PPgp/bayesLifeHIV`, downloaded on March 30, 2021

Table 3.1: Summary of projection scenarios

Scenario	Target 4.1 Assumptions	Target 3.7 Assumptions
Reference	No intervention	No intervention
Both SDGs (0% Unmet)	Universal lower secondary education by 2030	Unmet need reaches 0% in 2030; all unmet need is assumed to be met with modern methods
Both SDGs (75% DS)	Universal lower secondary education by 2030	Demand satisfied by modern methods reaches 75% by 2030; all increases in contraceptive prevalence are assumed to be for modern methods
Both SDGs 2040 (75% DS)	Universal lower secondary education by 2040	Demand satisfied by modern methods reaches 75% by 2040; all increases in contraceptive prevalence are assumed to be for modern methods
Education SDG Only	Universal lower secondary education by 2030	No intervention

3.3 Results

Projections of TFR and population size from the conditional projection model are presented in this section for the five-year time periods from 2020–2025 to 2095–2100 using Nigeria as a case study. Projections of TFR and population size for sub-Saharan Africa are also presented. Additional projection results, including projections of population size for the regional aggregate of all countries for which we create intervention-based projections of TFR, are available in the Appendix.

3.3.1 Case study: Nigeria

We present projections of TFR and population size using Nigeria as a case study. Nigeria is one of the most important countries for projections of future world population as it is a high-fertility country, with TFR in 2015–2020 of 5.42 children per woman, and is the most populous country in Africa, with estimated population size in 2020 of 206 million people.

Median projections of TFR and population for Nigeria are plotted for all projection scenarios in Figure 3.5. We also plot 95% projection intervals (PIs) for the reference and Both SDGs (0% Unmet) scenarios. Median and 95% PIs for projected values of TFR and population size for all scenarios are reported in Tables 3.2 and 3.3, where the projection intervals reflect both uncertainty about the parameters of the conditional TFR projection model and uncertainty about the future values of the covariates. Tables 3.2 and 3.3 also report differences in medians across the different projection scenarios for ease of comparison.

Table 3.2: Median TFR projections in 2030–2035, 2045–2050, and 2095–2100 for Nigeria in children per woman for all projection scenarios with 95% PIs. Rows indicating differences between projection scenarios show differences between median projected TFR.

	2030–2035	2045–2050	2095–2100
Reference	4.42 (3.26, 5.24)	3.56 (1.87, 4.78)	2.23 (1.15, 3.76)
Both SDGs (0% Unmet)	3.92 (2.72, 4.81)	3.29 (1.75, 4.44)	2.28 (1.23, 3.75)
Both SDGs (75% DS)	4.05 (2.90, 4.90)	3.34 (1.80, 4.54)	2.28 (1.23, 3.81)
Both SDGs 2040	4.10 (2.94, 4.95)	3.32 (1.77, 4.57)	2.26 (1.26, 3.80)
Educ SDG Only	4.24 (3.16, 5.02)	3.48 (1.97, 4.64)	2.32 (1.19, 3.74)
Reference − Both SDGs (0% Unmet)	0.50	0.27	−0.05
Reference − Both SDGs (75% DS)	0.37	0.22	−0.05
Reference − Both SDGS 2040	0.33	0.24	−0.02
Reference − Educ SDG Only	0.18	0.07	−0.09
Both SDGs (75% DS) − Both SDGs (0% Unmet)	0.13	0.05	0.00
Both SDGs 2040 − Both SDGs (75% DS)	0.05	−0.02	−0.02
Educ SDG Only − Both SDGs (0% Unmet)	0.32	0.19	0.04
Educ SDG Only − Both SDGs (75% DS)	0.19	0.14	0.04

In 2030–2035, the time period directly following the SDG intervention, the Both SDGs

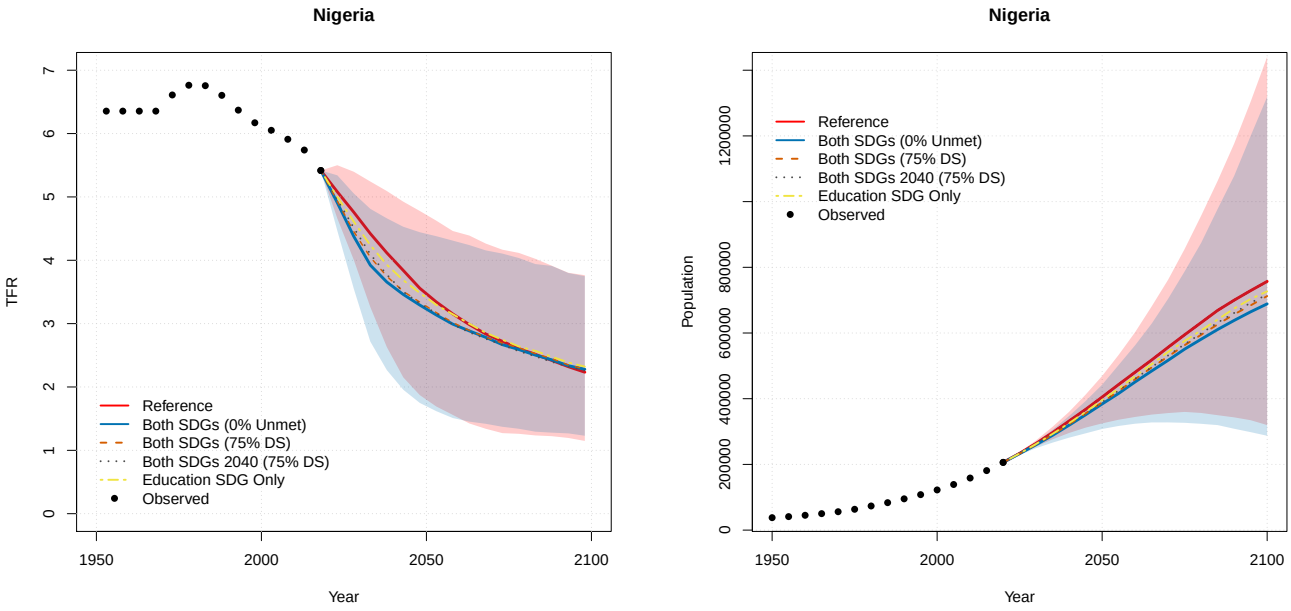


Figure 3.5: Comparison of median TFR and population projections for Nigeria from reference scenario in red, Both SDGs (0% Unmet) in dark blue, Both SDGs (75% DS) in orange dashed, Both SDGs 2040 (75% DS) in dark grey dotted, and Education SDG Only in yellow dash-dotted. 95% projection intervals for the reference and Both SDGs (0% Unmet) scenarios are also plotted.

(0% Unmet) scenario projects TFR to be 3.92 children per woman with a 95% PI of (2.72, 4.81), while the Both SDGs (75% DS) scenario projects TFR to be 4.05 (2.90, 4.90) children per woman. This is a reduction of 0.50 and 0.37 of a child, respectively, from the reference scenario projection of 4.42 (3.26, 5.24) children per woman. These differences translate into differences in median population size projections of 8.3 and 6.5 million fewer people in 2035, respectively, compared to the reference scenario. As expected, the most extreme interpretation of the SDGs, Both SDGs (0% Unmet), leads to the largest reduction in projected TFR and population size out of all intervention scenarios in 2030–2035, while the more conservative interpretations of meeting the SDGs lead to smaller reductions in projected TFR and

Table 3.3: Median population projections in 2035, 2050, and 2100 for Nigeria in millions of people for all projection scenarios with 95% PIs. Rows indicating differences between projection scenarios show differences between median projected population size.

	2035	2050	2100
Reference	297 (276, 315)	405 (325, 469)	757 (320, 1441)
Both SDGs (0% Unmet)	289 (266, 306)	384 (308, 443)	689 (287, 1319)
Both SDGs (75% DS)	290 (269, 308)	389 (312, 450)	712 (305, 1331)
Both SDGs 2040	292 (272, 308)	391 (317, 452)	717 (298, 1336)
Educ SDG Only	293 (274, 310)	396 (325, 458)	727 (326, 1411)
Reference − Both SDGs (0% Unmet)	8.3	21.0	68.4
Reference − Both SDGs (75% DS)	6.5	15.9	44.6
Reference − Both SDGS 2040	5.4	14.5	40.2
Reference − Educ SDG Only	3.9	9.1	30.3
Both SDGs (75% DS) − Both SDGs (0% Unmet)	1.8	5.1	23.7
Both SDGs 2040 − Both SDGs (75% DS)	1.2	1.4	4.5
Educ SDG Only − Both SDGs (0% Unmet)	4.4	11.9	38.0
Educ SDG Only − Both SDGs (75% DS)	2.6	6.8	14.3

population size.

From the Education SDG Only scenario, we see that attaining Target 4.1 in 2030 in the absence of any policy intervention for family planning still leads to a reduction of 0.18 of a child in median projected TFR for Nigeria compared to the reference scenario in 2030–2035. This projection incorporates both the direct effect of attaining Target 4.1 on fertility and the indirect effect of attaining Target 4.1 on fertility through the effect of education on family planning. Comparing the differences between the Education SDG Only scenario with the scenarios where both SDG targets are met allows us to quantify the additional effect of policy interventions for family planning on fertility decline. For the different interpretations of Target 3.7, the additional effect of policy interventions targeting family planning corresponds to additional reductions in median projected TFR of 0.32 or 0.19 in 2030–2035 compared to only attaining Target 4.1.

In 2045–2050, TFR is projected to be 3.29 (1.75, 4.44) in the Both SDGs (0% Unmet)

scenario and 3.34 (1.80, 4.54) in the Both SDGs (75% DS) scenario. Compared to the reference scenario projection of 3.56 (1.87, 4.78), this is a reduction of 0.27 and 0.22 of a child, respectively. These differences in TFR translate into differences in population size of 21.0 and 15.9 million fewer people, respectively, in 2050.

After mid-century, TFR projections for all scenarios begin to converge. TFR projections in 2095–2100 range between 2.23 (1.15, 3.76) in the reference scenario, 2.28 (1.23, 3.75) in Both SDGs (0% Unmet), 2.28 (1.23, 3.81) in Both SDGs (75% DS), 2.26 (1.26, 3.80) in Both SDGs 2040 (75% DS), and 2.32 (1.19, 3.74) in Education SDG Only. This convergence is due to the shared post-transition Phase III model, where once a country has entered Phase III, the country is projected to converge towards and fluctuate around a long-term mean TFR. As TFR projections across all scenarios eventually converge to the same country-specific mean, the most interesting comparisons of the projected effects of the SDG interventions refer to the period before mid-century.

Despite the convergence in TFRs, population projections remain relatively distinct across the projection scenarios out to 2100 due to population momentum. In 2100, Nigeria's population is projected to reach 757 (320, 1441) million people under the reference scenario. In the Both SDGs (0% Unmet) scenario, population is projected to reach 689 (287, 1319) million, a reduction of 68.4 million people compared to the reference scenario. In the Both SDGs (75% DS) scenario, population is projected to reach 712 (305, 1331) million, a reduction of 44.6 million people compared to the reference scenario. In the Both SDGs 2040 (75% DS) scenario, population is projected to be 717 (298, 1336) million in 2100, which is a reduction of 40.2 million people compared to the reference scenario. We note that the projected distributions of population in 2100 are very similar for the Both SDGs (75% DS) and Both SDGs 2040 (75% DS) scenarios, with substantial overlap between their 95% PIs. This overlap is unsurprising given the similarities of the TFR projections for these scenarios and the convergence of TFR projections across all scenarios after mid-century. However, it

is still notable as it suggests meeting the SDGs a decade later than the target year of 2030 could lead to similar long-term reductions in population growth for Nigeria as meeting the SDGs in 2030.

For Nigeria, the 95% PIs for population projections from the Education SDG Only scenario overlap significantly with the reference scenario projections. In 2100, population is projected to reach 727 (326, 1411) million people under Education SDG Only, compared to the reference scenario's projected 757 (320, 1441) million people. This still corresponds to a difference in median projected population size of 30.3 million people, but the boundaries of the 95% PI for the Education Only SDG scenario lie entirely within the boundaries of the 95% PI for the reference scenario. The large overlap suggests that while policy interventions targeting educational attainment in the absence of interventions for family planning may still result in some reductions in population size for Nigeria, there is also the possibility of negligible long-term impacts on population size compared to the reference scenario.

3.3.2 Sub-Saharan Africa

Next, we present projection results for sub-Saharan Africa as a whole. World population in the next century will be driven by population growth in high-fertility countries, the majority of which are in SSA. As described in Section 3.2.5, TFR projections for countries with current TFR ≥ 2.1 follow the conditional TFR projection model, while countries that currently have lower fertility are projected with the bayesTFR model. Countries without available covariate data are also projected using the bayesTFR model. Thus, our intervention-based projections of regional aggregates reflect only the impact of policy interventions in countries with TFR above replacement level that had available covariate data. For SSA, there are 41 countries that are projected using the conditional TFR projection model and nine countries that are projected using the bayesTFR model.

Figure 3.6 shows median projections of TFR and population for SSA for all projection

scenarios alongside 95% PIs for the reference and Both SDGs (0% Unmet) scenarios. Median and 95% PIs for projected values of TFR for SSA for all scenarios and differences in median projected TFR between different projection scenarios are shown in Table 3.4 for 2030–2035, 2045–2050, and 2095–2100. TFR results were aggregated for SSA as the average of the age-specific fertility rates for countries in SSA, weighted by the size of the female population for each country.[5] Table 3.5 summarizes the population projections corresponding to the TFR projection scenarios for SSA.

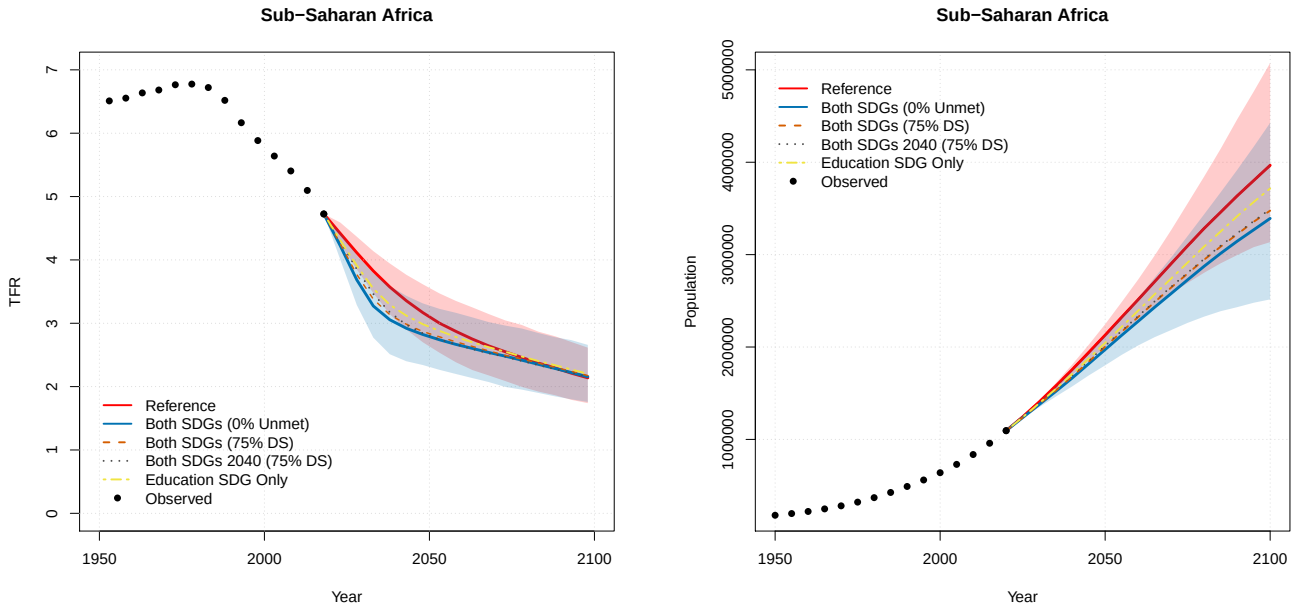


Figure 3.6: Comparison of median TFR and population projections for sub-Saharan Africa from reference scenario in red, Both SDGs (0% Unmet) in dark blue, Both SDGs (75% DS) in orange dashed, Both SDGs 2040 (75% DS) in dark grey dotted, and Education SDG Only in yellow dash-dotted. 95% projection intervals for the reference and Both SDGs (0% Unmet) scenarios are also plotted.

[5]Note that is not the exact TFR for SSA, but is likely to be a very close approximation.

Table 3.4: Median TFR projections in 2030–2035, 2045–2050, and 2095–2100 for sub-Saharan Africa in children per woman for all projection scenarios. Rows indicating differences between projection scenarios show differences between median projected TFR.

	2030–2035	2045–2050	2095–2100
Reference	3.83 (3.48, 4.14)	3.17 (2.70, 3.62)	2.14 (1.74, 2.62)
Both SDGs (0% Unmet)	3.27 (2.77, 3.79)	2.82 (2.34, 3.31)	2.16 (1.76, 2.66)
Both SDGs (75% DS)	3.38 (2.91, 3.85)	2.87 (2.38, 3.39)	2.17 (1.75, 2.70)
Both SDGs 2040	3.46 (3.04, 3.89)	2.84 (2.34, 3.35)	2.15 (1.75, 2.66)
Educ SDG Only	3.55 (3.09, 3.98)	2.99 (2.51, 3.46)	2.20 (1.77, 2.69)
Reference − Both SDGs (0% Unmet)	0.55	0.35	−0.02
Reference − Both SDGs (75% DS)	0.44	0.30	−0.04
Reference − Both SDGS 2040	0.36	0.33	−0.02
Reference − Educ SDG Only	0.28	0.18	−0.06
Both SDGs (75% DS) − Both SDGs (0% Unmet)	0.11	0.05	0.02
Both SDGs 2040 − Both SDGs (75% DS)	0.08	−0.03	−0.02
Educ SDG Only − Both SDGs (0% Unmet)	0.27	0.16	0.04
Educ SDG Only − Both SDGs (75% DS)	0.17	0.12	0.03

In 2030–2035, the Both SDGs (0% Unmet) scenario projects TFR to be 3.27 (2.77, 3.79) children per woman in SSA while the Both SDGs (75% DS) scenario projects TFR to be 3.38 (2.91, 3.85) children per woman. This is a reduction in medians of 0.55 and 0.44 of a child, respectively, from the reference scenario projection of 3.83 (3.48, 4.14). From the Education SDG Only scenario, attaining Target 4.1 in 2030 in the absence of any policy intervention for family planning leads to a reduction of 0.28 in median projected TFR compared to the reference scenario in 2030–2035. The additional effect of policy interventions for family planning leads to an additional reduction in median projected TFR of 0.27 or 0.17 for the different interpretations of Target 3.7.

In 2045–2050, the TFR for SSA is projected to be 2.82 (2.34, 3.31) in the Both SDGs (0% Unmet) scenario and 2.87 (2.38, 3.39) in the Both SDGs (75% DS) scenario. Compared to the reference scenario projection of 3.17 (2.70, 3.62), this is a reduction of 0.35 and 0.30 of a child, respectively. Attaining Target 3.1 in 2030 in the absence of any policy intervention for

Table 3.5: Median projections in 2035, 2050, and 2100 of population size for sub-Saharan Africa in millions of people for all projection scenarios with 95% PIs. Rows indicating differences between projection scenarios show differences between median projected population size.

	2035	2050	2100
Reference	1576 (1541, 1609)	2129 (2001, 2247)	3967 (3135, 5074)
Both SDGs (0% Unmet)	1518 (1462, 1569)	1973 (1801, 2121)	3392 (2515, 4427)
Both SDGs (75% DS)	1528 (1476, 1579)	2003 (1840, 2151)	3476 (2694, 4572)
Both SDGs 2040	1539 (1494, 1582)	2013 (1870, 2153)	3493 (2724, 4496)
Educ SDG Only	1546 (1496, 1592)	2044 (1894, 2197)	3716 (2887, 4738)
Reference − Both SDGs (0% Unmet)	57	156	575
Reference − Both SDGs (75% DS)	47	126	491
Reference − Both SDGS 2040	36	115	474
Reference − Educ SDG Only	30	85	251
Both SDGs (75% DS) − Both SDGs (0% Unmet)	10	30	84
Both SDGs 2040 − Both SDGs (75% DS)	11	10	17
Educ SDG Only − Both SDGs (0% Unmet)	27	72	324
Educ SDG Only − Both SDGs (75% DS)	17	41	240

family planning leads to a reduction of 0.18 of a child, and the additional effect of attaining Target 3.7 leads to a reduction of 0.16 or 0.12 of a child for the different interpretations of the family planning target.

After mid-century, we note the same convergence in TFR projections across projections scenarios that was observed in Nigeria. The TFR of SSA is projected to approach replacement level in 2095–2100 in all projection scenarios, with the slowest approach occurring in the Education SDG Only scenario. Population in 2100 in SSA is projected to reach 3.97 (3.14, 5.07) billion people in the reference scenario, 3.39 (2.52, 4.43) billion people in the Both SDGs (0% Unmet) scenario, 3.48 (2.69, 4.57) billion people in the Both SDGs (75% DS) scenario, 3.49 (2.72, 4.50) billion people in the Both SDGs 2040 (75% DS) scenario, and 3.72 (2.89, 4.74) billion people in the Education SDG Only scenario.

Population in 2100 is projected to be 575 million people lower in the Both SDGs (0%

Unmet) scenario and 491 million people lower in the Both SDGs (75% DS) scenario than the reference scenario. If the SDGs are instead met in 2040, this difference is 474 million people. We note that the projected distributions for TFR and population size for the Both SDGs (75% DS) and Both SDGs 2040 (75% DS) scenarios are very similar across time periods. The difference between median TFR projections between these two scenarios is largest at about 0.08 of a child in 2030–2035, but the projected TFR distributions become nearly indistinguishable for 2045–2050 and 2095–2100. This suggests that achieving the SDG targets a decade later in countries that are not currently on track to meet the SDGs in 2030 could still have substantial long-term demographic implications for SSA.

In 2100, if only the target of universal lower secondary education is attained with no intervention targeting family planning, population is projected to be 251 million people lower than the reference scenario for SSA. The additional effect of attaining the SDG corresponding to family planning is a reduction of 324 million people in 2100 if Target 3.7 is interpreted as 0% Unmet Need and 240 million people if the target is interpreted as 75% Demand Satisfied.

3.4 Validation

Our goal in developing the conditional TFR projection model is not to improve the predictive performance of the existing bayesTFR projection method, but to expand the utility of the method to allow for creation of policy-based intervention projections. The conditional TFR projection model should ideally extend the utility of the bayesTFR model for intervention scenario projections without reducing the predictive accuracy of bayesTFR. To evaluate this, the conditional TFR projection model was assessed using out-of-sample validation. Out-of-sample validation is a method frequently used to validate probabilistic forecasts and, in particular, was the method used to validate the original bayesTFR method (Alkema et al., 2011). The validation exercises conducted on the conditional TFR projection model are compared to an analogous exercise for bayesTFR to ensure the conditional TFR

projection model has similar predictive performance to bayesTFR.

For the first out-of-sample validation exercise, we created projections for the five-year time period 2015–2020 using the conditional projection model estimated using data spanning 1970–1975 through 2010–2015. Fertility observations came from WPP 2019 and were subject to the same "high-fertility" constraint as before, where only country-time pairs where the country was in Phase II and had TFR > 2.5 were included in estimation of the second stage. Covariate data were restricted to reflect the true data availability before 2015 as much as was possible for both estimation and projection. We used data from Version 1.2 of the Wittgenstein Centre Data Explorer to fit the educational attainment model to create out-of-sample estimates and projections of ΔLowSec+. For estimates and projections of contraceptive prevalence, the FPEMglobal model was re-estimated using only survey estimates that were available before 2015. The model for GDP was also re-estimated using only data that were available in the 1970–1975 through 2010–2015 estimation period. This first validation exercise checks the marginal predictive performance of the conditional TFR projection model, where we expect the conditional model to perform similarly to bayesTFR.

For the second validation exercise, we considered out-of-sample validation conditional on knowing the true values of the covariates for 2015–2020. The projection model was estimated using the same method as the first out-of-sample exercise, where the model is estimated leaving out data for the 2015–2020 time period. Then, TFR for 2015–2020 is projected using the left-out 2015–2020 values of the covariates as inputs. This second validation exercise checks the conditional predictive performance of the conditional TFR projection model, where we expect the conditional model to perform similarly or slightly better than bayesTFR.

The results of the two validation exercises are summarized in Table 3.6, where the results for the conditional TFR projection model are compared with analogous results for bayesTFR. Results for both models are averaged over the 97 countries included in the out-of-sample estimation. The out-of-sample validation exercises using the out-of-sample projections of

the covariates are denoted "OOS" in the "Validation Type" column, while the conditional out-of-sample validation using the left-out 2015–2020 values of the covariates is denoted "Conditional OOS." The root mean squared error (RMSE) evaluates the performance of the point predictions and is calculated as

$$\sqrt{\frac{1}{97}\sum_{c}\left(\hat{f}_{c,2015-2020} - f_{c,2015-2020}\right)^2},$$

where $\hat{f}_{c,2015-2020}$ denotes the median projection of TFR for country c and time 2015–2020, $f_{c,2015-2020}$ is the observed value of TFR for country c and time 2015–2020 from WPP 2019, and the sum is taken over all 97 countries included in the out-of-sample estimation. The conditional TFR projection model performed similarly to bayesTFR in terms of RMSE in both validation exercises. The OOS validation exercise resulted in slightly larger RMSE compared to bayesTFR, while the Conditional OOS validation exercise resulted in slightly smaller RMSE compared to bayesTFR.

The performance of the projection intervals was evaluated by computing the coverage of the 95% projection intervals with respect to the left-out observations, where coverage is averaged over all 97 countries and is calculated as the proportion of the intervals that contained the true 2015–2020 value of TFR from WPP 2019. The conditional TFR projection model performed similarly to bayesTFR, with both projection models having coverage for the 95% intervals slightly above the nominal level. The projection intervals were also evaluated by looking at the average interval width across all countries. The intervals for the Conditional OOS validation exercise were very similar in width to the bayesTFR intervals, but the intervals for the OOS validation exercise were wider on average than the intervals from bayesTFR. This follows expectations, as the OOS validation exercise incorporates uncertainty about the out-of-sample covariate projections into the projections of TFR. Based on these validation exercises and comparisons to bayesTFR, we find the conditional TFR projection model has similar predictive performance as bayesTFR. Thus, the conditional TFR projection model

is able to extend the utility of bayesTFR by enabling the creation of conditional projections based on policy intervention scenarios without sacrificing the predictive accuracy of bayesTFR.

Table 3.6: Out-of-sample (OOS) validation results for one five-year time period (2015–2020) left out for the conditional TFR projection model and bayesTFR using WPP 2019, where results are averaged across all 97 countries included in estimation of the second stage. The metrics shown are the root mean squared error (RMSE), the coverage of the 95% projection intervals (95% Cvg), and the average width of the 95% projection intervals (95% Width).

Model	Validation Type	RMSE	95% Cvg	95% Width
Conditional Projection Model	OOS	0.1234	0.9691	0.8318
Conditional Projection Model	Conditional OOS	0.1197	0.9691	0.8033
bayesTFR		0.1215	0.9691	0.8046

We also conducted a sensitivity analysis for the prior distributions of the $\boldsymbol{\beta}$ coefficients and found that the conditional TFR projection model is not sensitive to changes in the prior distributions. Further details of this analysis and other validation exercises, including checking of the conditional linearity assumption, can be found in the Appendix.

3.5 Discussion

We created a conditional projection model for TFR that extends the unconditional Bayesian hierarchical model that is the basis of the fertility projections published by the UN. The conditional TFR model enables the creation of probabilistic projections of TFR conditional on policy interventions that target educational attainment and contraceptive prevalence, such as meeting the SDG targets for education and family planning.

Previous work has explored potential ways to quantify the possible impact of the SDGs on fertility and population size. We compare our results with those of Abel et al. (2016) and Vollset et al. (2020). Abel et al. (2016) created population projections based on attaining the SDGs by building upon the population projection model developed by the Wittgenstein

Centre. The Wittgenstein Centre projections are based on global population scenarios corresponding to the Shared Socioeconomic Pathways (SSPs) used by the Intergovernmental Panel on Climate Change (Lutz et al., 2014). In lieu of reporting population projections with corresponding measures of uncertainty, the Wittgenstein Centre produces population projections following a number of different SSP scenarios corresponding to different levels of socioeconomic development.

Abel et al. (2016) extends this work to consider different SDG scenarios based on varying interpretations of the SDG targets. In particular, Target 4.1 is interpreted as either universal lower secondary education or universal upper secondary education and Target 3.7 is interpreted as meaning that education-specific fertility rates will either be 20% lower or 10% lower due to reductions in unmet need for family planning. These different SDG scenarios lead to a range of possible population projections. However, the individual SDG scenario projections do not come with measures of uncertainty.

Vollset et al. (2020) created probabilistic population projections for the Global Burden of Disease (GBD) project at IHME. The GBD model incorporates measures of education and family planning as covariates in the fertility projection model and incorporates uncertainty in projections. The GBD model projects completed cohort fertility at age 50 as a function of educational attainment (measured as years of education) and contraceptive met need. Vollset et al. consider scenarios for population projections based on different rates of change in educational attainment and contraceptive met need, including a scenario corresponding to meeting the SDGs.

Unlike the Abel et al. (2016) projections, the Vollset et al. (2020) projections do come with measures of uncertainty. However, the Vollset et al. projections have been criticized in the demographic community for questionable model assumptions and demographically implausible projection results (Gietel-Basten and Sobotka, 2020, 2021). The interpretation of their intervention-based projections as causal has also been questioned by Alkema (2020),

who highlighted a key flaw in the causal assumptions underlying the Vollset et al. model. By using met need for contraception as the measure of family planning, the Vollset et al. model does not distinguish between the effect of increased demand for family planning and the effect of improved access among those with a need for family planning in their SDG intervention scenario. We avoided this issue in our model by focusing on the relationship between contraceptive prevalence and fertility. In our model, increases in contraceptive prevalence that result from the SDG intervention correspond both to fertility reductions that are due to increases in demand for family planning and fertility reductions that are due to improved access. These two pathways leading to fertility reductions are also reflected in our implementation of the SDG intervention projections for contraceptive prevalence.

There are notable differences between our results and the two existing sets of intervention-based projections. We compare population size projections under reference and SDG intervention scenarios from our model, the Abel et al. model, and the Vollset et al. model for the regional aggregate of a subset of 72 out of the 83 countries for which we create intervention-based projections. We exclude 11 countries from the comparison due to lack of available projections for the Abel et al. model. The countries additionally excluded are Afghanistan, Angola, Bolivia, Botswana, Côte d'Ivoire, Israel, Oman, Sri Lanka, Sudan, Togo, and Yemen.

We are unable to directly compare the median population projections for the regional aggregate of the 72 countries across models, as Abel et al. (2016) and Vollset et al. (2020) do not publish individual projection trajectories. Instead, we compare an approximation for the median projected population of the regional aggregate. For the Abel et al. model, we approximate the median projected population for the regional aggregate with the sum of the (deterministic) population projections for each country in the aggregate. The SDG scenario projections for Abel et al. correspond to their SDG2 projection scenario,[6] while

[6]Population projections for individual countries were obtained from the supplementary material to Abel et al. (2016), available at `https://www.iiasa.ac.at/SDGscenarios2016` and downloaded on February 8, 2023

the reference scenario projections correspond to the SSP2 projection scenario.[7] For our model and the Vollset et al. model, we approximate the median projected population for the regional aggregate using the sum of the median population projections for each country in the aggregate.[8] The sum of median population projections over all countries in the aggregate is not equivalent to the median of the projected population distribution for the regional aggregate, but for our model we found the sum of medians to be a good approximation. Additional details of this approximation can be found in the Appendix.

Table 3.7 summarizes a comparison of the reference scenario and SDG intervention scenario population projection results from our model, the Abel et al. model, and the Vollset et al. model for the regional aggregate of 72 countries. SDG intervention results from our conditional projection model are shown for the Both SDGs (0% Unmet) and Both SDGs (75% DS) projection scenarios, where Both SDGs (0% Unmet) aligns most closely with the SDG assumptions used by Abel et al. and Vollset et al. Results from Both SDGs (75% DS) are included in the comparison to illustrate a more realistic interpretation of meeting the SDGs.

We note that there can be large differences between our reference scenario projections and those from Abel et al. (2016) and Vollset et al. (2020), particularly for 2100. One source of these differences is the input data used to estimate the three models. Our model and the Abel et al. model use similar sources of data, however Abel et al. use older versions of these sources of data. In particular, both our model and the Abel et al. model use estimates and projections of educational attainment from the Wittgenstein Centre. However, the Abel et al. results are based on the 2014 version of the educational attainment data while our

[7]Population projections for individual countries were obtained from version 1.2 of the Wittgenstein Centre Data Explorer, available at `https://dataexplorer.wittgensteincentre.org/wcde-v1/` and downloaded on February 8, 2023

[8]Reference and SDG scenario population projections for individual countries from Vollset et al. (2020) are available at `https://ghdx.healthdata.org/record/ihme-data/global-population-forecasts-2017-2100`, downloaded on February 8, 2023

results are based on the 2018 update. The older version of the Wittgenstein Centre database covers fewer countries, uses an earlier baseline year of data, and has some differences in methodology used for reconstruction of past educational attainment (Speringer et al., 2019).

There are also notable differences between the data used to estimate the Vollset et al. model and the data used to estimate our model. The Vollset et al. model uses estimates of past fertility from the GBD study at IHME, whereas our model is based on estimates from the UN. One important difference between these two sets of fertility estimates concerns SSA, where the estimates of fertility for countries in SSA from IHME tend to be lower than the estimates from the UN (Gietel-Basten and Sobotka, 2020). As countries in SSA account for about half of the countries in the regional aggregate, this impacts the overall reference scenario comparisons.

Another source of the differences between projections across the three models is the underlying differences in fertility projection assumptions in the low-fertility setting when TFR has reached below 2.1 children per woman. These low-fertility assumptions dictate the long-run fertility levels used to model the post-transition phase, which mostly impacts the population projections for countries in the regional aggregate in the latter half of the century. As our analysis is primarily conducted in the high-fertility setting and we use the post-transition projection model from bayesTFR without any modifications, details of these low-fertility modeling differences are out of the scope of this chapter but have been discussed by Wilmoth (2019), Vollset et al. (2020), and Kaneda et al. (2021), among others.

Due to these differences in reference scenario projections, we focus our comparisons on the projected differences between the reference scenario and the SDG intervention scenario from each set of projections. These projected differences are shown in the rows labeled "Difference" in Table 3.7 for 2050 and 2100. Additional comparisons of the projected differences for all projection years can be found in the Appendix.

We first consider comparisons between our results and the Abel et al. (2016) results. The

Table 3.7: Comparison of sum of median population projections for the regional aggregate of 72 countries in 2050 and 2100 under the reference model, the intervention scenario assuming both SDG targets are met in 2030, and the difference between the two scenarios from our conditional projection model (Cond. BHM), Abel et al. (2016), and Vollset et al. (2020) in billions of people. The SDG results from our conditional projection model follow the Both SDGs (0% Unmet) and Both SDGs (75% DS) scenarios. The SDG results from the Abel et al. model follow their SDG2 scenario.

Year	Scenario	Cond. BHM (0% Unmet)	Cond. BHM (75% DS)	Abel et al. (2016)	Vollset et al. (2020)
2050	Reference	5.52	5.52	5.02	5.42
	SDG	5.17	5.23	4.66	4.82
	Difference	0.35	0.28	0.36	0.60
2100	Reference	7.01	7.01	5.41	5.61
	SDG	5.98	6.17	4.66	3.73
	Difference	1.03	0.84	0.76	1.88

assumptions of the SDG2 scenario from Abel et al. align most closely with our Both SDGs (0% Unmet) scenario. Compared to Abel et al., we project larger reductions in population size in 2100 due to attaining both SDGs in 2030. Under SDG2, Abel et al. project the population of the regional aggregate to be 4.657 billion people in 2100, which is a reduction of 755 million people compared to their reference scenario projection of 5.412 billion people. In contrast, we project a reduction from 7.010 billion people to 5.984 billion people between our reference and Both SDGs (0% Unmet) scenarios, which is a difference of 1.026 billion people. The differences in projection results between our model and the Abel et al. model are much smaller in 2050. Our Both SDGs (0% Unmet) scenario projects a reduction from 5.516 billion people to 5.166 billion people, which is a difference of 350 million people from meeting the SDGs. The Abel et al. model projects a very similar difference of 360 million people in 2050 from meeting the SDGs, with a reference scenario population of 5.022 billion people and an SDG intervention scenario population of 4.662 billion people.

The projected reductions in population size from our Both SDGs (0% Unmet) scenario

roughly agree with the projected reductions in population size from Abel et al. up until mid-century, at which point the projected reductions diverge. A large part of the differences in the projected effect of the SDGs after mid-century is due to underlying differences between the fertility projection models used by the UN and the Wittgenstein Centre. Abel et al. project the population of the regional aggregate to peak slightly after mid-century in both the reference and SDG intervention scenarios, whereas our population projections continue to increase to 2100 in all scenarios. We note that the differences in projected reductions in population between our more realistic Both SDGs (75% DS) scenario and the Abel et al. projections are less pronounced, with the largest difference occurring at 2100. However, the two sets of projections are fairly similar even at 2100, with our model projecting a reduction of 835 million people and the Abel et al. model projecting a reduction of 755 million people.

Next, we consider comparisons with Vollset et al. (2020). The assumptions of the SDG intervention projection scenario from Vollset et al. also align most closely with our Both SDGs (0% Unmet) scenario. Vollset et al. found that meeting the SDG targets for education and contraceptive met need would result in population size in 2100 for the regional aggregate of 3.730 billion people compared to their reference scenario projection of 5.612 billion people, which is a reduction in population of 1.882 billion people. In comparison, we project a smaller reduction of 1.026 billion people in 2100 from our Both SDGs (0% Umet) scenario. Similar differences between the Vollset et al. projections and our projections occur in earlier time periods. In 2050, Vollset et al. projects population to be 5.418 billion people in the reference scenario and 4.820 billion people in the SDG intervention scenario, which is a difference of 598 million people. In our Both SDGs (0% Unmet) scenario, we project a smaller difference of 350 million people.

The projected reductions in population from Vollset et al. are much larger than the projected reductions from our results and from Abel et al. By 2100, the projected reduction in population from Vollset et al. is about 1.83 times as large as the projected reduction

in population from our Both SDGs (0% Unmet) scenario and about 2.25 times as large as the projected reduction from our Both SDGs (75% DS) scenario. Part of the differences between our projections and the Vollset et al. projections are due to underlying differences in fertility model assumptions. Vollset et al. project the population of the regional aggregate to peak around or slightly after mid-century in both reference and SDG scenarios, whereas our population projections continue to increase to 2100. A comparison of the reference scenario projections for world population from the GBD model, the Wittgenstein Centre model, and the UN model was conducted in Vollset et al. (2020). They found that for the world aggregate, the GBD projections align more closely with the results from the Wittgenstein Centre than with results from the UN for both fertility and population projections. However, despite these similarities in reference scenario projections, it is notable that Vollset et al. project the effect of meeting the SDGs for the regional aggregate to be about 1.66 times the effect projected by Abel et al. in 2050 and almost 2.5 times the effect projected by Abel et al. in 2100.

Our method improves upon the existing intervention-based projections of TFR and population in several ways. First, unlike the Abel et al. model, our model fully incorporates uncertainty about covariate projections into projections of TFR and provides probabilistic projections of TFR and population size. Second, unlike the Vollset et al. model, our model incorporates demographic background knowledge to ensure that results are demographically plausible and to inform the causal framework underlying the model.

Our projections also reflect only the impact of policy interventions in the high-fertility setting. While the accelerating effect of education and family planning expansion on fertility decline is well established in the high-fertility context, the effect is less clear in the low-fertility context, in particular for education. Although there is some evidence that the effect of education on fertility persists over the course of the fertility transition (Lutz et al., 2014), there is also evidence that the effect of education may be different in countries where a

majority of women have attained tertiary education (Adserá, 2017b). For example, there is evidence to suggest that the relationship between education and fertility may in fact be reversed in the low-fertility setting, where higher educated women have greater resources to attain their desired childbearing and thus have higher fertility compared to lower educated women (Sobotka et al., 2017). This uncertainty about the long-run relationship between education and fertility decline once a country reaches low fertility levels is not reflected in either the Abel et al. or Vollset et al. projections.

Unlike the existing projection models, our model also accounts for the fact that the associations between the covariates and fertility decline have been observed to be weaker in SSA compared to other regions of the world. This difference is especially important in the context of intervention-based projections, where, for example, the effect of eliminating unmet need for family planning may have a weaker effect on fertility decline in SSA compared to other regions of the world due to high ideal family sizes (Bongaarts and Casterline, 2013).

Despite these improvements on the existing SDG intervention-based projections, our results have several limitations. The policy intervention scenarios rely on statistical extrapolation for projections of educational attainment and contraceptive prevalence. The SDG projections of the covariates are extreme scenarios, where the amount of acceleration in educational attainment and contraceptive prevalence encoded in the SDG intervention projections has not been observed historically. The intervention scenarios also assume that the historical relationships between education, family planning, and fertility will hold in the extrapolation, which may not be the case. This limitation is shared with all existing SDG intervention projections, and indeed is acknowledged by both Abel et al. (2016) and Vollset et al. (2020).

Our projections are also limited in terms of interpretation as causal effects due to the simplifying assumptions needed to model the complex causal relationships between education, family planning, and fertility. The assumptions used in our model are outlined in Section

3.2.3 and are based on demographic background knowledge. However, this does not preclude the possibility that our model omits confounders or overly simplifies the underlying causal structure. The accuracy of our projections is further limited by the accuracy of the estimates and projections of the covariates. We used the best data available for globally and historically comparable estimates and projections of the covariates, but these still have limitations. For example, the estimates of contraceptive prevalence used in our model do not take the effectiveness of different contraceptive methods into account, which Bongaarts (2017) identifies as a key aspect of analyzing the relationship between contraceptive prevalence and TFR.

The proposed conditional TFR projection model is additionally constrained to the five-year time scale. This restriction is primarily due to the availability of the educational attainment data, where estimates of historical educational attainment from the Wittgenstein Centre are only available in five-year increments and the projection model for educational attainment was developed under the assumption of five-year increments. Due to this constraint, we used estimates of TFR from the 2019 revision of the UN *World Population Prospects*, which uses a five-year time scale, rather than the more recent 2022 revision, which uses a one-year time scale. Extending the conditional TFR projection methodology to the one-year time scale is of interest for future work. This could be done by building upon the one-year extension of the bayesTFR model developed by Liu et al. (2023) and would ideally use one-year estimates and projections of all covariates, including educational attainment from the Wittgenstein Centre.

Finally, our population projection results assume that the SDG intervention scenarios affect population size only through the impact of the interventions on fertility. All population projections are created using non-intervention projections of mortality and migration. The potential impacts of universal secondary education and universal access to family planning on mortality and migration are not known to be substantial, and seem likely to be much smaller than the effects on fertility.

3.6 Conclusion

We have developed a conditional Bayesian hierarchical model for projections of TFR that incorporates the effect of women's educational attainment, contraceptive prevalence, and GDP per capita. This model creates probabilistic projections of TFR conditional on projections of the covariates, where the covariate projections can correspond to policy intervention outcomes. These conditional TFR projections could be used to answer questions about the likely effect of education and family planning policies on future fertility and could be an informative tool for policymakers in high-fertility countries. Given a specific policy intervention scenario, the TFR and population projections resulting from the conditional TFR projection model could help policymakers plan for the future infrastructure needs of their constituencies such as schools, health care, and transportation. This, in turn, could be used to determine the allocation of resources in support of expanding girls' education and expanding access to family planning.

As an illustrative policy intervention, we focused on attaining the SDG Targets 3.7 and 4.1, which target universal access to family planning and universal secondary education, respectively. We created projections of TFR for five-year time periods covering 2020–2025 to 2095–2100 conditional on several policy-based intervention scenarios for differing rates of increase in educational attainment and contraceptive prevalence corresponding to different translations of the SDG targets. Using the intervention-based projections of TFR, we created corresponding probabilistic projections of population size for each intervention scenario.

One potential use of the conditional projection results could be to determine the relative allocation of resources to support girls' education and to support expansion of voluntary family planning programs through comparing the projected outcomes from the Education SDG Only scenario with the scenarios that assume both SDGs targets are met, such as the Both SDGs (75% DS) scenario. For sub-Saharan Africa, population size in 2100 is projected to reach 3.72 (2.89, 4.74) billion people under the Education SDG Only scenario,

3.48 (2.69, 4.57) billion people under the Both SDGs (75% DS) scenario, and 3.97 (3.14, 5.07) billion people in the reference scenario. While these results suggest notable acceleration in fertility decline can result from rapid expansion of educational attainment, it is likely that the combination of expansion of education *and* expansion of access to family planning is required to see the full effect of meeting the SDGs on future fertility and population size. Reaching the target of universal lower secondary education in 2030 in the absence of increased investment in family planning corresponds to a reduction in median projected population size for sub-Saharan Africa in 2100 of 251 million people. If the target of reaching 75% of demand for family planning satisfied by modern methods is also met in 2030, this corresponds to an additional reduction of 240 million people. Based on the median projections, about half of the potential reduction in population size in 2100 from meeting SDG Targets 4.1 and 3.7 can be attributed to the effect of meeting the education target and about half can be attributed to the effect of meeting the family planning target. These results indicate that expanding education for girls is a worthwhile policy goal to pursue on its own that, in addition to having other benefits to society, can contribute to accelerated fertility decline and a reduction in population growth. However, if accelerating fertility decline is a key policy goal, this comparison also suggests that increased access to family planning is a worthwhile investment.

Another potential application for the conditional TFR projection model is in supporting arguments for a slower, possibly more realistic expansion of education and family planning to achieve SDG Targets 4.1 and 3.7 in a later year. For many currently high-fertility countries, meeting these SDG targets in 2030 may be an unrealistic goal. We found that meeting the SDG targets a decade later in 2040 could still have substantial impacts on the future trajectory of TFR and population size. For sub-Saharan Africa, we found the median population size in 2100 in the Both SDGs (75% DS) scenario is projected to be 3.48 billion people with a 95% PI of (2.69, 4.57). If the same policy goals are met a decade later in 2040, population

in 2100 is instead projected to be 3.49 (2.72, 4.50) billion people. Compared to the reference scenario, both of these intervention scenarios correspond to a reduction in median projected population in 2100 of over 470 million people. While meeting the SDGs in 2030 does lead to a larger reduction in projected median population in 2100, there is a substantial overlap between the projected population distributions for meeting the SDGs in 2030 or in 2040. The results from the conditional TFR projection model show that a sustained, slower investment over a longer time period could have comparable long-term effects on fertility and population size as a shorter-term, larger investment in education and family planning while also being more realistic to achieve.

Conditional TFR projections could be of use not only within the context of empowering women and girls and expanding the ability for individuals to achieve their desired child-bearing goals, but also for policymakers to consider more generally in conversations about sustainable development. The negative environmental impact of rapid population growth is well-documented (e.g. Bongaarts, 2016; Lutz, 2023; O'Neill et al., 2010), and policy discussions related to environmental sustainability and climate change must necessarily take information about future population size into account. The population projections created using the conditional TFR projections could help guide these discussions by providing a probabilistic view of what future population might look like under different education and family planning intervention scenarios. These intervention-scenario-based projections of population could then be translated into potential futures regarding long-term issues like food production, carbon emissions, and climate change. In turn, these discussions regarding sustainable development could provide a further rationale for increased investment in education and family planning.

Our findings suggest that attainment of SDG Targets 3.7 and 4.1 are likely to have substantial long-term effects on fertility decline and population growth in the high-fertility setting. Notable reductions in population growth are projected to occur even if the targets are

met a decade later than the target achievement date of 2030 and even if increased investment in girls' education occurs without increased investment in family planning programs. These results show that pursuit of the SDGs, even at a slower pace than implied by the target year of 2030, is a worthwhile policy goal that could result in substantial shifts in future demographic trends.

Chapter 4

MULTIPLE IMPUTATION OF HIERARCHICAL NONLINEAR TIME SERIES DATA WITH AN APPLICATION TO SCHOOL ENROLLMENT DATA

Hierarchical time series data sets based on survey data, such as annual country-level estimates of school enrollment rates, can suffer from large amounts of missing data due to differing coverage of surveys across countries and across times. A popular approach for handling missing data in these settings is through multiple imputation, which can be especially effective when there is an auxiliary variable that is strongly predictive of and has a smaller amount of missing data than the variable of interest. However, standard methods for multiple imputation of hierarchical time series data can perform poorly when the auxiliary variable and the variable of interest have a nonlinear relationship. Performance of standard multiple imputation methods can also suffer if the substantive analysis model of interest is uncongenial to the imputation model, which can be a common occurrence when the imputation phase is conducted independently of the analysis phase.

In this chapter, we propose a Bayesian method for multiple imputation of hierarchical nonlinear time series data that uses a sequential decomposition of the joint distribution and incorporates smoothing splines to account for nonlinear relationships between variables. We compare the proposed method with existing multiple imputation methods through a simulation study and an application to secondary school enrollment data. We find the proposed method can lead to substantial performance increases for estimation of parameters in uncongenial analysis models and for prediction of individual missing values.

4.1 Introduction

Missing values within hierarchical time series data sets are a common occurrence in social science data, particularly for comparisons across countries and across times that rely on survey data. International surveys may only be conducted in selected years and coverage of countries may differ from year to year, while country-level surveys and censuses may be conducted annually in some countries but may only occur sporadically in others. The presence of missing data can be compounded in settings that require the compilation of data from multiple different sources. One example where this occurs is for school enrollment data, where the UNESCO Institute for Statistics compiles survey and administrative data to create annual, internationally comparable estimates of school enrollment rates. The differing availability of survey and administrative data for each country and year leads to a large number of country-years in the UNESCO database where enrollment data is missing. Researchers interested in questions comparing survey-based indicators like enrollment rates across countries and across times can thus encounter large amounts of missing hierarchical time series data.

For some hierarchical time series variables, there may be related variables that that measure similar underlying concepts but have different amounts of missing data. For example, this can occur if a survey is designed to first ask for more specific information, but, failing to obtain the specific information, then asks for more general information. If the variable of interest for analysis is the variable that has a greater amount of missing data, researchers might be interested in how to best leverage the information from the auxiliary variable that has less missing data to impute the variable of interest. This situation arises with school enrollment data, where two commonly reported measures of enrollment rates are the Net Enrollment Rate (NER) and the Gross Enrollment Ratio (GER). Measurement of NER requires knowledge of the number and age distribution for children who are enrolled, while measurement of GER only requires knowledge of the number of children enrolled. The com-

parative ease of measuring GER can result in more missing values for NER compared to GER. If researchers are more interested in analyses using NER, they may want to use the available information about the auxiliary variable GER to help impute missing values for NER using a multiple imputation procedure.

Multiple imputation, first developed by Rubin (1978, 1987), is a widely used approach for handling missing data. In multiple imputation, $M > 1$ imputed values for missing observations are sampled from the posterior predictive distribution of the missing data given the observed data. The M imputed values result in M completed data sets, each of which consists of the observed data and one set of imputed values for the missing observations. The completed data sets can each be analyzed separately using complete data methods and the results of the analyses can be combined into one final, pooled result using combining rules from Rubin (1987) that account for both within-imputation variation and between-imputation variation.

For hierarchical data, multiple imputation approaches that do not explicitly account for the hierarchical structure of the data can lead to biased results in downstream analyses (Enders et al., 2016; Lüdtke et al., 2017; Taljaard et al., 2008). Many approaches specifically for multiple imputation of hierarchical time series data have been developed (e.g. Enders et al., 2020; Grund et al., 2021; He et al., 2011; Liu et al., 2000; Speidel et al., 2018; among others), with two of the most widely used approaches for social science data being Amelia II and multilevel extensions of Multiple Imputation by Chained Equations (MICE). Amelia, originally developed by King et al. (2001) and extended as Amelia II by Honaker and King (2010), is a multiple imputation method designed specifically for hierarchical time series data. Amelia is based on the joint modeling approach to multiple imputation, where imputed values are sampled from a joint distribution for all variables with missing data. MICE is a multiple imputation method developed by van Buuren and Groothuis-Oudshoorn (2011) that uses the fully conditional specification (FCS) approach to multiple imputation. Rather

than explicitly specifying a joint imputation model, the FCS algorithm iteratively samples from univariate conditional imputation models for all variables with missing data until convergence is reached. Several methods that account for hierarchical data structures have been implemented within the MICE framework, including the linear mixed effects method developed by Schafer and Yucel (2002). These commonly used methods assume that variables in the imputation model have a linear relationship. In settings where variables have a strong nonlinear relationship and transformation to approximate linearity is not possible, these methods are misspecified and can result in poor imputations.

We focus on the setting where the imputation of missing values and the analysis of imputed values are conducted independently. The development of multiple imputation originated in this setting, where Rubin (1977, 1978) proposed the multiple imputation framework as a way to provide imputed values for missing responses in public-use releases of large data sets from sample surveys. An appealing feature of multiple imputation for practitioners is the ability to use the same imputed data set to conduct many different analyses (Rubin, 1987; Schafer, 1997a). However, using the same multiply imputed data for different analyses can lead to the analysis model being uncongenial to the imputation model in the sense of Meng (1994). The analysis model and the imputation model are congenial if the imputation model contains at least as much information as the analysis model. Many of the theoretical properties of multiply imputed estimates, such as the consistency of estimates using Rubin's combining rules, rely on congeniality (Meng, 1994; Rubin, 1996; Xie and Meng, 2017). In practice, uncongeniality of the analysis and imputation models can be a regular occurrence when the researchers that collect the survey data create imputed values for publication that are then used in analyses by external researchers. The external researchers may not have access to the same information or resources needed to create imputations of their own, for example if the imputation process is not well-documented or if the predictive variables used in the imputation model are not publicly available. In this scenario, the analysis models used

by the external researchers are not guaranteed to be congenial to the imputation model. The ability of a multiple imputation method to perform well for uncongenial analyses is thus of interest for practitioners.

In this chapter, we propose a multiple imputation method for continuous hierarchical time series data that can account for nonlinear relationships between variables. We refer to this method as MINTS for Multiple Imputation of hierarchical Nonlinear Time Series data. We focus on the bivariate setting, where one variable is the variable of interest for which imputations are desired and the second variable is an auxiliary variable that is easier to measure and has a nonlinear relationship with the variable of interest. We also focus on a specific type of nonlinear relationship where the nonlinearity takes the form of a piecewise linear function. We evaluate the out-of-sample validation performance of MINTS using a simulation study that compares the performance of the proposed method with existing multiple imputation methods using simulated data, where we focus on estimating parameters in analysis models that are uncongenial to the imputation model. We also conduct two validation exercises using a motivating data set on secondary school enrollment rates, where we evaluate the predictive performance for estimating individual missing values and for estimating parameters in uncongenial analysis models. Finally, we report the imputation results using the full enrollment data.

This chapter is structured as follows. In Section 4.2, we describe the motivating case study of secondary school enrollment data in further detail. Section 4.3 describes the proposed multiple imputation method. In Sections 4.4 and 4.5, we evaluate how well the proposed method predicts missing observations and how well the multiply imputed data sets from the proposed method perform at estimating parameters in uncongenial analysis models through a simulation study and an application to the enrollment data. Section 4.6 includes further discussion and comparison to existing imputation methods. Finally, we summarize the findings of this chapter in Section 4.7.

4.2 Motivating Case Study: Secondary School Enrollment Rates

The UNESCO Institute for Statistics collects internationally comparable data on education indicators on an annual basis for all countries of the world, based largely on survey and administrative data (UNESCO Institute for Statistics, 2023). The World Bank combines this education data with population data from the United Nations Population Division to create estimates of two types of enrollment rates: the Net Enrollment Rate (NER) and the Gross Enrollment Ratio (GER) (World Bank, 2021). We focus on secondary school enrollment. NER is the ratio of children of official secondary school age who are enrolled in secondary school to the population of official secondary school age children. NER is bounded between 0% and 100% and both the numerator and denominator reflect children of official secondary school age. GER is the ratio of total enrollment in secondary school, regardless of age, to the population of official secondary school age children. The numerator and denominator for GER potentially represent different populations, where children who are not of official secondary school age can be counted in the numerator but not in the denominator. Thus, GER can be greater than 100% if children who are enrolled in secondary school are not of official school age. The two measures of enrollment are subject to the boundary NER $\leq$ GER and have a strong nonlinear relationship that can be well-approximated by a piecewise linear function.

For substantive analyses, one measure of enrollment may be preferred over the other. NER can be thought of as a demographic rate, where the numerator counts the number of enrollments for the population of children of official secondary school age in a given year and the denominator counts the person-years lived in that population for the given year. NER can thus be preferable over GER for demographic analyses. However, historical time series of NER tend to have more missing values than time series of GER. Measurement of NER is more difficult than measurement of GER due to requiring knowledge of the age distribution for children enrolled in secondary school. Measurement of GER is comparatively easy, as

GER can be calculated using only the number of children who are currently enrolled. For school systems that do not have robust recordkeeping systems and countries that do not have good vital registration systems, knowledge of the age of all enrolled children can be difficult to obtain. The greater availability of estimates of GER and the strong relationship between NER and GER motivates the desire to impute missing values of NER using the relationship between NER and GER.

We obtain estimates of secondary school enrollment rates for both genders combined from World Bank (2021),[1] where the definition of secondary school used for each country is based on the International Standard Classification of Education. After excluding all countries and years in the World Bank data base with no observations for either NER or GER, the resulting data set includes 202 countries and 51 years spanning 1970 to 2020 for a total of 10,302 country-year combinations. The overall rate of missingness is about 73.0% for NER and about 37.6% for GER. Within countries, the rate of missingness for NER ranges from about 13.7% missing in Malta to 100% missing in 14 countries. For GER, the rate of missingness within countries ranges from about 2.0% missing in Peru to about 98.0% missing in Curaçao.

Figure 4.1 shows a scatter plot of the complete cases for NER and GER. The red line illustrates the B-spline of degree 1 fit to the complete cases using the A-splines methodology described in Section 4.3.2. There is a nonlinear relationship between NER and GER, with a shift in trend occurring around GER = 100. The variation of NER about the fitted spline also appears to vary with GER, with smaller variability at the lowest levels of GER and larger variability around GER = 100.

Time series of NER and GER are shown in Figure 4.2 for Afghanistan, Belgium, Spain, and Nigeria. Afghanistan is an example of the most common type of pattern seen for individual countries, where there is a larger number of observed values for GER compared

[1]Downloaded on August 5, 2021

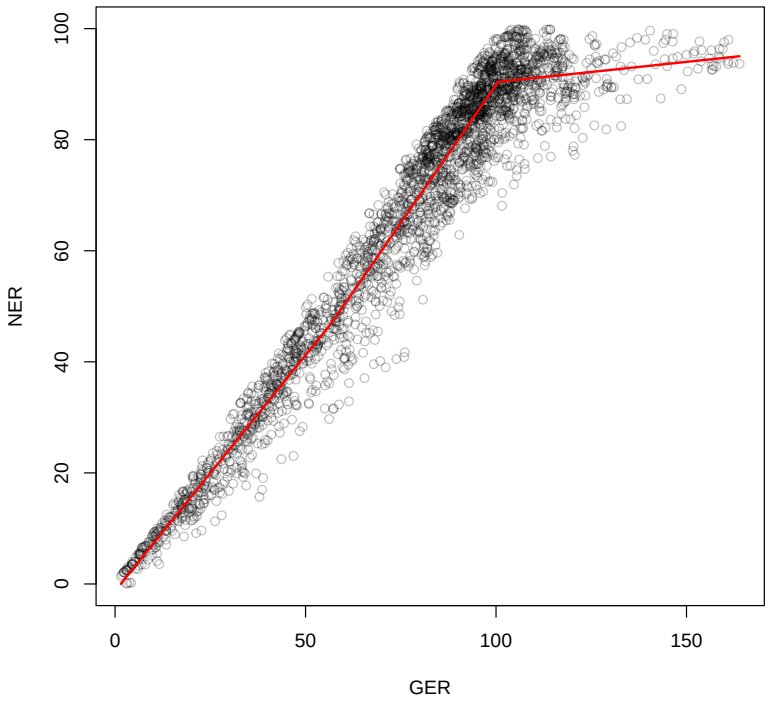


Figure 4.1: Scatter plot of complete cases for NER and GER from secondary enrollment data set, with the B-spline of degree 1 fit using A-splines superimposed in red.

to NER. Only one country, Brazil, has the opposite pattern with one more observed value of GER than observed values of NER. Belgium and Spain are two examples of countries where the nonlinear relationship between NER and GER is visible in the time series for the individual country. Both Belgium and Spain have relatively few missing values for GER, but have large stretches of time with no observations for NER. Finally, Nigeria is an example of a country that has some observed values for GER but has no observed values for NER. There are 14 countries in the enrollment data set that, like Nigeria, have at least one observation for GER but no observations for NER.

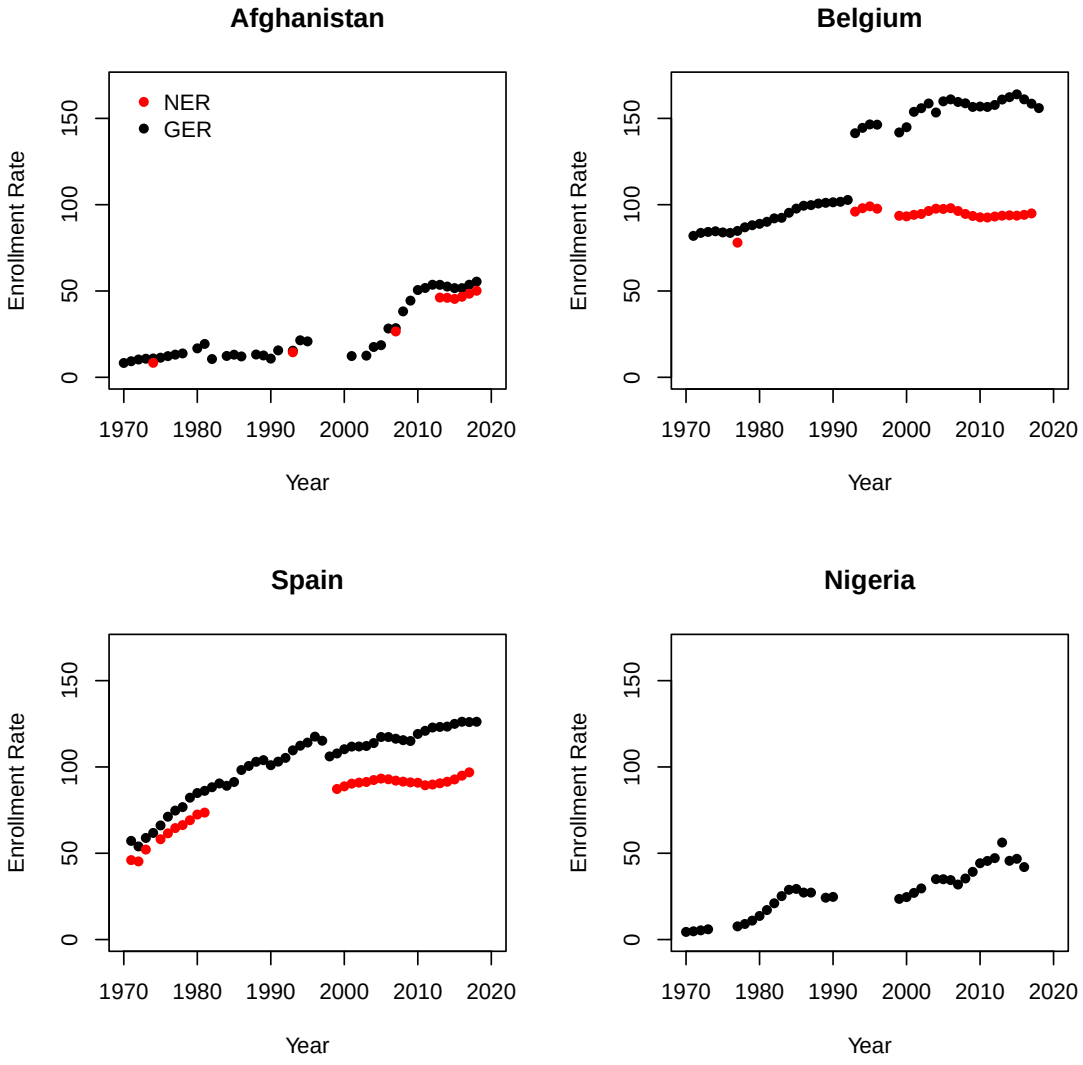


Figure 4.2: Observed values of NER and GER for selected countries from secondary enrollment data set. Solid black and red circles indicate observed values for GER and NER, respectively.

4.3 Methods

4.3.1 Notation

We consider hierarchical time series data where the clustering variable is country and the time variable is year. Let $X_{c,t}$ denote the auxiliary variable and let $Y_{c,t}$ denote the variable of interest for country c and year t, where $c \in 1, ..., C$ and $t \in 1, ..., T$. The vector of $X_{c,t}$ for all countries at year t is denoted by $\mathbf{X}_t = [X_{1,t}, X_{2,t}, ..., X_{C,t}]$. Similarly, the vector of $Y_{c,t}$ for all countries at year t is denoted by $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}, ..., Y_{C,t}]$. The C by T matrix of all $X_{c,t}$ is $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_T]$ and the C by T matrix of all $Y_{c,t}$ is $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_T]$.

Let the matrices $\mathbf{R}^X$ and $\mathbf{R}^Y$ denote the response matrices for $\mathbf{X}$ and $\mathbf{Y}$, respectively. If an element (c, t) is observed, then the corresponding element of the response matrix equals 1. For example, if $X_{c,t}$ is observed, then $R^X_{c,t} = 1$. If $X_{c,t}$ is missing, then $R^X_{c,t} = 0$. The matrices $\mathbf{X}$ and $\mathbf{Y}$ can also be written in terms of their observed and missing portions, e.g. $\mathbf{X} = [\mathbf{X}_{mis}, \mathbf{X}_{obs}]$ where the observed portion of $\mathbf{X}$ is denoted $\mathbf{X}_{obs}$ and the unobserved portion of $\mathbf{X}$ is denoted $\mathbf{X}_{mis}$.

For the enrollment data, the auxiliary variable $\mathbf{X}$ is GER and the variable of interest $\mathbf{Y}$ is NER. The enrollment data includes a total of $C = 202$ countries, where the assignment of countries to indices $c \in 1, ..., C$ was done alphabetically by country name. There are $T = 51$ years in the enrollment data set, where $t = 1$ corresponds to the year 1970 and $t = 51$ corresponds to the year 2020.

4.3.2 Model

Assumptions

We assume the missing data mechanism is ignorable, i.e. the missing data mechanism is Missing At Random (MAR, defined in Section 4.4.3) and the parameters of the complete data model for $(\mathbf{X}, \mathbf{Y})$ are distinct and a priori independent from the parameters of the missing

data model governing the response matrices $\mathbf{R}^X, \mathbf{R}^Y$ (Little and Rubin, 2002; Rubin, 1976; Schafer, 1997a). The assumption of ignorability allows for imputation without specification of a model for the missing data mechanism. Using $\mathbf{X}$ as an example, imputed values are said to be Bayesianly proper in the sense of Schafer (1997a) if they are independently drawn from the posterior distribution $p(\mathbf{X}_{mis}|\mathbf{X}_{obs}, \mathbf{R}^X)$. When the missing data mechanism is ignorable, proper imputations can instead be drawn from the posterior distribution $p(\mathbf{X}_{mis}|\mathbf{X}_{obs}) = \int p(\mathbf{X}_{mis}|\mathbf{X}_{obs}, \theta)p(\theta|\mathbf{X}_{obs})d\theta$ where θ is the vector of parameters for the complete data. The assumption of ignorability is ubiquitous for general-purpose imputation models, but in practice researchers cannot know if this assumption is met. We conducted sensitivity analyses in Sections 4.4 and 4.5 to evaluate how well the MINTS method performs when the ignorability assumption is violated.

Finally, we assume that the auxiliary variable is observed at least once for each country for which we are interested in imputing the variable of interest. That is, both $\mathbf{X}$ and $\mathbf{Y}$ can contain missing values for any country-year, but we assume that each country has at least one year in which $X_{c,t}$ is observed. This reflects the motivating setting for the proposed imputation method, where the auxiliary variable is more likely to be observed in each country than the variable of interest due to being easier to measure. This assumption is satisfied for the enrollment data, where after the exclusion of countries that have no observed values for either NER or GER, the remaining countries all have at least one observed value of GER.

Sequential decomposition of joint model

The MINTS method uses a variation of the joint modeling approach to multiple imputation. In the usual joint modeling approach, the complete data $(\mathbf{X}, \mathbf{Y})$ is assumed to follow a multivariate joint distribution and imputed values are sampled from the joint posterior predictive distribution of the missing data given the observed data. One downside to this approach is the difficulty of specifying a joint distribution for all variables used in the impu-

tation model. A common general-purpose choice for the joint distribution is the multivariate normal distribution, which has been found to still perform reasonably for imputation even when the assumption of normality is violated (Schafer, 1997a; Schafer and Olsen, 1998). However, if there are nonlinear relationships between variables in the imputation model, joint modeling assuming multivariate normality is unable to incorporate those relationships.

A more flexible approach to multiple imputation is the fully conditional specification (FCS) approach, also known as imputation via chained equations (van Buuren et al., 2006) and sequential regression multivariate imputation (Raghunathan et al., 2001), among other names. FCS does not explicitly specify a joint distribution for the variables with missing data. Instead, univariate conditional distributions are specified for each variable with missing data given all the other variables in the imputation model. Missing values are imputed by iteratively sampling from the univariate conditional distributions until convergence for all imputation model parameters is reached. The conditional distributions can take any form and can accommodate nonlinear relationships between variables. However, this level of flexibility can lead to cases where the univariate conditional distributions are not compatible. Although simulation studies show that FCS can result in reasonable imputations even when the conditional distributions are not compatible (e.g. van Buuren et al., 2006), ultimately FCS does not guarantee that the iterative conditional algorithm will converge to a proper joint distribution.

We use an alternative to the usual joint modeling approach that combines desirable features of joint modeling and FCS through a sequential decomposition of the joint model. The sequential decomposition approach for joint modeling was first proposed by Lipsitz and Ibrahim (1996) and has been extended by Ibrahim et al. (2002, 1999); Lee and Mitra (2016); Lüdtke et al. (2020); Xu et al. (2016); among others. The joint distribution of the variables with missing data is decomposed into a sequence of univariate conditional distributions. One

possible decomposition of the joint distribution for $(\mathbf{X}, \mathbf{Y})$ is

$$p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta}) = p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}_Y)p(\mathbf{X}|\boldsymbol{\theta}_X). \tag{4.1}$$

The vector of parameters for the joint distribution is $\boldsymbol{\theta}$, the vector of parameters for the conditional distribution of $\mathbf{Y}|\mathbf{X}$ is $\boldsymbol{\theta}_Y$, and the vector of parameters for the distribution of $\mathbf{X}$ is $\boldsymbol{\theta}_X$. Unlike in FCS, this sequence of univariate conditional distributions is guaranteed to correspond to a well-defined joint distribution by construction.

The sequential decomposition approach allows for greater flexibility compared to joint modeling, but is not quite as flexible as FCS due to the additional restriction of requiring an ordering for the conditional distributions. The choice of ordering can have a substantial impact on the performance of the imputation model due to the risk of conditional distributions being incorrectly specified. We follow standard guidelines proposed by Rubin and Schafer (1990) and choose the ordering based on the percentage of missing values. For the ordering in Equation 4.1, the variable of interest $\mathbf{Y}$ has a larger amount of missing data compared to the auxiliary variable $\mathbf{X}$.

Model specification

The joint distribution of $(\mathbf{Y}, \mathbf{X})$ is decomposed sequentially following Equation 4.1 as the product of the distribution of $\mathbf{Y}|\mathbf{X}$ and the distribution of $\mathbf{X}$. The distribution of $\mathbf{X}$ is further decomposed as

$$p(\mathbf{X}|\boldsymbol{\theta}_X) = \left(\prod_{t=1}^{T} p(\mathbf{X}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta}_X) \right) p(\mathbf{X}_0|\boldsymbol{\theta}_X),$$

where $\mathbf{X}_0$ is a parameter that represents the vector of $X_{c,0}$ for all countries c at the unobserved year $t = 0$. Similarly, the conditional distribution of $\mathbf{Y}|\mathbf{X}$ is further decomposed as

$$p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}_Y) = \left(\prod_{t=1}^{T} p(\mathbf{Y}_t|\mathbf{Y}_{t-1}, \mathbf{X}_t, \boldsymbol{\theta}_Y) \right) p(\mathbf{Y}_0|\mathbf{X}_0, \boldsymbol{\theta}_Y),$$

where $\mathbf{Y}_0$ is a parameter that represents the vector of $Y_{c,0}$ for all countries c at the unobserved year $t = 0$.

The distribution of $\mathbf{X}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta}_X$ is modeled as a random walk with a country-specific drift term γ_c. For country c and $t \in 1, ..., T$,

$$X_{c,t}|X_{c,t-1}, \boldsymbol{\theta}_X \sim TN_{[X_{low}, X_{up}]}(X_{c,t-1} + \gamma_c,\, \sigma_X^2), \tag{4.2}$$

where $TN_{[X_{low}, X_{up}]}$ refers to the truncated normal distribution with lower bound X_{low} and upper bound X_{up}. These boundaries can vary with (c, t), but the dependency is suppressed in the notation.

The conditional distribution of $\mathbf{Y}_t|\mathbf{Y}_{t-1}, \mathbf{X}_t, \boldsymbol{\theta}_Y$ is modeled with a country-specific intercept α_c, a nonlinear function f of $\mathbf{X}$ with coefficient β, and an AR(1) term with autoregressive parameter ρ. The variance model for $\mathbf{Y}_t|\mathbf{Y}_{t-1}, \mathbf{X}_t, \boldsymbol{\theta}_Y$ models heteroscedasticity as a function h of $\mathbf{X}$. For country c and $t \in 1, ..., T$,

$$Y_{c,t}|Y_{c,t-1}, X_{c,t}, \boldsymbol{\theta}_Y \sim TN_{[Y_{low}, Y_{up}]}\left(\alpha_c + \beta\, f(X_{c,t}) + \rho\, Y_{c,t-1},\, \sigma_Y^2\, h(X_{c,t})\right), \tag{4.3}$$

where $TN_{[Y_{low}, Y_{up}]}$ refers to the truncated normal distribution with lower bound Y_{low} and upper bound Y_{up}. These boundaries can vary with (c, t), but the dependency is suppressed in the notation.

Prior distributions

The prior distributions are specified as

$$\begin{aligned}
\sigma_X^2 &\sim InvGamma(2, \delta_X),\\
\gamma_c &\sim N(\mu_{drift}, \sigma_{drift}^2),\\
\mu_{drift} &\sim N(\nu_{drift}, \zeta_{drift}^2),\\
\sigma_{drift^2} &\sim InvGamma(2, \delta_{drift}),
\end{aligned}$$

$$
\begin{aligned}
\sigma_Y^2 &\sim InvGamma(2, \delta_Y), \\
\alpha_c &\sim N(\mu_0, \sigma_0^2), \\
\mu_0 &\sim N(0, \zeta_0^2), \\
\sigma_0^2 &\sim InvGamma(2, \delta_0), \\
\beta &\sim N(0, 1), \\
\rho &\sim U(0, 1),
\end{aligned}
$$

where the hyperparameters $\delta_X, \nu_{drift}, \zeta_{drift}, \delta_{drift}, \delta_Y, \zeta_0,$ and δ_0 are control parameters that are used to adjust the prior distributions to the appropriate scale for the data.

For all c, the joint prior distribution of $(Y_{c,0}, X_{c,0})$ is a truncated normal distribution with control parameters $\boldsymbol{\mu}_{early}$ and $\boldsymbol{\Sigma}_{early}$ given by

$$
\begin{bmatrix} Y_{c,0} \\ X_{c,0} \end{bmatrix} \sim TN\left(\boldsymbol{\mu}_{early}, \boldsymbol{\Sigma}_{early}\right).
$$

The truncation is such that $X_{c,0} \in [X_{0,low}, X_{0,up}]$ and $Y_{c,0} \in [Y_{0,low}, Y_{0,up}]$ where the boundaries can vary with c but this dependency is suppressed in the notation. The control parameters $\boldsymbol{\mu}_{early}$ and $\boldsymbol{\Sigma}_{early}$ are shared across all c.

Priors were chosen to be conjugate and diffuse for most parameters. An informative prior was used for the autoregressive parameter ρ in the imputation model for $\mathbf{Y}|\mathbf{X}$ to reflect the prior belief that $\mathbf{Y}$ is generally increasing over time for the motivating enrollment data, but should generally be specified based on the data being imputed. The values for all hyperparameters should ideally be determined using prior expert knowledge. However, in missing data problems where a general-purpose imputation method is used, the imputer generally does not possess such knowledge for the choice of hyperparameters. In the absence of expert information, we instead propose an algorithm for specifying diffuse priors dictated by data-based control parameters, details of which are in the Appendix. The prior distributions were chosen to be sufficiently diffuse such that the data-based control parameters do not over-

whelm the posterior inference following the philosophy of Edwards et al. (1963). We note that use of the data-based algorithm to specify prior distributions results in an approximate posterior distribution rather than a fully Bayesian posterior distribution.

A-splines

The nonlinear functions f and h in Equation 4.3 are estimated through spline regression using the complete cases in $(\mathbf{X}, \mathbf{Y})$. To estimate f, the model $Y_{c,t} = f(X_{c,t}) + \varepsilon^f_{c,t}$ is fit with Gaussian errors using a B-spline of degree 1. The residuals from the estimation of f are then used to estimate h by fitting the model $|Y_{c,t} - f(X_{c,t})| = h(X_{c,t}) + \varepsilon^h_{c,t}$ with Gaussian errors and using a B-spline of degree 1. After estimation, f is truncated to have range $[Y_{low}, Y_{up}]$ and h is truncated to have range $[\epsilon, \infty)$, where ϵ is a small positive value. The number and placement of knots for the B-splines used for f and h are selected using a method called adaptive splines, or A-splines (Goepp et al., 2018). A-splines automates the selection of knots using an iterative penalized likelihood approach and is implemented in the R package "aspline" (Goepp, 2022).

4.3.3 Estimation

Model estimation

The MINTS model is estimated using a Markov chain Monte Carlo (MCMC) algorithm with Gibbs sampling and Metropolis-Hastings steps in R. Multiple imputations are created in two phases. The parameters of the imputation model are first estimated in the estimation phase. In the imputation phase, additional iterations from the same MCMC algorithm are run and used to create multiply imputed data sets.

Estimation of the imputation model parameters occurs simultaneously with estimation of the missing values in a similar fashion to the data augmentation algorithm of Tanner and Wong (1987), with the MCMC algorithm resulting in samples from the joint posterior

distribution of the imputation model parameters and the missing values given the observed data. At each iteration of the MCMC algorithm, two steps are iterated until convergence is reached. First, values of the imputation model parameters are drawn from their posterior distributions given the observed data and the most recent estimates of the missing values. Second, estimates for the missing values are drawn from their posterior distributions given the observed data and the previously drawn imputation model parameters. Let $\boldsymbol{\theta}_X = (\boldsymbol{\gamma}, \sigma^2_X, \mathbf{X}_0, \mu_{drift}, \sigma^2_{drift})$ and $\boldsymbol{\theta}_Y = (\boldsymbol{\alpha}, \beta, \rho, \sigma^2_Y, \mathbf{Y}_0, \mu_0, \sigma^2_0)$ denote the parameters of the models for $\mathbf{X}$ and $\mathbf{Y}|\mathbf{X}$, respectively. The general approach of the MCMC algorithm proceeds as follows for iterations $i = 1, ..., n_{iter}$:

1. Draw $\boldsymbol{\theta}_X^{(i)}$ from $p(\boldsymbol{\theta}_X|\mathbf{X}_{obs}, \mathbf{X}_{mis}^{(i-1)})$

2. Let $j = 1, ..., J$ index the $X_{c,t}$ in $\mathbf{X}_{mis}$. For each j, draw $X_j^{(i)}$ from

$$p(X_j|\mathbf{X}_{obs}, X_1^{(i)}, X_2^{(i)}, ..., X_{j-1}^{(i)}, X_{j+1}^{(i-1)}, ..., X_J^{(i-1)}, \boldsymbol{\theta}_X^{(i)})$$

3. Draw $\boldsymbol{\theta}_Y^{(i)}$ from $p(\boldsymbol{\theta}_Y|\mathbf{Y}_{obs}, \mathbf{Y}_{mis}^{(i-1)}, \mathbf{X}_{obs}, \mathbf{X}_{mis}^{(i)})$

4. Let $k = 1, ..., K$ index the $Y_{c,t}$ in $\mathbf{Y}_{mis}$. For each k, draw $Y_k^{(i)}$ from

$$p(Y_k|\mathbf{Y}_{obs}, Y_1^{(i)}, Y_2^{(i)}, ..., Y_{k-1}^{(i)}, Y_{k+1}^{(i-1)}, ..., Y_K^{(i-1)}, \mathbf{X}_{obs}, \mathbf{X}_{mis}^{(i)}, \boldsymbol{\theta}_Y^{(i)})$$

The total number of iterations n_{iter} used in the estimation phase is determined based on convergence diagnostics such as inspection of trace plots and evaluation of the diagnostics of Raftery and Lewis (1996) and Gelman and Rubin (1992). Complete details of the MCMC algorithm can be found in the Appendix.

Imputation procedure

After the MCMC algorithm has converged for estimation of the imputation model parameters, the imputation phase begins. All iterations of the MCMC that were required for

convergence are treated as burn-in during the imputation phase. Imputed values for $\mathbf{X}_{mis}$ and $\mathbf{Y}_{mis}$ are created by continuing the MCMC algorithm with additional thinning steps. Thinning of the MCMC chains is required during the imputation phase to ensure that the imputed data sets are approximately independent draws from the posterior predictive distribution of the missing data given the observed data under the model described in Section 4.3.2 and priors described in Section 4.3.2. The amount of thinning is chosen so that the autocorrelation of the imputed values is approximately zero.

The number of iterations used in the imputation phase depends on the desired number of multiply imputed data sets, the number of chains of the MCMC, and the number of iterations used for thinning. To obtain M multiply imputed data sets from C chains with n_{thin} iterations between imputed values, the MCMC algorithm is run for an additional $\frac{M}{C} \times n_{thin}$ iterations for each chain.

4.4 Simulation Study

We conducted a simulation study to evaluate how well the MINTS method performs for estimation of analysis models that are uncongenial to the imputation model. We refer to this validation exercise as "analysis model validation."

Analysis model validation was conducted for a simulated data set where a nonlinear relationship was simulated between variables. We considered nine experiments corresponding to three rates of simulated missingness and three missing data mechanisms. Each experiment was replicated $N_{rep} = 1000$ times and the average performance across replications for MINTS was compared with the performance of existing multiple imputation methods for hierarchical time series data. Details of the simulation study using the nonlinear simulated data are presented in this section, while details of an analogous simulation study using linear simulated data are available in the Appendix.

4.4.1 Data generation

Variables **X** and **Y** were generated for 20 countries and 30 years for a total sample size of 600 country-years. **Y** is the variable of interest for substantive analyses, while **X** is an auxiliary variable that is only of interest for imputation of **Y**.

X was simulated independently for each country and is bounded in $[0, 100]$. For country c,

$$\begin{aligned} X_{c,1} &\sim U(X_{1,low}, X_{1,up}), \\ X_{c,t+1} &\sim TN_{[0,100]}(X_{c,t} + \gamma_c, \sigma^2) \text{ for } t = 1, ..., 29, \end{aligned}$$

where $TN_{[0,100]}$ refers to the truncated normal distribution with support $[0, 100]$ and γ_c is a country-specific drift term. For the nonlinear simulated data, we set $X_{1,low} = 0$, $X_{1,up} = 25$, $\sigma^2 = 1$, and $\gamma_c \sim U(1, 3)$.

Y was simulated to have a nonlinear relationship with **X** and is bounded as $Y_{c,t} \in [0, min(X_{c,t}, 60)]$. For country c,

$$\begin{aligned} Y_{c,t} &\sim TN_{[0,\min(X_{c,t},60)]}\left(\alpha_c + \frac{40}{\left(1 + \exp\left(-\frac{X_{c,t}-60}{8}\right)\right)} + 3\log(X_{c,t}), 1^2\right), \\ \alpha_c &\sim U(0, 5), \end{aligned}$$

where $TN_{[0,\min(X_{c,t},60)]}$ refers to the truncated normal distribution with support $[0, \min(X_{c,t}, 60)]$. **X** and **Y** were constructed to have a generally monotonically increasing relationship similar to the relationship observed for the enrollment data.

4.4.2 Analysis Model

We focused on the setting where the analysis model is uncongenial to the imputation model. The variable **Z** is treated as the outcome variable and was simulated to have a linear

relationship with $\mathbf{Y}$. For each country c and year t,

$$Z_{c,t} \sim N(\eta_c + 2Y_{c,t}, 10^2),$$
$$\eta_c \sim U(0, 15).$$

We considered the linear regression of $\mathbf{Z}$ on $\mathbf{Y}$. The parameter of interest is ω_1, the coefficient on $\mathbf{Y}$ in the regression

$$Z_{c,t} = \omega_0 + \omega_1 Y_{c,t} + \varepsilon^{\omega}_{c,t},$$
$$\varepsilon^{\omega}_{c,t} \sim N(0, \sigma^2_{\varepsilon_\omega}).$$

Additional analysis model validation results for a random intercept model are available in the Appendix.

4.4.3 Analysis model validation procedure

Data was simulated as missing following the three missing data mechanisms of Rubin (1976): Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not at Random (MNAR). For variable $\mathbf{X}$, MCAR occurs when $P(\mathbf{R}^X|\mathbf{X}) = P(\mathbf{R}^X)$. Data was simulated as missing under MCAR by assuming that each observation has the same probability of being missing. MAR occurs when $P(\mathbf{R}^X|\mathbf{X}) = P(\mathbf{R}^X|\mathbf{X}_{obs})$. Data was simulated as missing under MAR by assuming that observations in earlier years are more likely to be missing. Finally, MNAR occurs when the probability of being missing depends on both the missing and the observed data and $P(\mathbf{R}^X|\mathbf{X})$ cannot be simplified further. Data was simulated as missing under MNAR by assuming that the probability of being missing depends on the observed values. All methods compared in the simulation study assume that the missing data mechanism is at least MAR, so the simulations using MNAR act as a sensitivity analysis. Details of the MAR and MNAR implementations can be found in the Appendix.

For each missing data mechanism, data was simulated as missing at the 10%, 40%, and 80% rates. Each combination of missing data mechanism and rate was implemented simultaneously for $\mathbf{X}$ and $\mathbf{Y}$ and defines an experiment. For example, the MCAR 10% experiment corresponds to the setting where 10% of $\mathbf{X}$ was simulated as missing under MCAR and 10% of $\mathbf{Y}$ was simulated as missing under MCAR. We note that while 80% is a high rate of missingness, it is similar to the observed rate of missingness for NER in the enrollment data set.

For each experiment, the analysis model validation procedure is

1. For replication $r = 1, ..., N_{rep}$,

   (a) Simulate missing values according to the experiment's missing data mechanism and rate to separate the data into "observed" and "missing" data

   (b) Run the multiple imputation procedure using the observed data to create $M = 40$ completed data sets. Completed data sets consist of the observed data and the imputed values for the missing data.

   (c) Estimate quantities of interest Q using each of the M completed data sets

   (d) Pool the estimates of Q and $SE(Q)$ across the M completed data sets using combining rules from Rubin (1987) to obtain the pooled estimates $\bar{Q}_r$ and $SE(\bar{Q}_r)$

2. Calculate the true value of Q using the full data set

3. Calculate evaluation metrics for the pooled estimates $\bar{Q}_r$ averaged across replications

The number of imputations $M = 40$ was chosen to balance between having a large enough number of imputations to guarantee minimal contribution of simulation error to the variability of estimands and having a small enough number of imputations to be computationally feasible.

Pooled estimates for each scalar quantity of interest Q are created using combining rules from Rubin (1987) in Step 1(d) of the validation procedure. Let $\hat{Q}_m$ denote the point estimate of Q from imputation m and let $\hat{U}_m$ denote its associated variance. The pooled point estimate of Q across all M imputations is $\bar{Q}_M = \frac{1}{M}\sum_{m=1}^{M}\hat{Q}_m$. The pooled variance estimate of $\bar{Q}_M$ is $T_M = \bar{U}_M + \left(1+\frac{1}{M}\right)B_M$, where $\bar{U}_M = \frac{1}{M}\sum_{m=1}^{M}\hat{U}_m$ is the average within-imputation variance and $B_M = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{Q}_m - \bar{Q}_M)^T(\hat{Q}_m - \bar{Q}_M)$ is the between-imputation variance. The 95% confidence interval for the pooled estimate $\bar{Q}_M$ is constructed using $\bar{Q}_M \pm t^*_\nu\sqrt{T_M}$, where t^*_ν is the critical value for 95% confidence from the t distribution with ν degrees of freedom. The degrees of freedom ν for finite number of imputations M is $\nu = (M-1)(1+\frac{1}{r_M})^2$, where r_M is the relative increase in variance due to nonresponse given by $r_M = \left(1+\frac{1}{M}\right)\frac{B_M}{\bar{U}_M}$.

The performance of the point estimates $\bar{Q}_M$ was evaluated using the mean absolute error (MAE), calculated as $\frac{1}{N_{rep}}\sum_r |\bar{Q}_M - Q|$ where the sum is taken over all replications within each experiment. The performance of the 95% confidence intervals for the pooled estimates $\bar{Q}_M$ was evaluated by using the mean coverage across all replications within each experiment, calculated as the proportion of intervals that contained the true value. We also evaluated the mean fraction of Fisher information about Q that is missing due to nonresponse, which we abbreviate as FMI for Fraction of Missing Information. In each replication of each experiment, FMI is estimated as

$$\text{FMI} = \frac{r_M + \frac{2}{\nu+3}}{r_M + 1}.$$

The mean FMI across all replications within each experiment enables assessment of the amount of information about Q that is lost due to the presence of missing data (Savalei and Rhemtulla, 2012; Schafer, 1997a). We note that with $M = 40$ imputations, the estimates of FMI may be noisy, so in this case FMI should only be interpreted as an exploratory diagnostic (Bodner, 2008; Enders, 2010).

4.4.4 Model implementation

For each replication of each experiment, we created 40 imputations using the MINTS method by running 10 chains of the MCMC algorithm. The bounds of the model for $\mathbf{X}$ were set as $X_{low} = 0$ and $X_{up} = 100$ and the bounds of the model for $\mathbf{Y}|\mathbf{X}$ were set as $Y_{low} = 0$ and $Y_{up} = \min(X_{c,t}, 60)$. During the estimation phase, the MCMC was run for enough iterations to ensure convergence of all imputation model parameters. The number of iterations differed across experiments, but ranged from 10000 to 25000 iterations per chain. During the imputation phase, an additional 4000 iterations was run for each chain and four iterations from each chain were selected as the imputed values. The iterations selected as the final imputed values were chosen to be 1000 iterations apart to ensure autocorrelation was close to zero following the procedure described in Section 4.3.3.

We compared the MINTS method to six models based on existing methods for multiple imputation. Three of these models are based on the MICE methodology as implemented in the R package "mice" (van Buuren and Groothuis-Oudshoorn, 2011).[2] MICE uses the FCS algorithm, which allows for the specification of separate univariate conditional models for each variable with missing data. The univariate conditional models include all available variables as predictors, for example, the model for $\mathbf{X}$ includes $\mathbf{Y}$, year, and country as predictors. We first considered the default imputation method for continuous data in the mice package, which is predictive mean matching. We refer to this method as the MICE PMM method. Unlike the other methods considered, MICE PMM does not explicitly account for the hierarchical structure of the data but is included as a baseline for comparison.

We evaluated two models within the MICE framework that account for the hierarchical structure of the data using the method of Schafer (1997b) and Schafer and Yucel (2002), referred to as the pan method following its implementation in the R package "pan" (Zhao and Schafer, 2023). The function *mice.impute.2l.pan* within the mice package uses a Gibbs

[2]mice version 3.14.0 used

sampler to estimate the conditional linear mixed effects model with homogeneous within-group variances for each variable with missing data given the other variables. We evaluated a model that includes a country-specific intercept and fixed effects for all covariates in the imputation model, referred to as the pan Fixed Effects method. For example, imputed values for $X_{c,t}|Y_{c,t}$ are modeled with a country-specific intercept, a fixed effect of year, a fixed effect of $Y_{c,t}$, and a homogeneous normally distributed error term. We also considered the pan Random Effects method, which adds random effects of all covariates to the model used in the pan Fixed Effects method.

We evaluated three models using the Amelia II methodology as implemented in the R package "Amelia" (Honaker and King, 2010).[3] For complete data $(\mathbf{X}, \mathbf{Y})$, Amelia assumes the joint distribution is multivariate normal. The parameters of the joint distribution are estimated using a combination of bootstrapping and an EM algorithm. Amelia is designed specifically for imputation of hierarchical time series data, which Honaker and King refer to as time series cross-sectional data, through modeling features such as smooth trends over time and allowing for country-specific effects. For our comparisons, we evaluated three implementations of Amelia that were chosen to use the simplest form of the time-series-cross-sectional modeling features in Amelia: a time-series (TS) model, a cross-sectional (CS) model, and a time-series-cross-sectional (TSCS) model. All three of the imputation models are constructed by adding terms to the default multivariate normal joint model for $(\mathbf{X}, \mathbf{Y})$. In the Amelia TS method, a linear effect of time is added. In the Amelia CS method, a country-specific intercept term is added. In the Amelia TSCS method, a country-specific intercept term and a country-specific linear effect of time are added.

Imputations were created using the default settings in mice and Amelia, with the exception of setting the number of imputations as $M = 40$, specifying the form of the imputation models as described above, and specifying bounds of $\mathbf{X} \in [0, 100]$ and $\mathbf{Y} \in [0, 60]$. Scalar

[3]Amelia version 1.80 used

bounds were used for $\mathbf{Y}$ as the mice and Amelia packages do not allow for variable bounds.

4.4.5 Analysis model validation results

For the linear regression analysis model validation, we evaluated how well each multiple imputation method performs for estimation of $Q = \omega_1$, the regression coefficient on $\mathbf{Y}$ in the linear regression of $\mathbf{Z}$ on $\mathbf{Y}$. Table 4.1 summarizes the results of this validation for the nonlinear simulated data. Overall, MINTS results in the best balance between MAE, coverage, and FMI across the nine experiments. MINTS has the smallest MAE in all experiments except MAR 10%, where MINTS has the second smallest MAE behind MICE PMM. At the 10% and 40% rates, MINTS has reasonably close to nominal coverage for 95% intervals. While MINTS has the closest to nominal coverage out of the methods compared in the experiments at the 80% rate, coverage is below nominal in all three experiments. MINTS also has the smallest FMI at the 10% and 40% rates. As expected, FMI is much larger for all methods at the 80% rate, with FMI $> 50\%$ for several methods. This is higher than is typical for the setting of sample survey data that multiple imputation was originally designed for, where FMI is usually $\leq 30\%$ (Rubin 2007). Given the high FMI, it is thus unsurprising how poorly all multiple imputation methods perform in the experiments at the 80% rate.

None of the existing methods perform consistently well across experiments for estimation of ω_1. While MICE PMM can outperform the methods explicitly designed for hierarchical time series data at the lower rates of missingness, the performance of MICE PMM suffers greatly at the highest rate of missingness. The pan and Amelia methods perform similarly to one another at the 10% rate, but all have larger MAE than MICE PMM. Performance of the pan and Amelia methods generally is best under MCAR, with pan Random Effects and Amelia TSCS performing the best of the group in terms of MAE and coverage.

We note that the existing methods included in the comparisons assume a linear relationship between variables in the imputation model. We also conducted a simulation study

Table 4.1: Summary of analysis model validation for nonlinear simulated data for $Q = \omega_1$, the regression coefficient on Y in the linear regression of Z on Y. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 1000 replications of each experiment. The true value of Q is 2.060.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	1.16	100.0	6.4	**0.53**	100.0	3.4	0.81	100.0	4.4
	pan Fixed	1.78	100.0	6.0	6.34	93.3	7.1	3.43	100.0	5.5
	pan Random	1.18	100.0	4.1	4.87	99.9	4.9	2.35	100.0	3.5
	Amelia TS	1.87	100.0	7.1	5.64	99.8	7.6	3.37	100.0	6.0
	Amelia CS	2.30	100.0	8.1	6.83	**95.3**	9.1	3.97	100.0	6.8
	Amelia TSCS	1.28	100.0	4.7	5.13	99.9	5.2	2.45	100.0	3.6
	MINTS	**0.29**	100.0	**0.5**	0.56	100.0	**0.5**	**0.29**	100.0	**0.4**
40%	MICE PMM	13.76	30.5	37.8	4.71	100.0	23.0	10.25	56.7	32.3
	pan Fixed	10.35	50.7	24.9	27.75	0.0	22.7	17.86	0.0	22.2
	pan Random	4.95	97.5	16.6	20.05	0.0	15.9	9.95	16.9	14.7
	Amelia TS	11.77	51.0	23.8	23.35	0.0	21.3	16.87	0.0	21.7
	Amelia CS	19.61	1.0	29.0	27.31	0.0	22.2	22.57	0.0	24.3
	Amelia TSCS	5.19	**96.9**	22.3	21.23	0.0	23.8	9.94	21.3	17.4
	MINTS	**1.21**	100.0	**3.9**	**3.02**	**99.9**	**4.4**	**1.49**	**100.0**	**2.8**
80%	MICE PMM	60.62	0.1	60.2	29.59	4.8	56.9	59.98	0.0	54.4
	pan Fixed	35.58	2.9	49.0	56.03	0.0	35.8	49.51	0.0	43.6
	pan Random	15.30	43.2	54.1	43.08	0.0	51.7	26.93	0.1	52.7
	Amelia TS	41.08	4.8	50.4	54.29	0.0	**32.9**	46.55	0.0	**33.1**
	Amelia CS	99.19	0.0	66.6	79.16	0.0	70.0	77.02	0.0	54.4
	Amelia TSCS	22.73	48.2	75.8	57.15	0.0	78.9	35.91	4.9	76.6
	MINTS	**7.62**	**81.0**	**38.7**	**13.14**	**47.5**	47.0	**9.71**	**62.6**	35.6

for data simulated to follow a linear relationship where we considered two uncongenial and two congenial analysis models. Although our primary focus is on the uncongenial setting, the analysis model validation using congenial analysis models and linear simulated data allows us to compare the performance of MINTS with the existing imputation methods in

an "ideal" setting for multiple imputation. We found that the existing imputation methods perform better in the analysis model validation for the linear simulated data compared to the nonlinear simulated data. However, for estimation of uncongenial analysis models we found MINTS still outperforms the existing methods in the linear setting at the 10% and 40% rates of simulated missingness. All imputation methods were found to perform well for estimation of congenial analysis models using linear simulated data, with no method consistently standing out from the others across experiments. Details of the simulation study using linear simulated data can be found in the Appendix.

4.5 Application to Enrollment Data

We further evaluated the performance of MINTS by revisiting the secondary school enrollment rate data that was described in Section 2. We conducted two validation exercises using the enrollment data by simulating additional country-years as missing. In the first validation exercise, we evaluated how well MINTS performs for estimation of parameters of interest Q for uncongenial analysis models. This is analogous to the analysis model validation that was conducted for the simulation study.

In the second validation exercise, we evaluated the predictive performance of MINTS for predicting left-out observations of NER. Out-of-sample validation for prediction of individual missing values is less frequently conducted for multiple imputation methods compared to analysis model validation, but has been considered by Gelman et al. (1998); Honaker and King (2010); Nguyen et al. (2017); among others. Although the primary goal of multiple imputation is to create valid estimates of parameters of interest in the presence of missing data rather than recovering the missing values (Rubin, 1996), the prediction of individual missing values can still be of great interest in practice. A multiple imputation method that can perform well for prediction of individual missing values and for creating multiply imputed estimates of quantities of interest is thus of increased utility. We refer to this second

validation exercise as "out-of-sample validation."

Finally, we applied the MINTS method to the full enrollment data set without simulating any missingness and created 40 multiple imputations for the missing country-years for NER and GER that are present in the original data set. We briefly present the multiple imputation results and make available the 40 completed data sets in the Appendix.

4.5.1 Validation procedure

For both validation exercises using the enrollment data, we considered eight experimental conditions. Parameters varied in both validation exercises were the rate of simulated missingness (10%, 40%, 80%) and the missing data mechanism (MCAR, MAR, and MNAR), where the same missing data mechanisms used in the simulation study and described in the Appendix were also used for the enrollment data. We did not consider the MNAR 80% experiment for the enrollment data as this resulted in the majority of countries having no observations for NER or GER.

As we simulated additional missing values in a data set that began with missing values, the overall rate of missingness for each experiment is larger than the rate of simulated missingness. For example, the experiments with a 40% rate of simulated missingness correspond to an overall rate of missingness of 84.8% for NER. Similarly, the simulated missing data mechanisms used in the validation exercises describe the missing data mechanism only for observations that are simulated as missing. The true missing data mechanism for the observations that began as missing in the full enrollment data set is unknown.

For the analysis model validation, each experiment was replicated $N_{rep} = 100$ times following an analogous procedure as was described in Section 4.4.3 for the simulation study. A major difference in the analysis model validation procedure for the enrollment data is the need to distinguish between the country-years that start out as missing and the country-years that are simulated as missing, where only the country-years that are simulated as

missing are used for evaluation purposes. To facilitate this distinction, in each replication of each experiment we separated the enrollment data into "started-as-missing", "observed", and "simulated-as-missing" sets. The observed set is the training set for model estimation, while the simulated-as-missing set is the testing set for validation. Imputed values were still created for the started-as-missing set, but we are unable to evaluate the performance of these imputed values as the true value for the country-years that started as missing is unknown.

For the out-of-sample validation exercise, we considered one replication of each experiment. We used the same distinction between the country-years that started as missing and the country-years that were simulated as missing, where only the country-years that were simulated as missing are included in the testing set for validation. The out-of-sample validation compares performance metrics for prediction of missing values for NER averaged over all country-years in the testing set.

In both validation exercises, we compared the performance of MINTS with the MICE PMM, pan Fixed Effects, pan Random Effects, Amelia TS, Amelia CS, and Amelia TSCS methods that were described in Section 4.4.4.

4.5.2 Analysis model validation

Analysis model

We are interested in estimating the relationship between NER and the Total Fertility Rate (TFR), which is a period measure of the expected number of children a woman would bear in her lifetime if she were to experience the period-specific fertility rates at each age and if she lived through the reproductive age range of 15–49. There is a well-established negative association between education and fertility in the high-fertility setting (Hirschman, 1994). One mechanism through which education is posited to have a negative effect on fertility is through the educational enrollment of children, where increased enrollment increases the cost of raising children (Axinn and Barber, 2001). Children who are enrolled in school have

reduced capacity for work and may incur increased costs for caregivers through fees related to tuition, uniforms, and textbooks (Caldwell, 1982; Caldwell et al., 1985; Easterlin and Crimmins, 1985). To estimate this relationship, we use annual estimates of TFR from the United Nations World Population Prospects 2022 (United Nations, 2022). We restrict our analyses to the high-fertility context, defined here as the years where a country has TFR > 2.5 children per woman.

The analysis model is the linear regression of TFR on NER, where the quantity of interest $Q = \beta_1$ is the regression coefficient on NER in the regression

$$\begin{aligned} \text{TFR}_{c,t} =& \beta_0 + \beta_1 \text{NER}_{c,t} + \varepsilon^{\beta}_{c,t}, \\ \varepsilon^{\beta}_{c,t} \sim& N(0, \sigma^2_{\varepsilon_\beta}). \end{aligned}$$

For each replication of each experiment, the analysis model is estimated using only the country-years that were simulated as missing and where TFR > 2.5. We also conducted analysis model validation using a random intercept model for TFR on NER, results of which are available in the Appendix.

Model implementation

GER is the auxiliary variable $\mathbf{X}$ and NER is the variable of interest $\mathbf{Y}$. The bounds of the model for $\mathbf{X}$ were set as $X_{low} = 0$ and $X_{up} = \infty$, while the bounds of the model for $\mathbf{Y}|\mathbf{X}$ were set as $Y_{low} = 0$ and $Y_{up} = \min(X_{c,t}, 100)$. Imputations were created using MINTS for each replication of each experiment by running 10 chains of the MCMC algorithm until convergence was reached in the estimation phase. The total number of iterations differed across experiments, with the experiments with larger rates of simulated missingness requiring a larger number of iterations to achieve convergence. After burn-in, the number of iterations per chain ranged from 5000 to 35000. In the imputation phase, an additional 4000 iterations was run for each chain and four iterations were selected from each chain to produce a total of $M = 40$ completed data sets following the procedure described in Section 4.3.3.

Imputations from the MICE and Amelia methods were created using the mice and Amelia packages in R. We set the number of imputations to $M = 40$, specified the form of the imputation models as described in Section 4.4.4, and specified bounds of $\mathbf{X} \in [0, \infty)$ and $\mathbf{Y} \in [0, 100]$, but otherwise used the default software settings.

Results

Table 4.2 summarizes the results of the analysis model validation for estimation of $Q = \beta_1$, the regression coefficient on NER in the linear regression of TFR on NER. MINTS results in the best overall performance, with the smallest MAE in all experiments except MAR 80%, where MINTS has the second smallest MAE after pan Fixed Effects. MINTS also has the smallest FMI in all experiments and close to nominal coverage in all but the MNAR 40% experiment. Out of the previously existing methods, pan Fixed Effects performs the best overall, with the smallest MAE in the MAR 80% experiment and the second smallest MAE in all other experiments. Pan Fixed Effects also results in good coverage for the MCAR and MAR experiments and has the closest to nominal coverage for the MNAR 40% experiment.

4.5.3 Out-of-sample validation

Next, we evaluated the predictive performance of each multiple imputation method for imputing the individual values of NER that were simulated as missing. In each experiment, the performance of point estimates was evaluated using the mean absolute error, where the mean was taken over all country-years in the testing set. The performance of the predictive intervals was evaluated by checking the average interval widths and coverage of the 95% intervals with respect to the observations in the testing set, where coverage is calculated as the proportion of the intervals that contained the true left-out values of NER. We also evaluated how well each imputation method balances the trade-off between interval width and coverage for the 95% intervals using the average negatively-oriented interval score from

Table 4.2: Summary of analysis model validation for enrollment data for $Q = \beta_1$, the regression coefficient on NER in the linear regression of TFR on NER. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 100 replications of each experiment. The value of Q estimated using the observed country-years from the full enrollment data set is -0.043.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	0.61	90	20.9	0.86	71	26.3	7.29	0	28.2
	pan Fixed	0.08	100	7.2	0.11	100	6.8	2.69	0	29.9
	pan Random	0.07	100	5.3	0.12	100	7.8	4.08	2	88.4
	Amelia TS	0.29	100	24.2	0.51	**99**	26.7	7.16	0	30.6
	Amelia CS	0.21	100	11.4	0.40	**99**	11.9	6.03	0	58.9
	Amelia TSCS	0.09	100	8.3	0.16	100	12.9	6.42	0	69.2
	MINTS	**0.04**	100	**2.2**	**0.08**	100	**2.4**	**0.42**	**100**	**10.8**
40%	MICE PMM	1.82	0	35.0	1.94	0	38.1	5.22	0	43.0
	pan Fixed	0.05	100	13.4	0.08	**100**	12.8	1.02	**76**	74.9
	pan Random	0.13	**99**	21.0	0.20	**100**	31.4	4.73	0	93.7
	Amelia TS	1.34	0	32.4	1.48	0	35.5	4.77	0	40.5
	Amelia CS	0.69	3	24.6	0.86	0	24.6	6.04	0	69.1
	Amelia TSCS	0.22	**99**	22.4	0.43	74	32.4	6.42	0	71.1
	MINTS	**0.04**	100	**4.6**	**0.07**	**100**	**5.4**	**0.63**	13	**31.9**
80%	MICE PMM	2.96	0	52.1	3.08	0	54.2			
	pan Fixed	0.07	**100**	41.4	**0.09**	**100**	45.9			
	pan Random	1.20	0	69.3	1.29	0	76.6			
	Amelia TS	2.55	0	44.6	2.64	0	47.0			
	Amelia CS	1.45	0	57.8	1.52	0	57.3			
	Amelia TSCS	1.01	2	62.4	1.57	0	64.3			
	MINTS	**0.06**	**100**	**21.8**	0.11	**100**	**25.5**			

Gneiting and Raftery (2007). For a 95% prediction interval, the interval score is defined as

$$IS = \frac{1}{n}\sum_{x}\left[(u-l) + \frac{2}{0.05}1\{x < l\} + \frac{2}{0.05}1\{x > u\}\right],$$

where (l, u) is the prediction interval, n is the number of observations in the testing set, and the sum is over all true values x for observations in the testing set.

The out-of-sample validation evaluates the posterior predictive distribution of the missing values given the observed values. Samples from the posterior predictive distribution of the missing values from the MINTS method were obtained using a slight modification of the imputation procedure, details of which can be found in the Appendix. Details of how samples from the posterior predictive distributions were obtained for the MICE and Amelia methods can also be found in the Appendix. Medians from the estimated posterior predictive distributions were used as point estimates for the imputed values and the 0.025 and 0.975 quantiles of the estimated posterior predictive distributions were used as 95% interval estimates.

Results

The results of the out-of-sample validation exercise are summarized in Table 4.3. MINTS results in the smallest MAE, narrowest interval width, and smallest interval score in all experiments. MINTS has close to nominal coverage in all experiments except for MNAR 40%, where coverage is below nominal. MINTS generally also produces narrower interval widths compared to the existing methods; these intervals appear to be sufficiently wide under MCAR and MAR, but may be too narrow under MNAR.

MICE PMM performs the worst overall, with comparatively large MAE and undercoverage in all experiments. This is perhaps unsurprising, as MICE PMM is the only method that does not account for the hierarchical structure of the data. Amelia TS also suffers from similarly large MAE as MICE PMM, but manages to retain close to nominal coverage in experiments under MCAR and MAR thanks to its wider intervals. The previously existing methods that have the best performance are pan Random Effects and Amelia TSCS. These two methods perform similarly in terms of MAE within experiments, with Amelia TSCS having larger MAE overall. Amelia TSCS has narrower intervals than pan Random Effects in most experiments, where Amelia TSCS tends towards undercoverage while pan Random

Effects tends towards overcoverage. Despite performing well overall, both these methods lead to larger MAE and interval scores than MINTS.

Table 4.3: Summary of out-of-sample validation for enrollment data for the country-years where NER was simulated as missing. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, Width denotes the average width of 95% intervals, and IS denotes the interval score for 95% intervals. Results are averaged over all NER observations simulated as missing in each experiment.

Simulated Missingness Rate	Method	**MCAR**				**MAR**				**MNAR**			
		MAE	Cvg	Width	IS	MAE	Cvg	Width	IS	MAE	Cvg	Width	IS
10%	MICE PMM	5.79	89.6	22.7	38.1	7.06	86.7	22.9	60.4	12.66	74.1	30.9	104.7
	pan Fixed	5.76	96.8	29.2	31.8	5.96	95.7	31.6	39.5	15.21	82.0	42.4	76.6
	pan Random	2.02	98.6	19.4	20.5	2.33	98.6	20.3	21.0	3.28	98.9	30.1	37.0
	Amelia TS	7.29	96.8	34.4	35.6	7.13	**95.0**	34.5	40.2	18.03	76.3	44.4	110.8
	Amelia CS	3.50	**95.3**	19.8	27.7	4.29	91.7	19.2	31.7	8.71	87.4	29.4	125.9
	Amelia TSCS	2.18	95.7	13.4	16.6	2.33	93.9	13.3	17.3	6.07	90.6	30.9	133.6
	MINTS	**1.27**	96.8	**9.7**	**11.7**	**1.42**	**95.0**	**9.6**	**12.3**	**1.77**	**94.6**	**8.5**	**13.1**
40%	MICE PMM	11.03	88.3	39.0	56.0	12.67	87.8	37.8	73.3	29.73	41.8	38.3	493.8
	pan Fixed	10.08	97.7	47.1	51.0	9.43	97.8	46.7	50.3	31.06	28.5	36.5	616.2
	pan Random	3.14	99.7	30.5	31.0	3.39	99.5	31.1	32.9	17.11	**97.3**	60.3	70.9
	Amelia TS	12.45	**93.9**	49.1	57.6	11.25	**94.6**	46.6	52.9	32.54	28.7	37.2	637.9
	Amelia CS	6.28	91.7	28.8	48.3	6.68	87.1	25.2	67.5	28.83	36.7	31.8	705.9
	Amelia TSCS	3.28	92.1	15.8	31.5	3.89	90.5	16.3	52.2	27.28	41.3	24.1	818.4
	MINTS	**2.08**	**96.1**	**12.9**	**14.8**	**2.17**	**95.4**	**12.5**	**16.0**	**4.44**	88.1	**16.2**	**33.2**
80%	MICE PMM	18.36	89.8	61.5	87.9	20.14	88.7	62.0	92.8				
	pan Fixed	16.89	97.6	71.6	76.0	16.80	**95.4**	69.5	77.6				
	pan Random	5.49	99.9	60.5	60.6	6.08	100.0	58.6	58.6				
	Amelia TS	18.82	92.2	67.7	80.3	18.08	91.9	65.3	80.8				
	Amelia CS	10.32	88.7	40.5	90.9	10.49	77.0	31.5	153.3				
	Amelia TSCS	6.40	82.9	25.0	114.1	8.47	74.4	24.2	165.1				
	MINTS	**4.14**	**93.7**	**20.1**	**24.8**	**4.86**	95.5	**23.1**	**28.3**				

We illustrate the out-of-sample validation results for the MINTS method for selected experiments in Figures 4.3 and 4.4. Observations in the training set are shown as solid circles in black for GER and red for NER. Posterior medians for the imputed values of NER are shown as red open circles, while the red shaded regions represent the 95% posterior predictive intervals. Imputed values are plotted for both the country-years that started as

missing in the enrollment data set and the country-years that were simulated as missing for the testing set. The true values of NER for the country-years in the testing set are shown as solid blue diamonds. Figure 4.3 shows the out-of-sample validation results for Afghanistan, Belgium, Spain, and Nigeria for the MCAR 40% experiment. Overall, the posterior predictive distributions for imputed values of NER result in plausible time series trends for each country, with the majority of the observations of NER that were simulated as missing captured within the 95% intervals. Results for the same example countries are shown in Figure 4.4 for the MAR 80% experiment. The predictive intervals are much wider overall for the MAR 80% experiment compared to the MCAR 40% experiment. This is especially apparent for Spain, which has many observations of NER and GER in early years and thus had a high proportion of observed values that were simulated as missing in the MAR 80% experiment.

4.5.4 Full enrollment data results

We used MINTS to create 40 multiple imputations for the missing country-years in the original enrollment data set without simulating any additional country-years as missing. Figure 4.5 shows the multiple imputations for Afghanistan, Belgium, Spain, and Nigeria for NER as translucent red circles along with the observed values of NER and GER from the original data set in solid red and black circles, respectively. Analogous figures illustrating the multiple imputation results for NER for all countries can be found in the Appendix.

4.6 Discussion

We developed a multiple imputation method for hierarchical time series data that can accommodate a specific type of nonlinear relationship between variables and evaluated the performance of the proposed method using a simulation study and an application to a data set on secondary school enrollment rates. Through comparisons with existing methods for

multiple imputation of hierarchical time series data, we found that the proposed MINTS method can lead to substantial gains in performance when variables in the imputation model have a nonlinear relationship that can be well-approximated by a piecewise linear function.

In the simulation study, we found that MINTS performed better for estimation of parameters in uncongenial analysis models using nonlinear simulated data compared to existing models based on the MICE and Amelia methodologies. MINTS generally resulted in the smallest MAE and FMI across experiments, while retaining close to nominal coverage in all but the experiments at the highest rate of simulated missingness. The simulation study results suggest that when there is a nonlinear relationship between variables, MINTS is a preferable method over the MICE and Amelia methods. We also found MINTS still performed well when the variables in the imputation model have a linear relationship, but the difference in performance between MINTS and the existing methods was less pronounced.

For the application to the enrollment data, we evaluated how well the MINTS method performs for estimation of parameters in uncongenial analysis models and for prediction of individual missing values. Based on the analysis model validation exercises, we found that MINTS resulted in improved performance for estimation of the regression coefficient in the linear regression of TFR on NER compared to existing models based on the MICE and Amelia methodologies, with the smallest or second smallest MAE in all experiments and close to nominal coverage in all but one experiment. MINTS also had good performance for estimation of the random intercepts regression of TFR on NER at the 10% rate of simulated missingness. However at higher rates of missingness we found that the pan Random Effects method performed better for estimation of the fixed effect coefficient. At all rates of missingness, MINTS tended to have the smallest MAE for point estimation of the variance of the random intercepts. This suggests that for random intercept models, MINTS is likely to be a better choice than the existing methods if a components of variance analysis is of interest, but the pan Random Effects model could be a better choice if only the fixed effect

coefficient is of interest. In the out-of-sample validation exercise for the enrollment data, we evaluated how well each imputation method predicted individual missing values for NER. We found that MINTS resulted in substantial improvements in performance compared to the existing methods, with smaller MAE, narrower intervals, and smaller interval scores across experiments. Despite the narrower intervals, MINTS still had close to nominal coverage in all experiments except MNAR 40%. Overall, our results suggest that MINTS is able to balance good predictive performance for individual missing values with good performance for multiply-imputed estimates of parameters in substantive analysis models.

To facilitate easier comparison across validation exercises, we summarize the average MAE across experiments for each multiple imputation method within each validation exercise in Table 4.4. A version of Table 4.4 that includes all validation exercises conducted is available in the Appendix. We note that Table 4.4 should not be interpreted as inferential and is only of interest as an exploratory comparison tool to enable a simple comparison of one metric from the full evaluation results. MINTS consistently results in the smallest average MAE across experiments in each validation exercise. Out of the previously existing imputation methods, we found the models using the pan methodology tend to have the smallest average MAE. For the validation exercises using enrollment data, the pan Random Effects method tends to perform the best out of the previously existing methods, with the second smallest average MAE for the out-of-sample validation exercise and the second or third smallest average MAE for the analysis model validation estimation of β_1. The pan Fixed Effects method also performs well in terms of average MAE for analysis model validation with the enrollment data, but has much larger average MAE for the out-of-sample validation exercise. For the validation exercises using the nonlinear simulated data, the pan Random Effects method has the smallest average MAE out of the existing imputation methods for estimation of ω_1.

We note that all of the imputation methods compared assume that the missing data mechanism is ignorable. We conducted sensitivity analyses to evaluate how robust each

Table 4.4: Average mean absolute error (MAE) across all experiments for each multiple imputation method within each validation exercise. For the enrollment data, OOS denotes to the out-of-sample validation and β_1 is the parameter from the linear regression analysis model validation. For the nonlinear simulated data, ω_1 is the parameter from the linear regression analysis model validation. MAE for the β_1 and ω_1 columns are multiplied by 100 before reporting.

Method	Enrollment Data		Nonlinear Data
	OOS	β_1	ω_1
MICE PMM	14.68	2.97	20.16
pan Fixed	13.90	0.52	23.18
pan Random	5.36	1.48	14.30
Amelia TS	15.70	2.59	22.76
Amelia CS	9.89	2.15	37.55
Amelia TSCS	7.49	2.04	17.89
MINTS	**2.77**	**0.18**	**4.15**

imputation model is to one type of violation of the ignorability assumption through the MNAR experiments, where the probability of a country-year being simulated as missing was dependent on the observed value. In the simulation studies, we found that all imputation methods considered were relatively robust to the violation of ignorability when the rate of simulated missingness was low. However, substantial increases in bias and undercoverage were seen at the 40% and 80% rates. MINTS tended to be the most robust to these violations for the simulated data but still suffered from bias and undercoverage, especially at the 80% rate. For the enrollment data validation exercises, no imputation method performed consistently well in the MNAR experiments. MINTS and pan Fixed Effects were the most robust to the violation of ignorability for the analysis model validation exercises, but both methods resulted in far below nominal coverage in more than one experiment. In the out-of-sample validation exercise, MINTS and pan Random Effects fared the best in the MNAR experiments. MINTS had the most robust performance in terms of point estimates, while pan Random Effects had the most robust coverage for interval estimates.

In this chapter, we focused on the setting where the analysis model is uncongenial to the imputation model. This is a common occurrence for social science data, particularly when the imputation phase is conducted independently from the analysis phase. However, if the analysis of interest is known during the imputation phase, congeniality between analysis and imputation models is a worthwhile goal. Several methods have been proposed for multiple imputation of hierarchical data under congeniality where the imputation model incorporates features of the substantive analysis model, such as Enders et al. (2020); Goldstein et al. (2014); Grund et al. (2021); Lüdtke et al. (2020); among others. Of particular note, Lüdtke et al. (2020) and Grund et al. (2021) propose a method called "mdmb" that uses a sequential decomposition of the joint model. The mdmb method ensures congeniality between analysis and imputation models by incorporating a term representing the substantive analysis model into the sequential decomposition following the substantive-model-compatible philosophy of Bartlett et al. (2015).

MINTS could be extended to congeniality using the same strategy as mdmb by adding a term representing the substantive analysis model into the decomposition of the joint model as

$$p(\mathbf{Z}, \mathbf{Y}, \mathbf{X}|\boldsymbol{\theta}_Z, \boldsymbol{\theta}) = p(\mathbf{Z}|\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}_Z)p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}_Y)p(\mathbf{X}|\boldsymbol{\theta}_X),$$

where $\mathbf{Z}$ is the outcome variable for the analysis of interest with parameters $\boldsymbol{\theta}_Z$. When the exact substantive analysis model of interest is included in the decomposition, MINTS could be adapted to a fully Bayesian approach and parameter estimates for the analysis model could be obtained directly from the estimation phase of the MCMC.

4.7 Conclusion

We have proposed the MINTS method for multiple imputation of hierarchical nonlinear time series data in the bivariate setting where one auxiliary variable is used to impute the variable of interest and the variable of interest has a larger amount of missing data than the

auxiliary variable. This setting is motivated by a school enrollment rate data set from the World Bank that includes two measures of enrollment. The Net Enrollment Rate (NER) is the variable of interest for an analysis of the relationship between the Total Fertility Rate and NER, but suffers from a large amount of missing data due to the difficulty of measuring NER. A second measure of enrollment, the Gross Enrollment Ratio (GER), is easier to measure and has a smaller amount of missing data. The MINTS method leverages the strong piecewise linear relationship between NER and GER to impute missing values in both NER and GER using a combination of A-splines, country-specific intercepts, and time series methods using a sequential decomposition of the joint model.

We compared MINTS with existing methods for multiple imputation of hierarchical time series data through a simulation study and an application to the enrollment data set. We considered three models within the MICE framework of van Buuren and Groothuis-Oudshoorn (2011) and three models within the Amelia II framework of Honaker and King (2010). We found that MINTS resulted in better performance overall for estimation of parameters in uncongenial linear regression and random intercept models compared to the imputation models based on the MICE and Amelia methodologies in both a simulation study using nonlinear simulated data and the application to school enrollment data, though models based on the pan method of Schafer and Yucel (2002) were strong alternatives. We also conducted an out-of-sample validation exercise for prediction of individual missing values using the enrollment data and found that MINTS resulted in better predictive performance compared to the existing methods. Based on these validation exercises, we believe MINTS has a promising capability to improve the quality of multiply imputed data sets for hierarchical nonlinear time series data.

One limitation of the MINTS method is the use of spline estimation to model the nonlinear relationship between variables in the imputation model. The A-splines method of Goepp et al. (2018) helps to automate the estimation procedure, but as with all spline estimation

methodologies there is the risk of overfitting. The MINTS algorithm uses linear splines and modifies the number of starting knots used in the A-spline algorithm when the sample size is small to reduce the risk of overfitting, but the spline estimation settings are likely to require manual tuning for applications of MINTS to other data sets. We note that other curve-fitting methods, such as LOESS, could be used instead.

MINTS also has several practical limitations compared to MICE and Amelia. Currently, MINTS is restricted to the bivariate setting with continuous variables. In principle, MINTS could be extended to the multivariate setting by adding additional univariate conditional terms to the sequential decomposition of the joint distribution, where careful consideration is needed for the ordering of the added conditional distributions. The imputation models for the added variables could be assumed to be linear or could incorporate a separate spline term for each marginal relationship in the conditional imputation model. MINTS could also be extended to accommodate categorical variables through the use of generalized regression models, for example using similar methodology as Lee and Mitra (2016). Another practical limitation of MINTS is computation time, where the estimation of the MINTS model is much more computationally intense than estimation of the MICE models and the two simpler Amelia models. Computation time could be improved by coding the MCMC algorithm in a more efficient language than R. Extending the MINTS methodology to the multivariate and mixed data type setting and improving the computational efficiency of the MINTS sampling algorithm is of interest for future work.

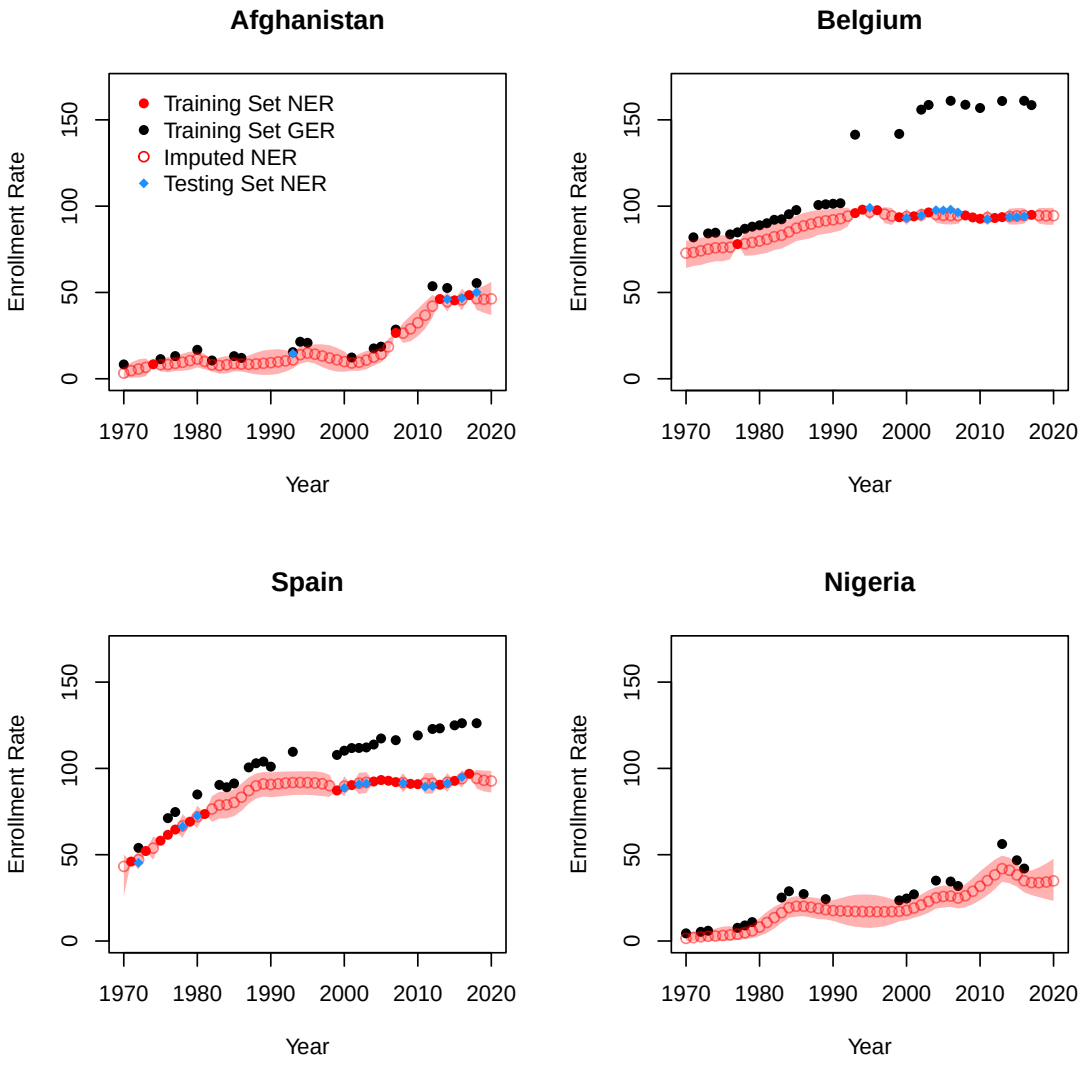


Figure 4.3: MCAR 40% experiment results for selected countries from the out-of-sample validation for enrollment data. Solid black and red circles indicate values in the training set for GER and NER, respectively. Solid blue diamonds indicate the true values in the testing set for NER. Open red circles indicate the median imputed values of NER and the red shaded regions indicate the 95% posterior quantiles for imputed values of NER, where values are imputed for country-years in the testing set and country-years that started as missing in the enrollment data set.

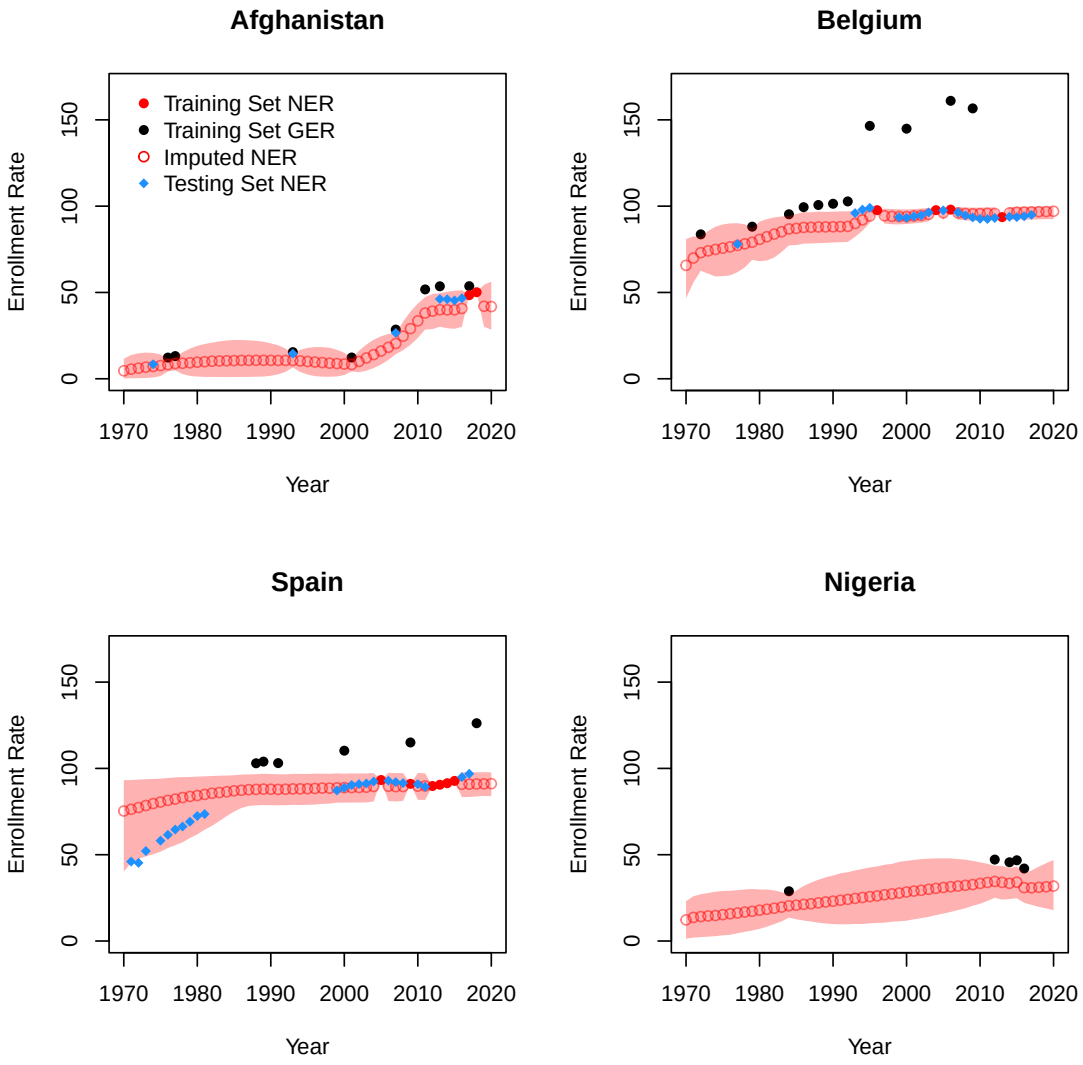


Figure 4.4: MAR 80% experiment results for selected countries from the out-of-sample validation for enrollment data. Solid black and red circles indicate values in the training set for GER and NER, respectively. Solid blue diamonds indicate the true values in the testing set for NER. Open red circles indicate the median imputed values of NER and the red shaded regions indicate the 95% posterior quantiles for imputed values of NER, where values are imputed for country-years in the testing set and country-years that started as missing in the enrollment data set.

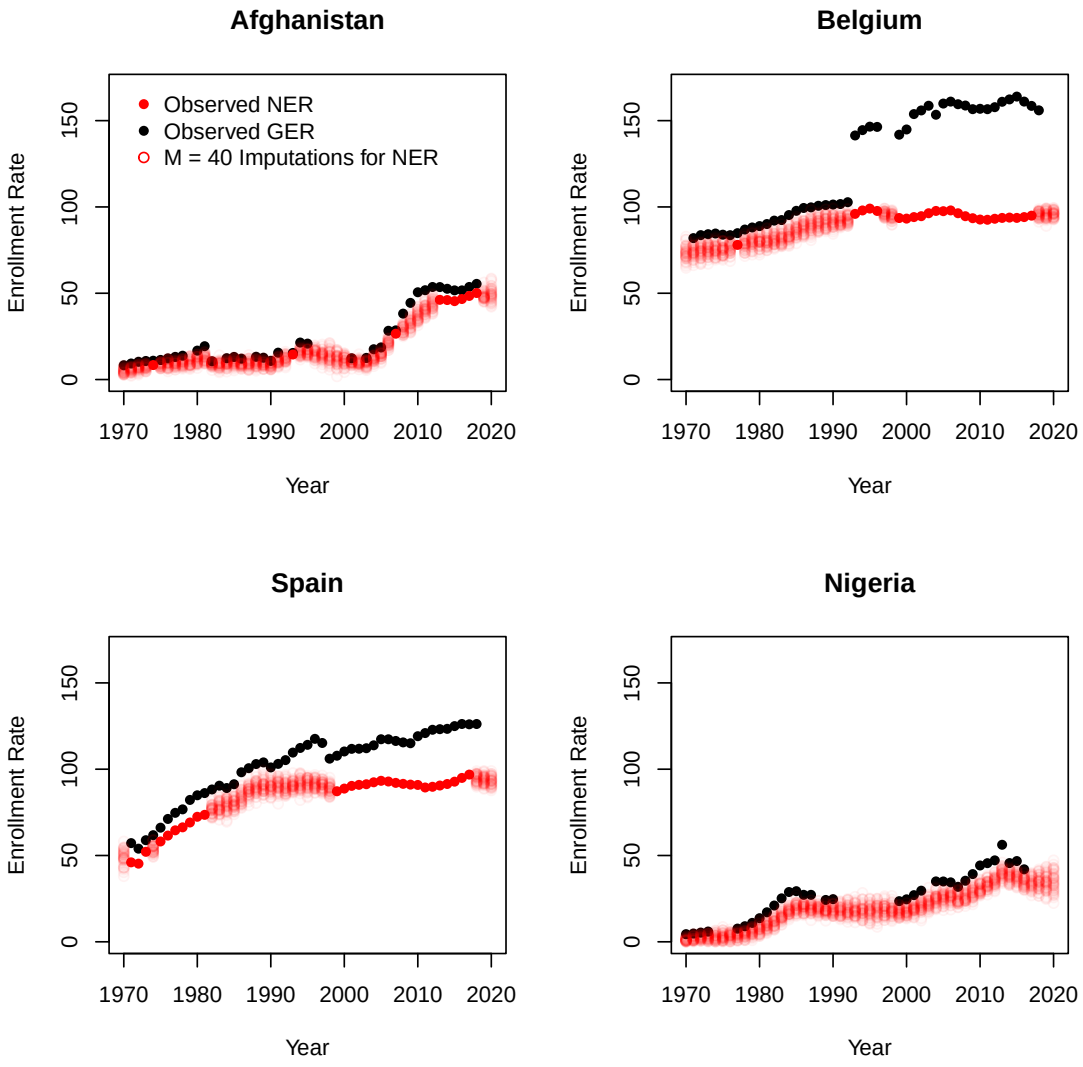


Figure 4.5: Results of $M = 40$ imputations for NER for selected countries. Solid black and red circles indicate observed values of GER and NER, respectively, from the enrollment data set. Translucent open red circles indicate imputed values of NER, where a total of 40 imputations were created for each missing value.

Chapter 5
DISCUSSION AND FUTURE WORK

This chapter reviews the main contributions of the dissertation and discusses several ideas for possible future research. Additional discussion can be found within each chapter in Sections 2.5, 3.5, and 4.6.

5.1 Summary of Contributions

In this dissertation, we have developed statistical methods to address two applied problems in demographic research involving hierarchical time series data. The first line of work aims to answer the question "what would happen to future fertility and population size if policy interventions targeting education and family planning are implemented?" in a probabilistic way. In Chapter 2, we developed methods for estimation of the potential accelerating effect of education and family planning on fertility decline using a framework inspired by Granger causality that measures the effect of expansions in education and family planning on fertility decline beyond what we would already expect the decline to look like based on past trends in fertility. We identified the mechanisms by which education and family planning can accelerate fertility decline and estimated their direct and indirect effects. We found that it is women's educational attainment of lower secondary education or higher that drives the accelerating effect of education on fertility decline. For family planning, we found that the accelerating effect on fertility decline operates primarily through increasing contraceptive prevalence rather than through reducing unmet need for family planning. We then built upon this work in Chapter 3 to develop a Bayesian hierarchical model for conditional probabilistic projections of TFR given policy interventions targeting women's educational attainment

and access to family planning. The conditional TFR projection model creates projections based only on the estimated accelerating effect of increasing educational attainment and contraceptive prevalence on fertility decline, thus enabling the creation of intervention-based probabilistic projections of TFR. The assumptions needed for a causal interpretation of the conditional TFR projections were also specified using the accelerating effect framework. Using the conditional projection model, we created probabilistic projections of fertility and population size given a variety of policy intervention scenarios corresponding to meeting the SDG targets for universal secondary education and universal access to family planning.

Chapter 4 addresses the second line of work, where we developed a multiple imputation method for hierarchical nonlinear time series data with a motivating application to a secondary school enrollment data set. Estimates of demographic quantities are typically based on censuses and surveys that may be conducted annually in some countries but may only occur sporadically in others. For analyses that involve comparisons of demographic quantities across countries and across times, this can result in hierarchical data sets with large amounts of missing data. This is the case for the school enrollment data set, where historical estimates of two related measures of school enrollment rates are available with differing amounts of missing data. The two measures of school enrollment rates have a strong, nonlinear relationship that could be used to impute one enrollment rate measure from the other. However, previously existing multiple imputation methods for hierarchical time series data generally assume the variables with missing data have a linear relationship and thus can result in poor imputations for the school enrollment data. Multiple imputation methods can also perform poorly for estimation of substantive analysis models if the analysis model is uncongenial to the imputation model, which can occur when imputation and analysis are conducted in independent phases. Uncongeniality is common in practice for analyses using survey data, especially if the organization that publishes the data also publishes imputed values for the missing data that are then used in analyses by external researchers.

We proposed a multiple imputation method based on a Bayesian hierarchical model that accounts for the hierarchical and time series nature of the data and is able to accommodate a nonlinear relationship between variables with missing data. We compared the proposed method with several previously existing multiple imputation methods for hierarchical time series data through an application to the enrollment data and a simulation study, We found that the proposed method led to substantial improvements in estimation of parameters in uncongenial analysis models and for prediction of individual missing values.

5.2 Future Research

5.2.1 Conditional probabilistic projection model for TFR

The Bayesian hierarchical model for conditional TFR projections could be improved in several ways. Starting with the 2022 revision of the UN World Population Prospects, the UN has begun producing estimates and projections of fertility and population size on an annual scale rather than on the scale of five-year time periods. However, the conditional TFR projection model is currently constrained to the five-year time scale. This restriction is primarily due to the availability of the educational attainment data, where estimates of historical educational attainment from the Wittgenstein Centre are only available in five-year increments and the projection model for educational attainment was developed under the assumption of five-year increments. Extending the conditional TFR projection methodology to the one-year time scale is of interest for future work, as policymakers may be interested in exploring the potential impact of policy interventions targeting education and family planning on a more granular time scale. This extension could be built upon the one-year extension of the fertility projection model used by the UN and developed by Liu et al. (2023). The one-year extension of the conditional projection model would ideally use one-year estimates and projections of all covariates, including educational attainment from the Wittgenstein Centre. In the absence of one-year estimates and projections of the covariates,

interpolation of the five-year estimates and projections could be used as an approximation.

The conditional TFR model currently ignores uncertainty about past values of TFR and the covariates. However, there can be substantial uncertainty about the true value of these estimates of the past. This may be of particular concern for estimates of contraceptive prevalence, as historical survey data on contraceptive use is sparse for many countries. The estimates of past contraceptive prevalence obtained using the methodology of Kantorová et al. (2020) can be primarily model-based for some countries rather than reflecting observed survey estimates. Uncertainty about these model-based estimates of the past is not currently accounted for in estimation of the accelerating effect of the covariates on fertility decline. Uncertainty about past values of TFR is also of concern, especially as the time periods when countries have high fertility often correspond to the time periods when countries do not have robust vital registration systems. Liu and Raftery (2020b) extended the fertility projection model used by the UN to account for uncertainty about past values of TFR by adding an additional level of hierarchy to the Bayesian hierarchical model that represents the different primary data sources. The extension of the conditional fertility projection model to incorporate uncertainty about past values in TFR and the covariates could be conducted in a similar way.

Another possible extension of the conditional TFR projection model that may be of particular interest for policymakers is the ability to create one-country projections. Estimation of the full Bayesian hierarchical model for all countries is computationally intensive, requiring several hours for convergence of the MCMC algorithm. Policymakers may only be interested in exploring the potential impact of specific policy interventions on future fertility and population size for a single country. They may also have their own estimates of TFR, women's educational attainment, contraceptive prevalence, or GDP per capita for their country that they wish to use to estimate the conditional projection model. For this use case, running the full MCMC algorithm may be computationally restrictive. One possible way to increase

utility of the conditional TFR projection model could be to create a one-country version. Most of the global parameters of the Bayesian hierarchical model could be fixed at values obtained from a previous run of the model for all countries. An exception could be made for the coefficients on the covariates, which could be estimated in the one-country version using informative priors based on the values of the coefficients obtained from a previous run of the model for all countries.

5.2.2 Multiple imputation for hierarchical nonlinear time series data

The multiple imputation method for hierarchical nonlinear time series data developed in Chapter 4 could be extended for greater practical utility. The method is currently restricted to the bivariate setting with continuous variables, where one variable is the variable of interest for analyses and the other variable is an auxiliary variable that has a smaller amount of missing data than and has a nonlinear relationship with the variable interest. The method could be extended to the multivariate setting by adding additional univariate conditional terms to the sequential decomposition of the joint distribution. The method could also be extended to accommodate categorical variables through the use of generalized regression models, for example using similar methodology as Lee and Mitra (2016). The choice of the ordering of the added conditional distributions would need to be carefully considered, but could follow standard recommendations from Rubin and Schafer (1990) based on variable type and percentage of missing values. The imputation models for the added variables could be assumed to be linear or could incorporate a separate spline term for each marginal relationship in the conditional imputation model.

BIBLIOGRAPHY

Abel, G. J., Barakat, B., KC, S., and Lutz, W. (2016). Meeting the sustainable development goals leads to lower world population growth. *Proceedings of the National Academy of Sciences*, 113(50):14294–14299.

Adserà, A. (2017a). Education and fertility in the context of rising inequality. *Vienna Yearbook of Population Research*, 15:63–92.

Adserà, A. (2017b). The future fertility of highly educated women: the role of educational composition shifts and labor market barriers. *Vienna Yearbook of Population Research*, 15:19–25.

Ainsworth, M., Beegle, K., and Nyamete, A. (1996). The impact of women's schooling on fertility and contraceptive use: A study of fourteen sub-Saharan African countries. *The World Bank Economic Review*, 10(1):85–122.

Alkema, L. (2020). The Global Burden of Disease fertility forecasts: Summary of the approach used and associated statistical concerns. OSF Preprints.

Alkema, L., Kantorová, V., Menozzi, C., and Biddlecom, A. (2013). National, regional, and global rates and trends in contraceptive prevalence and unmet need for family planning between 1990 and 2015: a systematic and comprehensive analysis. *The Lancet*, 381:1642–1652.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T., and Heilig, G. K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography*, 48(3):815–839.

Angeles, G., Guilkey, D. K., and Mroz, T. A. (2005). The effects of education and family planning programs on fertility in Indonesia. *Economic Development and Cultural Change*, 54(1):165–201.

Axinn, W. G. and Barber, J. S. (2001). Mass education and fertility transition. *American Sociological Review*, 66(4):481–505.

Azose, J. J. and Raftery, A. E. (2015). Bayesian probabilistic projection of international migration rates. *Demography*, 52(5):1627–1650.

Barakat, B. (2016). wicedproj. `https://github.com/bifouba/wicedproj`.

Bartlett, J. W., Seaman, S. R., White, I. R., and Carpenter, J. R. (2015). Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Statistical Methods in Medical Research*, 24(4):462–487.

Behrman, J. A. (2015). Does schooling affect womens desired fertility? evidence from Malawi, Uganda, and Ethiopia. *Demography*, 52(3):787–809.

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, 15(4):651–675.

Bongaarts, J. (1978). A framework for analyzing the proximate determinants of fertility. *Population and Development Review*, 4(1):105–132.

Bongaarts, J. (1987). The proximate determinants of exceptionally high fertility. *Population and Development Review*, 13(1):133–139.

Bongaarts, J. (2003). Completing the fertility transition in the developing world: The role of educational differences and fertility preferences. *Population Studies*, 57(3):321–335.

Bongaarts, J. (2010). The causes of educational differences in fertility in sub-Saharan Africa. Poverty, Gender, and Youth Working Paper no. 20. New York: Population Council.

Bongaarts, J. (2013). Demographic trends and implications for development. IUSSP 2013 Meeting, Busan.

Bongaarts, J. (2016). Development: Slow down population growth. *Nature*, 530:409–412.

Bongaarts, J. (2017). The effect of contraception on fertility: Is sub-Saharan Africa different? *Demographic Research*, 37(6):129–146.

Bongaarts, J. and Casterline, J. (2013). Fertility transition: Is sub-Saharan Africa different? *Population and Development Review*, 38(s1):153–168.

Bongaarts, J., Frank, O., and Lesthaeghe, R. (1984). The proximate determinants of fertility in sub-Saharan Africa. *Population and Development Review*, 10(3):511–537.

Bongaarts, J. and Hardee, K. (2019). Trends in contraceptive prevalence in sub-Saharan Africa: The roles of family planning programs and education. *African Journal of Reproductive Health*, 23(3):96–105.

Bongaarts, J., Mensch, B. S., and Blanc, A. K. (2017). Trends in the age at reproductive transitions in the developing world: The role of education. *Population Studies*, 71(2):139–154.

Bricker, D. and Ibbitson, J. (2019). *Empty Planet: The Shock of Global Population Decline*. Crown, New York.

Cahill, N., Weinberger, M. W., and Alkema, L. (2020). What increase in modern contraceptive use is needed in FP2020 countries to reach 75% demand satisfied by 2030? An assessment using the Accelerated Transition Method and Family Planning Estimation Model. *Gates Open Research*, 4(113).

Caldwell, J. C. (1982). *Theory of Fertility Decline*. Academic Press, New York.

Caldwell, J. C., Reddy, P. H., and Caldwell, P. (1985). Educational transition in rural south India. *Population and Development Review*, 11(1):29–51.

Cleland, J. and Wilson, C. (1987). Demand theories of the fertility transition: An iconoclastic view. *Population Studies*, 41(1):5–30.

Cochrane, S. H. (1979). Fertility and education: What do we really know? World Bank Staff Occasional Papers No. OCP 26. Baltimore: The Johns Hopkins University Press.

Cygan-Rehm, K. and Maeder, M. (2013). The effect of education on fertility: Evidence from a compulsory schooling reform. *Labour Economics*, 25:35–48.

de Silva, T. and Tenreyro, S. (2017). Population control policies and fertility convergence. *Journal of Economic Perspectives*, 31(4):205–228.

Easterlin, R. A. and Crimmins, E. M. (1985). *The Fertility Revolution: A Supply-Demand Analysis*. University of Chicago Press, Chicago.

Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological review*, 70(3):193242.

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.

Enders, C. K., Du, H., and Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods*, 25(1):88–112.

Enders, C. K., Mistler, S. A., and Keller, B. T. (2016). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, 21(2):222–240.

Fabic, M. S., Choi, Y., Bongaarts, J., Darroch, J. E., Ross, J. A., Stover, J., Tsui, A. O., Upadhyay, J., and Starbird, E. (2015). Meeting demand for family planning within a generation: the post-2015 agenda. *Lancet*, 385(9981):1928–1931.

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

Friedman, J., York, H., Graetz, N., Woyczynski, L., Whisnant, J., Hay, S. I., and Gakidou, E. (2020). Measuring and forecasting progress towards the education-related SDG targets. *Nature*, 580:636–639.

Garenne, M. M. (2008). Fertility Changes in Sub-Saharan Africa. DHS Comparative Reports No. 18. Calverton, Maryland, USA: Macro International Inc.

Gelman, A., King, G., and Liu, C. (1998). Not asked and not answered: Multiple imputation for multiple surveys. *Journal of the American Statistical Association*, 93(443):846–857.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.*, 7(4):457–472.

Gertler, P. J. and Molyneaux, J. W. (1994). How economic development and family planning programs combined to reduce Indonesian fertility. *Demography*, 31(1):3363.

Gietel-Basten, S. and Sobotka, T. (2020). Uncertain population futures: Critical reflections on the IHME scenarios of future fertility, mortality, migration and population trends from 2017 to 2100. *SocArXiv*.

Gietel-Basten, S. and Sobotka, T. (2021). Trends in population health and demography.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Godwin, J. and Raftery, A. E. (2017). Bayesian projection of life expectancy accounting for the HIV/AIDS epidemic. *Demographic Research*, 37:1549–1610.

Goepp, V. (2022). aspline: Spline regression with adaptive knot selection. R package version 0.2.0.

Goepp, V., Bouaziz, O., and Nuel, G. (2018). Spline regression with automatic knot selection. *arXiv preprint arXiv:1808.01770*.

Goldstein, H., Carpenter, J. R., and Browne, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 177(2):553–564.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438.

Grant, M. J. (2015). The demographic promise of expanded female education: Trends in the age at first birth in Malawi. *Population and Development Review*, 41(3):409–438.

Grund, S., Lüdtke, O., and Robitzsch, A. (2021). Multiple imputation of missing data in multilevel models with the r package mdmb: a flexible sequential modeling approach. *Behavior Research Methods*, 53:2631–2649.

He, Y., Yucel, R., and Raghunathan, T. E. (2011). A functional multiple imputation approach to incomplete longitudinal data. *Statistics in Medicine*, 30(10):1137–1156.

Herzer, D., Strulik, H., and Vollmer, S. (2012). The long-run determinants of fertility: one century of demographic change 1900-1999. *Journal of Economic Growth*, 17(4):357–385.

Hirschman, C. (1994). Why fertility changes. *Annual Review of Sociology*, 20:203–233.

Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581.

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (2002). Bayesian methods for generalized linear models with covariates missing at random. *Canadian Journal of Statistics*, 30(1):55–78.

Ibrahim, J. G., Lipsitz, S. R., and Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):173–190.

Kaneda, T., Falk, M., and Patierno, K. (2021). Understanding and comparing population projections in sub-Saharan Africa. `https://www.prb.org/resources/understanding-and-comparing-population-projections-in-sub-saharan-africa/`. Accessed: 2021-07-23.

Kantorová, V., Wheldon, M. C., Ueffing, P., and Dasgupta, A. N. (2020). Estimating progress towards meeting women's contraceptive needs in 185 countries: A Bayesian hierarchical modelling study. *PLoS Med*, 17(2): e1003026.

King, G., Honaker, J., Joseph, A., and Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1):49–69.

Kirk, D. and Pillet, B. (1998). Fertility Levels, Trends, and Differentials in Sub-Saharan Africa in the 1980s and 1990s. *Studies in Family Planning*, 29(1):1–22.

Lee, M. C. and Mitra, R. (2016). Multiply imputing missing values in data sets with mixed measurement scales using a sequence of generalised linear models. *Computational Statistics & Data Analysis*, 95:24–38.

Lee, R. and Mason, A. (2006). What is the demographic dividend? *Finance and Development*, 43(3):16–24.

Lipsitz, S. R. and Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika*, 83(4):916–922.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons Inc.

Liu, D. H. and Raftery, A. E. (2020a). How do education and family planning accelerate fertility decline? *Population and Development Review*, 46:409–441.

Liu, D. H. and Raftery, A. E. (in press). Bayesian projections of total fertility rate conditional on the United Nations sustainable development goals. *The Annals of Applied Statistics*.

Liu, M., Taylor, J. M. G., and Belin, T. R. (2000). Multiple imputation and posterior simulation for multivariate missing data in longitudinal studies. *Biometrics*, 56:1157–1163.

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

Liu, P., Ševčíková, H., and Raftery, A. E. (2023). Probabilistic estimation and projection of the annual total fertility rate accounting for past uncertainty: A major update of the bayestfr r package. *Journal of Statistical Software*, 106:1–36.

Lüdtke, O., Robitzsch, A., and Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, 22(1):141–165.

Lüdtke, O., Robitzsch, A., and West, S. G. (2020). Regression models involving nonlinear effects with missing data: A sequential modeling approach using bayesian estimation. *Psychological Methods*, 25(2):157–181.

Lutz, W. (2023). Population decline will likely become a global trend and benefit long-term human wellbeing. *Vienna Yearbook of Population Research*, 21.

Lutz, W., Butz, W. P., and KC, S., editors (2014). *World Population and Human Capital in the Twenty-First Century*. Oxford University Press, Oxford.

Lutz, W., Goujon, A., KC, S., Stonawski, M., and Stilianakis, N., editors (2018). *Demographic and Human Capital Scenarios for the 21st Century: 2018 assessment for 201 countries*. Publications Office of the European Union, Luxembourg.

Lutz, W. and Skirbekk, V. (2014). How education drives demography and knowledge informs projections. In Lutz, W., Butz, W. P., and KC, S., editors, *World Population and Human Capital in the 21st Century*, pages 14–38. Oxford University Press, Oxford.

Maathuis, M. H. and Colombo, D. (2015). A generalized back-door criterion. *The Annals of Statistics*, 43(3):1060–1088.

Maddison Project Database, version 2018. Bolt, Jutta and Inklaar, Robert and de Jong, Herman and van Zanden, Jan Luiten (2018). Rebasing 'Maddison': new income comparisons and the shape of long-run economic development, Maddison Project working paper 10.

Martín, T. C. (1995). Women's education and fertility: Results from 26 Demographic and Health Surveys. *Studies in Family Planning*, 26(4):187–202.

Masih, A. M. M. and Masih, R. (2000). The dynamics of fertility, family planning and female education in a developing economy. *Applied Economics*, 32(12):1617–1627.

Mason, A. and Lee, R. (2006). Reform and support systems for the elderly in developing countries: capturing the second demographic dividend. *Genus*, 62:11–35.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, 9(4):538–558.

Nguyen, C. D., Carlin, J. B., and Lee, K. J. (2017). Model checking in multiple imputation: an overview and case study. *Emerging Themes in Epidemiology*, 14(8).

O'Neill, B. C., Dalton, M., Fuchs, R., Jiang, L., Pachauri, S., and Zigova, K. (2010). Global demographic trends and future carbon emissions. *PNAS*, 107:17521–17526.

Osili, U. O. and Long, B. T. (2008). Does female schooling reduce fertility? evidence from Nigeria. *Journal of Development Economics*, 87(1):57–75.

Pearl, J. (1993). Comment: Graphical models, causality and intervention. *Statist. Sci.*, 8(3):266–269.

Preston, S. H., Heuveline, P., and Guillot, M., editors (2001). *Demography: Measuring and Modeling Population Processes*. Blackwell Publishers, Oxford.

Raftery, A. E. (1995). Bayesian model selection for social research. *Sociological Methodology*, 29:111–163.

Raftery, A. E., Alkema, L., and Gerland, P. (2014). Bayesian population projections for the United Nations. *Statistical Science*, 29(1):58–68.

Raftery, A. E. and Lewis, S. M. (1996). Implementing MCMC. In Gilks, W., Richardson, S., and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*, pages 115–130. Chapman and Hall, London.

Raftery, A. E., Lewis, S. M., and Aghajanian, A. (1995). Demand or ideation? evidence from the Iranian marital fertility decline. *Demography*, 32(2):159–182.

Raftery, A. E., Zimmer, A., Frierson, D. M. W., Startz, R., and Liu, P. (2017). Less than 2 C warming by 2100 unlikely. *Nature Climate Change*, 7:637–641.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27(1):85–95.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association*, 72(359):538–543.

Rubin, D. B. (1978). Multiple imputations in sample surveys-a phenomenological bayesian approach to nonresponse. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, volume 1, pages 20–34. American Statistical Association.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473–489.

Rubin, D. B. and Schafer, J. L. (1990). Efficiently creating multiple imputations for incomplete multivariate normal data. In *Proceedings of the Statistical Computing Section of the American Statistical Association*, pages 83–88. American Statistical Association.

Savalei, V. and Rhemtulla, M. (2012). On obtaining estimates of the fraction of missing information from full information maximum likelihood. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(3):477–494.

Schafer, J. L. (1997a). *Analysis of Incomplete Multivariate Data*. CRC Press.

Schafer, J. L. (1997b). Imputation of missing covariates under a multivariate linear mixed model. Technical report, Dept. of Statistics, The Pennsylvania State University.

Schafer, J. L. and Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33(4):545–571.

Schafer, J. L. and Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, 11(2):437–457.

Schoumaker, B. (2014). Quality and Consistency of DHS Fertility Estimates, 1990 to 2012. DHS Methodological Reports No. 12. Rockville:ICF International.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464.

Ševčíková, H., , and Raftery, A. E. (2016). bayesPop: Probabilistic population projections. *Journal of Statistical Software*, 75(5):1–29.

Ševčíková, H., Alkema, L., and Raftery, A. E. (2011). bayesTFR: An R package for probabilistic projections of the total fertility rate. *Journal of Statistical Software*, 43(1):1–29.

Sobotka, T., Beaujouan, E., and Van Bavel, J. (2017). Introduction: education and fertility in low-fertility settings. *Vienna Yearbook of Population Research*, 15:1–16.

Speidel, M., Drechsler, J., and Jolani, S. (2018). R package hmi: A convenient tool for hierarchical multiple imputation and beyond. Technical Report 16, IAB-Discussion Paper.

Speringer, M., Goujon, A., KC, S., Potančoková, M., Reiter, C., Jurasszovich, S., and Eder, J. (2019). Global reconstruction of educational attainment, 1950 to 2015: Methodology and assessment. Vienna Institute of Demography Working Papers No. 02/2019.

Subbarao, K. and Raney, L. (1995). Social gains from female education: A cross-national study. *Economic Development and Cultural Change*, 44(1):105–128.

Taljaard, M., Donner, A., and Klar, N. (2008). Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biometrical Journal*, 50(3):329–345.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Tsui, A. O. (2001). Population policies, family planning programs, and fertility: The record. *Population and Development Review*, 27:184–204.

UNESCO Institute for Statistics (2023). Background information on education statistics in the UIS database.

United Nations (2015). Transforming our world: the 2030 Agenda for Sustainable Development . `https://sdgs.un.org/2030agenda`.

United Nations, Department of Economic and Social Affairs, Population Division (2018). World Urbanization Prospects: The 2018 Revision. Online edition.

United Nations, Department of Economic and Social Affairs, Population Division (2019a). Estimates and Projections of Family Planning Indicators 2019.

United Nations, Department of Economic and Social Affairs, Population Division (2019b). World Contraceptive Use 2019 (POP/DB/CP/Rev2019).

United Nations, Department of Economic and Social Affairs, Population Division (2019c). World Population Prospects: The 2019 Revision. Online edition.

United Nations, Department of Economic and Social Affairs, Population Division (2019d). World Population Prospects 2019: Methodology of the United Nations population estimates and projections.

United Nations, Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022. Online edition.

van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., and Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12):1049–1064.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45:1–67.

Vollset, S. E., Goren, E., Yuan, C.-W., Cao, J., Smith, A. E., Hsiao, T., Bisignano, C., Azhar, G. S., Castro, E., Chalek, J., Dolgert, A. J., Frank, T., Fukutaki, K., Hay, S. I., Lozana, R., Mokdad, A. H., Nandakumar, V., Pierce, M., Pletcher, M., Robalik, T., Steuben, K. M., Wunrow, H. Y., Zlavog, B. S., and Murray, C. J. L. (2020). Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease study. *Lancet*, 396(10258):1285–1306.

Westoff, C. F. and Bankole, A. (2001). The Contraception Fertility Link in Sub-Saharan Africa and in Other Developing Countries. DHS Analytical Studies No. 4. Calverton: ORC Macro.

White, M. J., Muhidin, S., Andrzejewski, C., Tagoe, E., Knight, R., and Reed, H. (2008). Urbanization and fertility: An event-history analysis of coastal Ghana. *Demography*, 45(4):803–816.

Wilmoth, J. (2019). Global population projections: A critical comparison of key methods and assumptions. `https://www.un.org/en/development/desa/population/about/director/pdf/Wilmoth_APC_Nov2019_Script.pdf`.

Wittgenstein Centre for Demography and Global Human Capital (2018). Wittgenstein

centre data explorer version 2.0. Available at: http://www.wittgensteincentre.org/dataexplorer.

World Bank (2019). World Development Indicators.

World Bank (2021). World bank open data: School enrollment, secondary (% gross and % net). Available at: https://data.worldbank.org/.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7):557–585.

Xie, X. and Meng, X.-L. (2017). Dissecting multiple imputation from a multi-phase inference perspective: What happens when god's, imputer's and analyst's models are uncongenial? *Statistica Sinica*, 27:1485–1545.

Xu, D., Daniels, M. J., and Winterstein, A. G. (2016). Sequential bart for imputation of missing covariates. *Biostatistics*, 17(3):589–602.

Zhao, J. H. and Schafer, J. L. (2023). pan: Multiple imputation for multivariate panel or clustered data. R package version 1.8.

Appendix A

APPENDICES FOR CHAPTER 2

A.1 Verification of Model Selection for Education Using Non-cumulative Levels of Attainment

In the model selection for education, we only considered the cumulative levels of change over time in women's educational attainment for interpretation purposes. However, we note that the cumulative levels of attainment are subsets of one another. For example, LowSec+ Change is the change over time in the proportion of women who have attained lower secondary education or higher. Both UppSec+ Change, the change over time in the proportion of women who have attained upper secondary education or higher, and PostSec+ Change, the change over time in the proportion of women who have attained post secondary education, are subsets of LowSec+ Change. Although we selected from the cumulative parameterization of the change over time in women's attainment ("Attain+ Change"), we also looked at the estimated effects for the non-cumulative parameterization ("Attain Change") as to verify our variable selection results.

Table A.1 summarizes the model including variables measuring women's attainment and children's enrollment. For comparison purposes, Table A.1 contains both the model that uses Attain+ Change and the model that uses Attain Change. The different levels of women's attainment are abbreviated analogously to "LowSec" for lower secondary. The change over time in Net Enrollment Rate is abbreviated to NER Change.

In both the Attain+ Change and Attain Change versions of the model in Table A.1, we found the only significant education variables corresponded to the change over time in women's attainment for lower secondary education or higher. In the cumulative attainment

model, this was LowSec+ Change. In the non-cumulative attainment model, this was LowSec Change, UppSec Change, and PostSec Change. Neither of the variables measuring children's enrollment were significant. Since we constructed the "change over time" education variables to be positive when education is increasing, we expected to find positive coefficient estimates for the education variables. However, we found that several of the coefficient estimates, including the coefficient estimates for both enrollment variables, were negative.

Table A.1: Summary of model fit with all education variables and all control variables, fit by GLS with TFR decrement as the dependent variable

Model with Attain+ Change and NER Change			Model with Attain Change and NER Change		
	Estimate	t-value		Estimate	t-value
(Intercept)	0.33	17.4***	(Intercept)	0.33	17.4***
Expected TFR Decr	0.91	18.4***	Expected TFR Decr	0.91	18.4***
IncPri+ Change	−0.66	−1.3	IncPri Change	−0.66	−1.3
Pri+ Change	0.79	1.3	Pri Change	0.13	0.3
LowSec+ Change	2.11	3.2**	LowSec Change	2.25	3.7**
UppSec+ Change	−0.78	−1.0	UppSec Change	1.47	2.1*
PostSec+ Change	0.63	0.5	PostSec Change	2.10	2.4*
NER Change (Pri)	−0.16	−1.1	NER Change (Pri)	−0.16	−1.1
NER Change (Sec)	−0.16	−1.6	NER Change (Sec)	−0.16	−1.6
GDP Growth	−0.38	−1.4	GDP Growth	−0.38	−1.4
GDP Growth Change	0.55	2.9**	GDP Growth Change	0.55	2.9**
Urban Change	−0.31	−0.6	Urban Change	−0.31	−0.6
Child Mortality Decr	0.07	0.1	Child Mortality Decr	0.07	0.1
SSA	−0.08	−2.5*	SSA	−0.08	−2.5*
Within-cluster correlation	0.23		Within-cluster correlation	0.23	
R^2	0.52		R^2	0.52	
BIC	-52.30		BIC	-52.30	
Country-time pairs	550		Country-time pairs	550	

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

From the results in Table A.1, we have verified that women's attainment was selected over children's enrollment regardless of the parameterization of women's attainment. We also found the only significant levels of attainment corresponded to the levels lower secondary

or higher for both parameterizations of women's attainment. However, due to the limited availability of data on children's enrollment, the models in Table A.1 were fit using only 550 country-time pairs. Given that the selected education mechanism was attainment, we confirmed that the selected levels of attainment were truly lower secondary or higher once we considered all 666 country-time pairs. Table A.2 summarizes the model not including enrollment variables. We considered versions with Attain+ Change and Attain Change for comparison purposes. In the model with Attain+ Change, we found LowSec+ Change was the only significant education variable. In the model with Attain Change, we found the education level with the largest effect size was lower secondary. The next largest effect sizes corresponded to upper secondary and post secondary. Although PostSec Change was not significant in the model with Attain Change, we justified the choice of LowSec+ Change since the women who have attained post secondary education have necessarily also attained lower secondary and upper secondary education, both of which were significant. The results of the comparison in Table A.2 support our choice of LowSec+ Change as the selected education variable and confirm that lower secondary is the most important level of education in terms of effect size.

A.2 Contraceptive Prevalence for All Women

In our analyses, we used estimates of contraceptive prevalence from the UN for married or in-union women aged 15–49 years, available from 1970 onwards. A technical limitation of using these estimates is an incomplete exposure to risk of pregnancy, as the TFR measures births to all women and not only women who are married or in-union (Bongaarts, 2017). This limitation may have a downward bias on the estimated effect of contraceptive prevalence on TFR within SSA compared to the rest of the world (Bongaarts, 2017). The UN also provides estimates of contraceptive prevalence for all women (whether or not they are married or in-union), however these estimates are only available from 1990 onwards. This restriction

Table A.2: Summary of model fit with attainment variables and all control variables, fit by GLS with TFR decrement as the dependent variable

Model with Attain+ Change			Model with Attain Change		
	Estimate	t-value		Estimate	t-value
(Intercept)	0.30	16.3***	(Intercept)	0.30	16.3***
Expected TFR Decr	0.91	19.2***	Expected TFR Decr	0.91	19.2***
IncPri+ Change	−0.73	−1.5	IncPri Change	−0.73	−1.5
Pri+ Change	1.00	1.8	Pri Change	0.27	0.7
LowSec+ Change	2.38	3.9***	LowSec Change	2.65	4.8***
UppSec+ Change	−1.04	−1.5	UppSec Change	1.62	2.6**
PostSec+ Change	−0.49	−0.5	PostSec Change	1.13	1.4
GDP Growth	−0.45	−1.7*	GDP Growth	−0.45	−1.7*
GDP Growth Change	0.46	2.6*	GDP Growth Change	0.46	2.6*
Urban Change	−0.47	−1.0	Urban Change	−0.47	−1.0
Child Mortality Decr	0.70	0.9	Child Mortality Decr	0.70	0.9
SSA	−0.09	−2.6**	SSA	−0.09	−2.6**
Within-cluster correlation	0.30		Within-cluster correlation	0.30	
R^2	0.49		R^2	0.49	
BIC	-86.45		BIC	-86.45	
Country-time pairs	666		Country-time pairs	666	

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

reduces the size of our dataset substantially. When using the married or in-union estimates, we have 666 country-time pairs with observations from 121 countries. However, when using the estimates for all women, we only have 344 country-time pairs with observations from 104 countries.

To check if using contraceptive prevalence for married or in-union women has a downward bias on the estimated effect of CP (Modern) Change on TFR Decr within SSA compared to non-SSA, we compared the model with CP (Modern) Change for married or in-union women with the analogous model fit using CP (Modern) Change for all women. The model was fit using GLS with the UN region × time point clustering scheme and includes LowSec+ Change, the SSA indicator, Expected TFR Decr, GDP Growth, GDP Growth Change, Urban

Change, and Child Mortality Decr as additional covariates. All continuous variables were centered prior to model fitting. Table A.3 provides a comparison of the results of the model using contraceptive prevalence for married or in-union women with the equivalent model fit using contraceptive prevalence for all women.

When considering the contraceptive prevalence of married or in-union women, the estimated effect of CP (Modern) Change on TFR Decr was 1.83 in SSA and 3.38 in non-SSA, which is an estimated effect about 1.85 times larger in non-SSA than in SSA. When considering the contraceptive prevalence of all women, the estimated effect of CP (Modern) Change on TFR Decr was 2.29 in SSA and 4.06 in non-SSA, which is an estimated effect about 1.77 times larger in non-SSA than in SSA. We found a notable and significant difference in the effect of CP (Modern) Change on TFR Decr between SSA and non-SSA regardless of which measure of contraceptive prevalence we considered. The difference in relative effect sizes between the two regions was smaller when we consider the contraceptive prevalence of all women, however the results still indicate the accelerating effect of increased contraceptive prevalence on fertility decline was weaker in SSA than in non-SSA. As the resulting conclusions about whether SSA is different remain the same regardless of which measure of contraceptive prevalence was used, we chose to use contraceptive prevalence of married or in-union women to make use of data from 1970-1990.

A.3 Omitted Path Coefficients

The bidirectional arrows connecting GDP Growth and GDP Growth Change to SSA, Expected TFR Decr, Urban Change, LowSec+ Change, Child Mortality Decr, and CP (Modern) Change are omitted from the path diagram for readability. The path coefficients for the omitted bidirectional arrows correspond to the correlations reported in Table A.4.

Error terms for Urban Change, LowSec+ Change, Child Mortality Decr, CP (Modern) Change, and TFR Decr are omitted for readability. The path coefficients for the omitted

Table A.3: Comparison of models using contraceptive prevalence for married or in-union women and contraceptive prevalence for all women, fit by GLS with TFR decrement as the dependent variable

	Married or in-union women		All women	
	Estimate	t-value	Estimate	t-value
(Intercept)	0.31	19.6***	0.32	15.0***
Expected TFR Decr	0.81	18.3***	1.11	15.6***
LowSec+ Change	1.79	4.6***	2.22	4.6***
CP (Modern) Change	3.38	9.3***	4.06	5.9***
GDP Growth	−1.20	−3.3**	−1.50	−2.7**
GDP Growth Change	0.77	3.7***	0.94	3.3**
Urban Change	−1.46	−2.5*	−1.19	−1.5
Child Mortality Decr	0.44	0.4	2.93	1.4
SSA	−0.06	−2.1*	−0.11	−3.1
SSA:LowSec+ Change	−0.57	−0.8	−1.25	−1.7
SSA:CP (Modern) Change	−1.55	−2.8**	−1.76	−2.0*
SSA:GDP Growth	1.08	2.1*	1.48	2.3*
SSA:GDP Growth Change	−0.61	−1.7	−0.94	−2.2*
SSA:Urban Change	1.81	2.1*	2.15	1.7
SSA:Child Mortality Decr	−1.40	−0.9	−3.65	−1.5
Within-cluster correlation	0.23		0.31	
R^2	0.57		0.61	
BIC	-168.49		-147.40	

*** denotes P< 0.001, ** denotes P< 0.01, and * denotes P< 0.05

error terms are reported in Table A.5.

Arrows corresponding to effects with P $>$ 0.05 are also omitted for readability. The path coefficients for the unidirectional arrows going from the "starting" variable to the "ending" variable are reported in Table A.6

Table A.4: Path coefficients for bidirectional arrows (omitted from path diagram for readability) from GDP Growth and GDP Growth Change to all other covariates

	GDP Growth Coefficient	GDP Growth Change Coefficient
SSA	-0.03	0.02
Expected TFR Decr	-0.06	-0.07
Urban Change	0.11	-0.09
LowSec+ Change	0.05	-0.08
Child Mortality Decr	0.11	-0.02
CP (Modern) Change	0.08	-0.01
GDP Growth	—	0.53
GDP Growth Change	0.53	—

Table A.5: Path coefficients for error terms (omitted from path diagram for readability)

	Error Term Coefficient
Urban Change	0.97
LowSec+ Change	0.93
Child Mortality Decr	0.94
CP (Modern) Change	0.94
TFR Decr	0.67

Table A.6: Path coefficients for unidirectional arrows from "starting" variable to "ending" variable corresponding to effects with $P > 0.05$ (omitted from path diagram for readability)

Starting Variable	Ending Variable	Path Coefficient
SSA Indicator	Child Mortality Decr	0.21
SSA Indicator	CP (Modern) Change	0.01
Urban Change	CP (Modern) Change	0.04
Urban Change	TFR Decr	-0.02
Child Mortality Decr	TFR Decr	-0.01

Appendix B

APPENDICES FOR CHAPTER 3

B.1 Full Model Specification

The full specification of the conditional TFR projection model is provided in this section. The Bayesian hierarchical model for Phase II has three levels (observation, country, and world) and is given by:

$$
\begin{aligned}
\Delta f_{c,t+1} &= f_{c,t+1} - f_{c,t} \\
&= -\, g(f_{c,t}|\boldsymbol{\theta}_c) + \boldsymbol{\Delta X}_{c,t}\boldsymbol{\beta} + \varepsilon_{c,t}, \\
\boldsymbol{\theta}_c &= (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)\,, \\
\boldsymbol{\beta} &= \begin{bmatrix} \beta_E \\ \beta_F \\ \beta_G \\ \beta_{E,SSA} \\ \beta_{F,SSA} \\ \beta_{G,SSA} \end{bmatrix}, \qquad \boldsymbol{\Delta X}_{c,t}^T = \begin{bmatrix} (\Delta\text{LowSec+})_{c,t} \\ (\Delta\text{CP})_{c,t} \\ (\Delta\text{GDP})_{c,t} \\ (\Delta\text{LowSec+})_{c,t} \times SSA_c \\ (\Delta\text{CP})_{c,t} \times SSA_c \\ (\Delta\text{GDP})_{c,t} \times SSA_c \end{bmatrix}
\end{aligned}
$$

The double logistic function $g(f_{c,t}|\boldsymbol{\theta}_c)$ takes on the current value of TFR ($f_{c,t}$) and a vector of country-specific parameters $\boldsymbol{\theta}_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$ as inputs and is defined as

$$
\begin{aligned}
g(f_{c,t}|\boldsymbol{\theta}_c) = & \frac{-d_c}{1 + \exp\left(-2\frac{\ln(9)}{\Delta_{c1}}\left(f_{c,t} - \sum_i \Delta_{ci} + 0.5\Delta_{c1}\right)\right)} \\
& + \frac{d_c}{1 + \exp\left(-2\frac{\ln(9)}{\Delta_{c3}}\left(f_{c,t} - \Delta_{c4} - 0.5\Delta_{c3}\right)\right)}
\end{aligned}
$$

The error term is specified as follows:

$$\begin{aligned}
\boldsymbol{\varepsilon} &\sim N(\tilde{\boldsymbol{\mu}}, \Sigma), \\
\Sigma &= \text{diag}(\tilde{\boldsymbol{\sigma}}) \cdot R \cdot \text{diag}(\tilde{\boldsymbol{\sigma}}), \\
R[i,j] &= \begin{cases} 1 & \text{if } i = j \\ \rho^{[bc]} & \text{if } i, j \in \text{same UN region and same time period} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

where $\tilde{\boldsymbol{\mu}}$ is a vector of $\mu_{c,t}$ ordered by UN region and time period and $\tilde{\boldsymbol{\sigma}}$ is a vector of $\sigma(f_{c,t})$, also ordered by UN Region and time period. The (i,j)th term of the correlation matrix R represents the correlation between country-time pair i and country-time pair j. If observations i and j are from the same time period and refer to countries within the same UN region, the between-country correlation is given as $\rho^{[bc]}$. The mean $\mu_{c,t}$ is defined as:

$$\mu_{c,t} = \begin{cases} m_\tau & \text{if } t = \tau_c \\ 0 & \text{if } t > \tau_c \end{cases}$$

where τ_c denotes the start time period of Phase II for country c and m_τ is the mean of the error in the start period. The variance $\sigma_{c,t}$ is defined as:

$$\sigma_{c,t} = \begin{cases} s_\tau^2 & \text{if } t = \tau_c \\ c_{1975}(t)\left(\sigma_0 + (f_{c,t} - S)\left(-aI_{[S,\infty)}(f_{c,t}) + bI_{[0,S)}(f_{c,t})\right)\right) & \text{if } t > \tau_c \end{cases}$$

where s_τ is the standard deviation of the error in the start period. The maximum standard deviation of the distortions is denoted σ_0, which is attained at TFR level S. Parameters a and b are multipliers of the standard deviation used to model the linear decrease for larger and smaller outcomes of TFR. The parameter $c_{1975}(t)$ is a constant that is added to model

the higher error variance of the distortions before 1975:

$$c_{1975}(t) = \begin{cases} c_{1975} & \text{if } t \leq 1970 - 1975 \\ 1 & \text{if } t > 1970 - 1975 \end{cases}$$

In the second level of the model, the country-specific parameters $\{\gamma_{ci}, U_c, d_c, \Delta_{c4}\}$ for $i = 1, 2, 3$ are specified as follows:

$$\begin{aligned}
U_c =& \begin{cases} f_{c,\tau} & \text{if } \tau_c \geq 1975 - 1955 \\ \sim U(\min\{5.5, \max_t f_{c,t}\}, 8.8) & \text{for } \tau_c < 1950 - 1955, \end{cases} \\
d_c^* =& \log\left(\frac{d_c - 0.25}{2.5 - d_c}\right), \\
d_c^* \sim& N(\chi, \Psi^2), \\
\Delta_{c4}^* =& \log\left(\frac{\Delta_{c4} - 1}{2.5 - \Delta_{c4}}\right), \\
\Delta_{c4}^* \sim& N(\Delta_4, \delta_4^2) \\
p_{ci} =& \frac{\Delta_{ci}}{U_c - \Delta_{c4}} \text{ for } i = 1, 2, 3, \\
p_{ci} =& \frac{\exp(\gamma_{ci})}{\sum_{j=1}^{3} \exp(\gamma_{cj})}, \\
\gamma_{ci} \sim& N(\alpha_i, \delta_i^2).
\end{aligned}$$

In the third level of the model, the world-level hyperparameters $\{\boldsymbol{\beta}, \rho^{[bc]}, \chi, \Psi^2, \Delta_4, \delta_4, \boldsymbol{\alpha}, \boldsymbol{\delta}\}$ and $\{a, b, S, \sigma_0, c_{1975}, m_\tau, s_\tau\}$ are specified via the following prior distributions:

$$\beta_j \sim N\left(0, 0.25 \times \frac{Var(\Delta f_{c,t})}{Var(\Delta X_j)}\right) \quad \text{for } j \in (E; F, G; E, SSA; F, SSA; G, SSA),$$

$$\begin{aligned}
\rho^{[bc]} \sim& \text{Uniform}(0, 1), \\
\chi \sim& N(-1.5, 0.6), \\
1/\Psi^2 \sim& \text{Gamma}(1, 0.6^2), \\
\alpha_1 \sim& N(-1, 1),
\end{aligned}$$

$$
\begin{aligned}
\alpha_2 &\sim N(0.5, 1), \\
\alpha_3 &\sim N(1.5, 1), \\
1/\delta_i^2 &\sim \text{Gamma}(1, 1), \text{ for } i = 1, 2, 3, \\
1/\delta_4^2 &\sim \text{Gamma}(1, 0.8^2), \\
\Delta_4 &\sim N(0.3, 0.8^2), \\
a &\sim U[0, 0.2], \\
b &\sim U[0, 0.2], \\
\sigma_0 &\sim U[0.01, 0.6], \\
c_{1975} &\sim U[0.8, 2], \\
S &\sim U[3.5, 6.5], \\
m_\tau &\sim N(-0.25, 0.4^2) \\
1/s_\tau^2 &\sim \text{Gamma}(1, 0.4^2).
\end{aligned}
$$

The covariates included in the observation level of the model are constructed as changes over time on the same scale as the TFR decrements $\Delta f_{c,t+1}$. For example, the change over time from t to $(t+1)$ for covariate X is denoted $\Delta X_{c,t+1}$. The covariates included are the change over time in the proportion of women aged 20-39 who have attained at least lower secondary, denoted ΔLowSec+, the change over time in the contraceptive prevalence of modern methods for all women of reproductive age, denoted ΔCP, and the GDP per capita percent change, denoted ΔGDP. We also consider interaction terms between all covariates and the indicator function SSA_c for whether country c is located in sub-Saharan Africa.

The continuous covariates $X = \{\text{LowSec+}, \text{CP}, \text{GDP}\}$ were centered at values X^* corresponding to the conditional expected value of X assuming no additional policy intervention. For GDP, X^* was defined as the mean for each country's ΔGDP values. For educational at-

tainment, $X^*_{c,t}$ was constructed conditional on the last observed value $X^*_{c,t-1}$ and conditional on the estimated model parameters for Bayesian hierarchical model for attainment from the Wittgenstein Centre (Lutz et al., 2018). The centering was as follows:

$$X^*_{c,t} = \text{probit}^{-1}(\text{probit}(X_{c,t-1}) + \hat{\mu} + \hat{b}_c)$$

where $\hat{\mu}$ and $\hat{b}_c$ are estimated by fitting the model

$$\text{probit}(\text{LowSec+}_{c,t}) = \text{probit}(\text{LowSec+}_{c,t-1}) + \mu + b_c + \varepsilon_{c,t}$$

where LowSec+ is the proportion of women aged 20-39 attaining lower secondary education or higher. Country-time pairs with observed LowSec+ of 0 were omitted from the estimation of the X^* model for LowSec+. The centered values for those country-time pairs was set to 0 for use in the estimation of the conditional TFR projection model.

For contraceptive prevalence, $X^*_{c,t}$ was constructed conditional on the estimated model parameters for the Bayesian hierarchical model for contraceptive prevalence and unmet need from Kantorová et al. (2020). Using outputs $\{\tilde{R}_c, \psi_c, \Psi_c, \tilde{P}_c, \omega_c, P_{c,t}\}$ from the Kantorová et al. model, the conditional expectation is constructed as:

$$\begin{aligned}
X^*_{c,t} &= R^*_{c,t} P^*_{c,t} \\
R^*_{c,t} &= \frac{\tilde{R}_c}{(1 + \exp(-\psi_c(t - \Psi_c)))} \\
P^*_{c,t} &= \\
\text{if } t > t^* : &\begin{cases} \tilde{P}_c \, \text{logit}^{-1}\left(\text{logit}\left(\frac{P_{c,t-1}}{\tilde{P}_c}\right) + \omega_c\right) & \text{if } P_{c,t-1} < \tilde{P}_c \\ \text{logit}(P_{c,t-1}) & \text{ow} \end{cases} \\
\text{if } t < t^* : &\begin{cases} \tilde{P}_c \, \text{logit}^{-1}\left(\text{logit}\left(\frac{P_{c,t+1}}{\tilde{P}_c}\right) - \omega_c\right) & \text{if } P_{c,t+1} < \tilde{P}_c \\ \text{logit}(P_{c,t+1}) & \text{ow} \end{cases}
\end{aligned}$$

As the term $P^*_{c,t}$ is constructed outwards starting from $t^* = 1990$, the conditional expected value X^* is constructed given the last observed value $X^*_{c,t-1}$ for years after 1990 and given the last observed value $X_{c,t+1}$ in years before 1990.

B.2 SDG Intervention Projections of Covariates

B.2.1 Women's educational attainment

Probabilistic projections of educational attainment corresponding to the SDG intervention were obtained from Abel et al. (2016). Abel et al. create SDG intervention projections using a modification of the Bayesian hierarchical model used to create the Wittgenstein Centre's (non-intervention) estimates and projections of educational attainment. The Wittgenstein Centre model projects educational attainment by country and sex for five-year age groups and five-year time steps. The modified model treats the SDG target levels of attainment as future observations at 2030 in the likelihood, allowing for the projected trend in attainment to increase at whatever rate is necessary to reach the target level in 2030. Abel et al. consider two translations of SDG Target 4.1, universal lower secondary education in 2030 and universal upper secondary education in 2030, corresponding to differences in policymaker interpretations of the "universal secondary education" target. We use the slightly more realistic translation of universal lower secondary education for our projections, though we found negligible differences between the two translations on our TFR projections.

B.2.2 Contraceptive prevalence

For projections of contraceptive prevalence assuming the SDGs are met in 2030, we modified the non-intervention projections from a converged simulation of the Bayesian hierarchical model for contraceptive prevalence and unmet need developed by Kantorová et al. (2020) and implemented in the "FPEMglobal" R package. From the causal framework underlying our model and from existing work such as Bongaarts and Hardee (2019), we know there is

an indirect effect of women's educational attainment on fertility decline through the impact educational attainment has on contraceptive prevalence. We consider both the indirect impact of SDG Target 4.1 on increasing the demand for family planning and the direct impact of SDG Target 3.7 on reducing unmet need for family planning when constructing SDG intervention projections of contraceptive prevalence.

In the Both SDGs (0% Unmet) scenario, we interpret the direct impact of Target 3.7 on family planning as unmet need will decline linearly to 0% in 2030, where we assume all unmet need is met by modern contraceptive use. This interpretation is implemented by modifying each trajectory of the non-intervention projections of contraceptive prevalence. For trajectory i, the corresponding SDG-intervention trajectory i^{SDG} is created as follows. For 2030, the trajectory is modified as:

$$\begin{aligned} p_{c,2030,3}^{(i^{SDG})} =& 0 \\ p_{c,2030,2}^{(i^{SDG})} =& p_{c,2030,2}^{(i)} + p_{c,2030,3}^{(i)} \end{aligned}$$

where $p_{c,2030,3}^{(i)}$ = the proportion of women experiencing unmet need in country c in year 2030 and $p_{c,2030,2}^{(i)}$ = the proportion of contraceptive prevalence of modern methods in country c in year 2030. For years 2020-2030 in i^{SDG}, we linearly interpolate between the observed 2020 values and the intervention-based 2030 values as follows:

$$p_{c,t,2}^{(i^{SDG})} = p_{c,t,2}^{(i)} + \lambda_t p_{c,t,3}^{(i)}$$

where λ_t goes from 0 in 2020 to 1 in 2030. We assume the family planning effort required to attain the SDG in 2030 will be sustained out to 2100 i.e. we set $\lambda_t = 1$ for $t \geq 2030$.

For the Both SDGs (75% Demand Satisfied) scenario, we interpret the direct impact of Target 3.7 on family planning as the demand for family planning satisfied by modern methods will increase to 75% by 2030. We use the accelerated transition method developed by Cahill et al. (2020) with some modifications. Demand for family planning satisfied by

modern methods is defined as:

$$\text{Demand Satisfied} = \frac{\text{Contraceptive Prevalence Modern}}{\text{Contraceptive Prevalence Total + Unmet Need}}$$

where the denominator includes contraceptive prevalence of both modern and traditional methods. We assume the expansion of family planning from meeting Target 3.7 applies only to modern contraceptive methods. We implement this intervention by modifying each trajectory of the non-intervention projections of contraceptive prevalence. For trajectory i, the corresponding SDG-intervention trajectory i^{SDG} was created using the following steps. Let $p_{c,t,3}$ denote the proportion of women experiencing unmet need in country c in year t, let $p_{c,t,2}$ denote the proportion of CP modern in country c in year t, and let $p_{c,t,1}$ denote the proportion of CP traditional in country c in year t. For each trajectory i, we first find the year t^* where the country is projected to reach 75% demand satisfied in the non-intervention reference projection. If the country is not projected to reach 75% demand satisfied, we set $t^* = 2100$ for trajectory i. If $t^* > 2030$, we create i^{SDG} as follows: Let u^* denote the level of unmet need for family planning and let r^* denote the level of contraceptive prevalence of traditional methods in year t^* under the non-intervention projections scenario for trajectory i. Then in 2030, we modify the trajectories as:

$$\begin{aligned}
p_{c,2030,3}^{(i^{SDG})} &= u^* \\
p_{c,2030,1}^{(i^{SDG})} &= r^* \\
p_{c,2030,2}^{(i^{SDG})} &= p_{c,2030,2}^{(i)} + \left(p_{c,2030,3}^{(i)} - u^*\right) + \left|p_{c,2030,1}^{(i)} - r^*\right|
\end{aligned}$$

where the absolute value allows us to accommodate for countries where CP traditional is projected to increase in the non-intervention projection. We assume any projected increase in CP traditional in the non-intervention scenario is instead met by modern methods under the SDG intervention. For years 2020-2030 in i^{SDG}, we linearly interpolate between the observed 2020 values and the intervention-based 2030 values:

$$p_{c,t,2}^{(i^{SDG})} = p_{c,t,2}^{(i)} + \lambda_t \left(p_{c,t,3}^{(i)} + p_{c,t,1}^{(i)}\right)$$

where λ_t goes from 0 in 2020 to $\lambda^* = \frac{\left(p^{(i)}_{c,2030,3} - u^*\right) + \left(p^{(i)}_{c,2030,1} - r^*\right)}{p^{(i)}_{c,2030,3} + p^{(i)}_{c,2030,1}}$ in 2030. We assume the family planning effort required to attain 75% demand satisfied in 2030 will be sustained out to 2100 i.e. $\lambda_t = \lambda^*$ for $t \geq 2030$. If $t^* \leq 2030$, the country is already projected to meet the target demand satisfied by 2030 in the non-intervention scenario, so no modification to trajectory i is needed and we set

$$p^{(i,SDG)}_{c,t,2} = p^{(i)}_{c,t,2}.$$

In both Both SDGs scenarios, we interpret the indirect effect of Target 4.1 on contraceptive prevalence as reducing the number of women who have no demand for family planning. This assumption follows the analytic framework for the determinants of fertility illustrated in Bongaarts (2010). First, we estimate the relationship between the proportion of women attaining lower secondary education or higher and the proportion of all women of reproductive age with no demand for family planning via OLS, pooled across countries and times. We find a regression coefficient of $\alpha = -0.2752$. This can be interpreted as for each 0.1 increase in the proportion of women who have attained lower secondary education or higher, we expect a decrease of 0.02752 in the proportion of women with no demand for family planning. Next, we use the median SDG intervention projection of ΔLowSec+ to find the expected changes in the proportion of women attaining lower secondary education or higher assuming Target 4.1 is met in 2030. The expected increase in contraceptive prevalence that comes from Target 4.1 was calculated as $\alpha \times \Delta$LowSec+ for the intervention time periods 2020-2025 to 2035-2040. These contributions were then added to the existing SDG intervention projections of contraceptive prevalence.

B.3 Model Diagnostics

The conditional projection model was estimated using Markov chain Monte Carlo (MCMC) in two stages. In the first stage, we estimate the double logistic parameters using

3 chains of length 360000 with burn-in of 20000 and thinned by 30. In the second stage, we estimate the $\boldsymbol{\beta}$ coefficients conditional on the results of the first stage. The second stage was estimated using 3 chains of length 300000 with burn-in of 2000 and thinned by 30. Convergence was checked using visual inspection of trace plots and assessment of the Raftery-Lewis diagnostic and the Gelman-Rubin diagnostic. The same diagnostics were used the check the convergence of the model used for the Education SDG Only projection scenario.

The run length diagnostic of Raftery and Lewis (1996) was assessed to determine if the MCMC has been run for enough iterations to estimate the 2.5% and 97.5% percentiles of the posterior distributions of all model parameters to within 0.0125 of the true quantile with probability of at least 0.95. In both stages of estimation, the MCMC chains exceeded the required sample size, indicating that both stages have been run for enough iterations. The convergence diagnostic of Gelman and Rubin (1992) was assessed using all three chains, where convergence is indicated by values below 1.1. In both stages of estimation, the potential scale reduction factors (PSRFs) fell below 1.1 for all parameters. Table B.1 summarizes convergence diagnostics for the $\boldsymbol{\beta}$ and $\rho^{[bc]}$ parameters from the second stage of estimation. Trace plots for the country-independent parameters from stage one of estimation are shown in Figure B.1. Trace plots for the $\boldsymbol{\beta}$ and $\rho^{[bc]}$ parameters from the second stage of estimation are shown in Figure B.2.

B.4 Additional Validation Exercises

B.4.1 Sensitivity of prior distributions

The majority of the prior distributions in the conditional TFR projection model reflect domain-specific prior knowledge. The choice of prior distributions for the double logistic parameters θ_c is discussed in Alkema et al. (2011). The choice of prior distribution for the between-country correlation parameter $\rho^{[bc]}$ reflects prior knowledge that the between-country correlation in TFR errors for countries that are geographically contiguous is positive

Table B.1: Convergence diagnostics for the $\boldsymbol{\beta}$ and $\rho^{[bc]}$ parameters from the second stage of estimation for the conditional TFR projection model. Columns PSRF and Upper CI give the point estimate and and upper bound of the 95% CI of the Gelman-Rubin diagnostic. Columns Burn-in, Total, and DF give the burn-in length, required sample size, and dependence factor for the Raftery-Lewis diagnostic for one randomly selected chain.

Parameter	PSRF	Upper CI	Burn-in	Total	DF
β_E	1.00	1.00	60	18240	30.4
β_{FP}	1.00	1.00	30	18060	30.1
β_{GDP}	1.00	1.00	60	18240	30.4
$\beta_{E,SSA}$	1.00	1.00	60	18690	31.2
$\beta_{FP,SSA}$	1.00	1.00	60	18240	30.4
$\beta_{GDP,SSA}$	1.00	1.00	60	17790	29.6
$\rho^{[bc]}$	1.00	1.00	60	18840	31.4

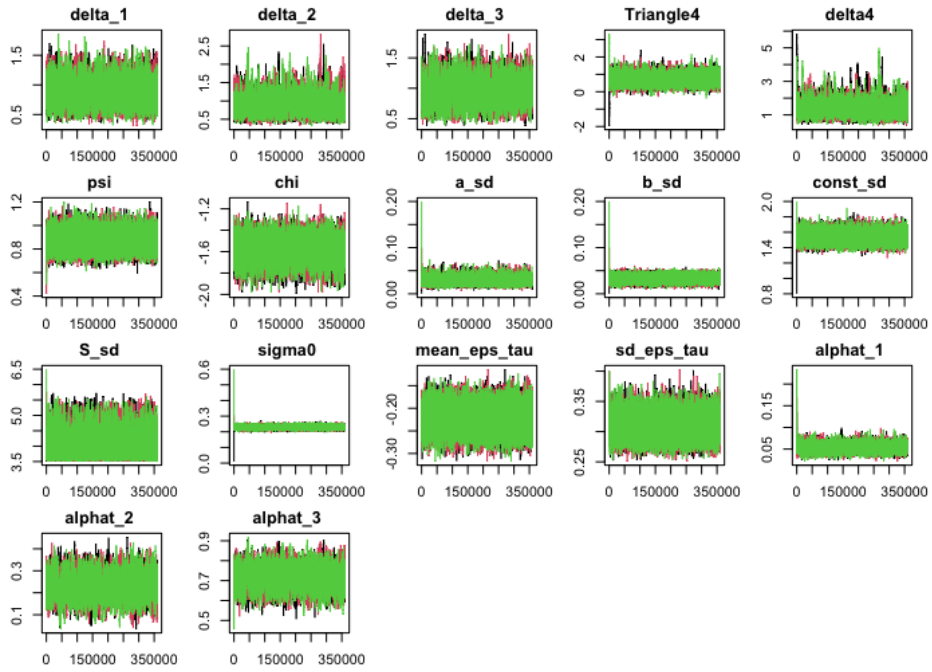


Figure B.1: Trace plots for country-independent parameters from first stage of estimation

(Fosdick and Raftery, 2014). However, the prior distribution for the $\boldsymbol{\beta}$ coefficients,

$$\beta_j \sim N\left(0, 0.25 \times \frac{Var(\Delta f_{c,t})}{Var(\Delta X_j)}\right) \quad \text{for } j \in (E; F, G; E, SSA; F, SSA; G, SSA)\,,$$

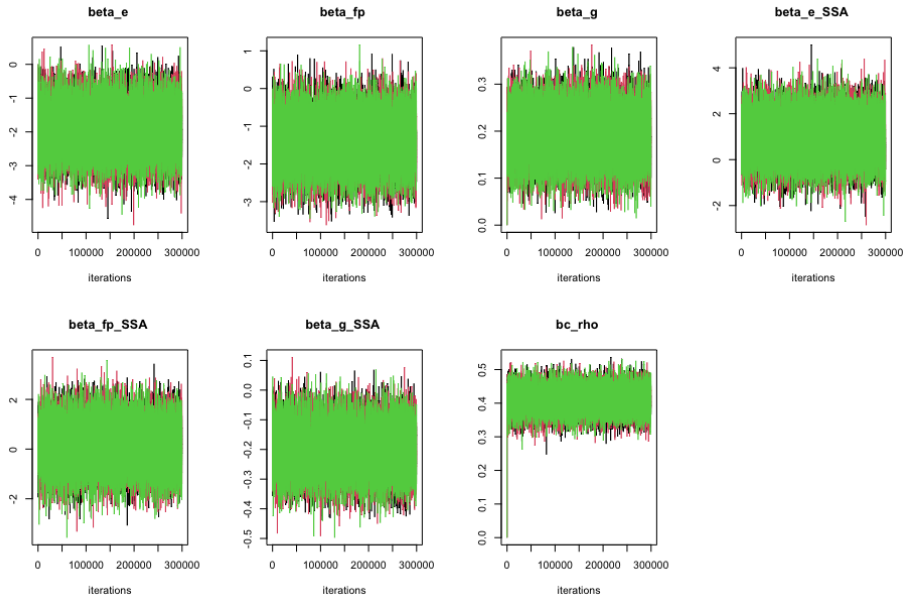


Figure B.2: Trace plots for $\boldsymbol{\beta}$ and ρ^{bc} from second stage of estimation

was chosen to be diffuse rather than to reflect domain-specific knowledge, where the prior variance is the ratio of the sample variance of observed changes in $f_{c,t}$ to the sample variance of observed changes in the covariate X_j.

We conduct a sensitivity analysis for the prior variance by re-estimating the second stage of the conditional projection model using wider prior variances and narrower prior variances. For each β_j, these prior variances are specified as

$$\text{Wider:} \quad \beta_j \sim N\left(0, 2 \times 0.25 \times \frac{Var(\Delta f_{c,t})}{Var(\Delta X_j)}\right)$$
$$\text{Narrower:} \quad \beta_j \sim N\left(0, 0.5 \times 0.25 \times \frac{Var(\Delta f_{c,t})}{Var(\Delta X_j)}\right).$$

We found that changing the prior variance for the β_j did not substantially impact the estimated posterior distributions for β_j. Figure B.3 shows a comparison of the density of the posterior distributions for β_j from all three prior specifications. Table B.2 compares the posterior quantiles of the distributions for β_j from all three prior specifications.

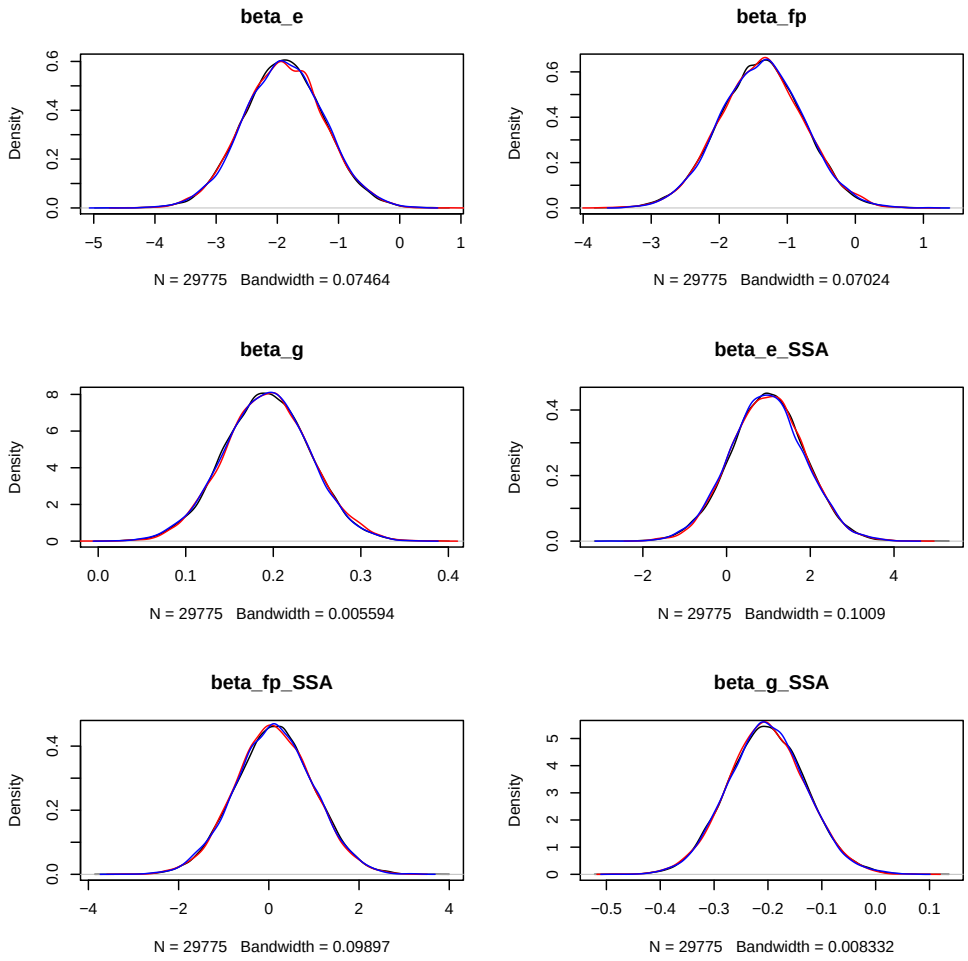


Figure B.3: Comparison of density of posterior distributions for $\boldsymbol{\beta}$ coefficients for original (black), wider (red), and narrower (blue) choices for prior variances

B.4.2 Validation of functional form of model

The conditional TFR projection model assumes the relationship between the covariates and fertility decline is linear. This assumption implies, for example, that changes in the education covariate have the same effect on fertility decline regardless of which stage of the fertility decline a country is in, conditional on the value of the other covariates. To check

Table B.2: Comparison of posterior quantiles for $\boldsymbol{\beta}$ parameters for original, wider, and narrower choices for prior variances

Parameter	Prior Choice	2.5%	25%	50%	75%	97.5%
β_E	Original	-3.1617	-2.3381	-1.8966	-1.4560	-0.6143
	Wider	-3.1758	-2.3437	-1.9031	-1.4617	-0.6004
	Narrower	-3.1853	-2.3356	-1.8878	-1.4403	-0.5952
β_F	Original	-2.5919	-1.7983	-1.3794	-0.9660	-0.1930
	Wider	-2.5792	-1.7935	-1.3731	-0.9628	-0.1524
	Narrower	-2.5707	-1.7913	-1.3667	-0.9603	-0.1619
β_G	Original	0.0961	0.1594	0.1927	0.2257	0.2882
	Wider	0.0989	0.1609	0.1935	0.2268	0.2926
	Narrower	0.0950	0.1590	0.1926	0.2250	0.2879
$\beta_{E,SSA}$	Original	-0.7688	0.3998	0.9878	1.5783	2.7099
	Wider	-0.6824	0.3911	0.9915	1.5821	2.6975
	Narrower	-0.7550	0.3670	0.9550	1.5463	2.7066
$\beta_{F,SSA}$	Original	-1.5474	-0.4459	0.1331	0.7117	1.8335
	Wider	-1.5633	-0.4572	0.1065	0.6939	1.7884
	Narrower	-1.5923	-0.4500	0.1211	0.7019	1.8163
$\beta_{G,SSA}$	Original	-0.3411	-0.2511	-0.2019	-0.1524	-0.0573
	Wider	-0.3430	-0.2509	-0.2032	-0.1545	-0.0598
	Narrower	-0.3403	-0.2499	-0.2024	-0.1536	-0.0583

this assumption empirically, we compared the linear model to an alternative model that uses a proportional relationship, namely

$$\Delta f_{c,t+1} = - g(f_{c,t}|\boldsymbol{\theta}_c) \exp\left(\boldsymbol{\Delta X_{c,t}\beta}\right) + \varepsilon_{c,t}.$$

The proportional relationship follows the same shape as the double logistic expected TFR decrement term, representing a larger impact of education on fertility decline at the peak of the decline compared to when the decline has just started or when the decline is ending.

We validated the functional form of our model by considering outcome variable

$$Y_{c,t+1} = \Delta f_{c,t+1} + g(f_{c,t}|\boldsymbol{\theta}_c).$$

Under the linear model, we have

$$E[Y_{c,t+1}] = \boldsymbol{\Delta X_{c,t}} \boldsymbol{\beta}.$$

Under the proportional model, we have

$$E[Y_{c,t+1}] = g(f_{c,t}|\boldsymbol{\theta}_c)\boldsymbol{\Delta X_{c,t}} \boldsymbol{\beta}$$

where we use the first-order power series approximation of the proportional model since the $\boldsymbol{\Delta X_{c,t}}$ are small. Using linear regression, we found the linear model fit slightly better based on R^2 (0.0715 vs. 0.0573). The two models performed similarly based on residuals plots and quantile-quantile plots. Figure B.4 shows the residuals from the linear model. Overall, the residuals look evenly dispersed with no clear trend. Based on this validation exercise, we conclude the linear form of the conditional projection model is appropriate.

B.4.3 Disaggregated out-of-sample validation

In addition to looking at the aggregated out-of-sample validation results, we also considered the results disaggregated by different levels of TFR and the covariates in the last time period used for estimation (2010–2015). For the TFR disaggregation, we created four groups based on 2010–2015 TFR level. The first group is denoted the"Lowest" group. These countries have TFR < 2.1 in 2015–2020, so based on our projection criterion these countries do not satisfy our "high-fertility" criterion and we do not create TFR projections using the conditional projection model. Instead, projections are created following the bayesTFR model. Thus, we expect the out-of-sample performance for the Lowest group to be the same between the conditional projection model and the bayesTFR model. All other countries are divided into three groups that are chosen to be roughly similar in size. We denote these the "Low," "Medium," and "High" groups. The TFR ranges for the disaggregation are as follows:

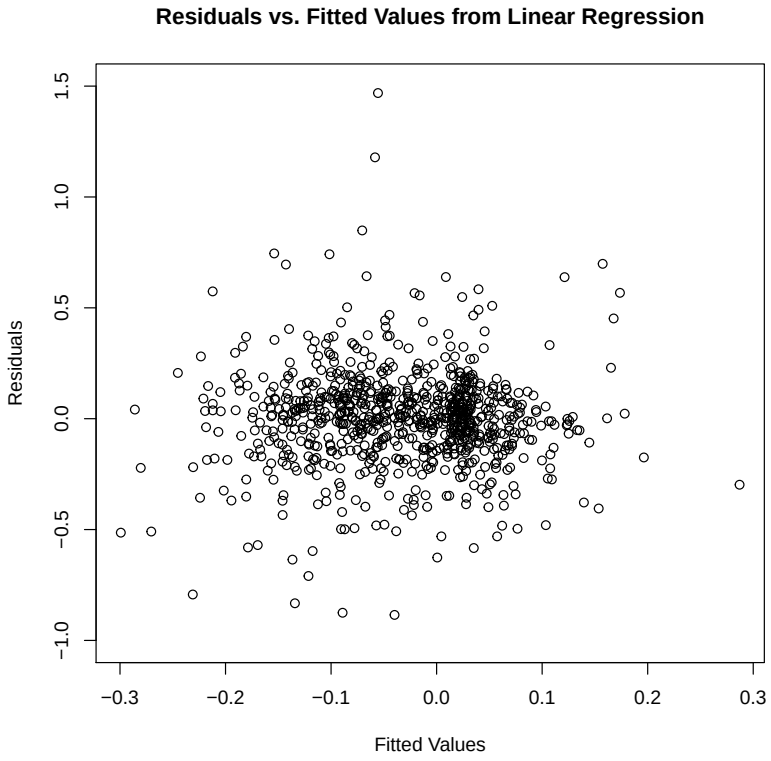


Figure B.4: Residuals from linear regression for outcome $Y_{c,t+1} = \Delta f_{c,t+1} + g(f_{c,t}|\boldsymbol{\theta}_c)$

1. Lowest (18 countries): TFR < 2.1
2. Low (27 countries): $2.1 \leq$ TFR < 2.75
3. Medium (25 countries): $2.75 \leq$ TFR < 4.6
4. High (27 countries): TFR ≥ 4.6

A summary of the OOS validation results disaggregated by TFR group is in Table B.3.

Table B.3: Out-of-sample validation results disaggregated by TFR group

Model	Validation Type	TFR Group	RMSE	95% Cvg	95% Width
Conditional Model	OOS	Lowest	0.0999	0.9444	0.5801
Conditional Model	OOS	Low	0.0658	1.0000	0.8006
Conditional Model	OOS	Medium	0.1887	0.9200	0.9500
Conditional Model	OOS	High	0.1039	1.0000	0.9215
Conditional Model	Conditional OOS	Lowest	0.0999	0.9444	0.5801
Conditional Model	Conditional OOS	Low	0.0800	1.0000	0.7594
Conditional Model	Conditional OOS	Medium	0.1832	0.9200	0.9121
Conditional Model	Conditional OOS	High	0.0856	1.0000	0.8953
bayesTFR		Lowest	0.0999	0.9444	0.5801
bayesTFR		Low	0.0697	1.0000	0.7588
bayesTFR		Medium	0.1890	0.9200	0.9159
bayesTFR		High	0.0916	1.0000	0.8972

We see very similar performance across models. Performance for the "Lowest" TFR group is identical across models, which follows expectations since the conditional model projects these countries following the same methodology as bayesTFR. Across all validation exercises, RMSE is the largest for the "Medium" TFR group. The mean width of the 95% intervals is also largest for this group. The "High" TFR group has the second largest RMSE and the second widest mean interval widths within each validation exercise, while the "Low" TFR group has the smallest RMSE within each validation exercise. Overall, the out-of-sample results disaggregated by TFR level indicate the conditional TFR projection model performs very similarly to the bayesTFR model.

We also checked the out-of-sample validation results disaggregated by different levels of educational attainment and contraceptive prevalence in the last time period used for estimation (2015–2020). We created groups based on levels of attainment and contraceptive prevalence rather than the changes over time in attainment and contraceptive prevalence. In other words, groups were based on X rather than ΔX.

For attainment, we created three equally sized groups based on the level of the proportion

of women who attained at least Lower Secondary education in 2010–2015. These groups were defined as follows:

1. Low (33 countries): attainment ≤ 0.345
2. Medium (32 countries): $0.345 <$ attainment ≤ 0.704
3. High (32 countries): attainment > 0.704

The results of the out-of-sample validation disaggregated by attainment level is summarized in Table B.4.

Table B.4: Out-of-sample validation results disaggregated by attainment group

Model	Validation Type	Attainment Group	RMSE	95% Cvg	95% Width
Conditional Model	OOS	Low	0.1186	1.0000	0.9359
Conditional Model	OOS	Medium	0.1043	1.0000	0.8285
Conditional Model	OOS	High	0.1443	0.9062	0.7279
Conditional Model	Conditional OOS	Low	0.0999	1.0000	0.9035
Conditional Model	Conditional OOS	Medium	0.1107	1.0000	0.7959
Conditional Model	Conditional OOS	High	0.1444	0.9062	0.7074
bayesTFR		Low	0.1062	1.0000	0.9080
bayesTFR		Medium	0.1074	1.0000	0.7946
bayesTFR		High	0.1468	0.9062	0.7081

The out-of-sample performance for the different attainment groups is similar across validation exercises. Within each validation exercise, the "Low" and "Medium" attainment groups tend to have similar performance with one another while the "High" attainment group has the worst performance in terms of RMSE and coverage of the 95% intervals.

For contraceptive prevalence, we created three equally sized groups based on the level of contraceptive prevalence of modern methods in 2010–2015. These groups were defined as follows:

1. Low (33 countries): CP ≤ 0.22

2. Medium (32 countries): $0.22 < \text{CP} \leq 0.37$
3. High (32 countries): $\text{CP} > 0.37$

The results of the out-of-sample validation disaggregated by contraceptive prevalence level is summarized in Table B.5.

Table B.5: Out-of-sample validation results disaggregated by contraceptive prevalence group

Model	Validation Type	CP Group	RMSE	95% Cvg	95% Width
Conditional Model	OOS	Low	0.0917	1.0000	0.8740
Conditional Model	OOS	Medium	0.1650	0.9375	0.8636
Conditional Model	OOS	High	0.1014	0.9688	0.7566
Conditional Model	Conditional OOS	Low	0.0743	1.0000	0.8508
Conditional Model	Conditional OOS	Medium	0.1683	0.9375	0.8298
Conditional Model	Conditional OOS	High	0.0969	0.9688	0.7279
bayesTFR		Low	0.0796	1.0000	0.8535
bayesTFR		Medium	0.1688	0.9375	0.8312
bayesTFR		High	0.0984	0.9688	0.7277

The out-of-sample performance for the difference contraceptive prevalence groups is similar across validation exercises, with the worst performance in terms of RMSE and coverage of 95% intervals occurring for countries in the "Medium" group for all exercises. The "Low" group had the best performance in terms of RMSE within each validation exercise. Based on the out-of-sample results disaggregated by covariate levels, the conditional model performs very similarly to the bayesTFR model.

B.5 Population Projection Results for Above Replacement Countries Aggregate

We present population projection results for the regional aggregate of all countries where we created intervention-based projections using the conditional TFR projection model. These are the 83 countries shown in blue in Figure B.5. All of these countries have available

covariate data and have at or above replacement level fertility in 2015-2020, where replacement level TFR is estimated at 2.1 children per woman. We refer to this aggregate as the "Above Replacement Countries" aggregate, although we do still include Kuwait, which has estimated TFR of exactly 2.1 in 2015-2020, in the aggregate.

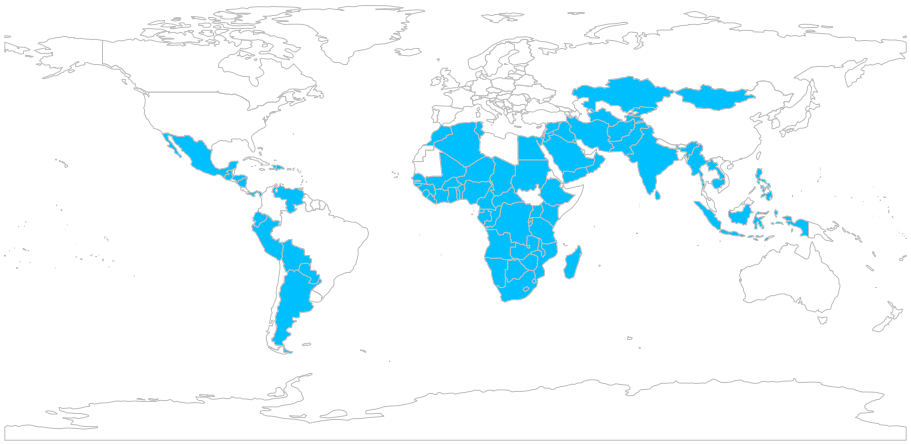


Figure B.5: TFR projections are created using the conditional TFR projection model for the 83 countries highlighted in blue.

Population projection trajectories for all 83 countries are aggregated to produce the distribution of population projections for the Above Replacement Countries aggregate. Figure B.6 shows median projections of population size for the Above Replacement Countries aggregate for all projection scenarios alongside 95% PIs for the reference and Both SDGs (0% Unmet) scenarios. Median and 95% PIs for projected values of population size for the aggregate for all scenarios and differences in median projected population size between scenarios are shown in Table B.6 for 2035, 2050, and 2100.

In 2100, population size for the Above Replacement Countries aggregate is projected to reach 7.86 (6.53, 9.61) billion people in the reference scenario, 6.66 (5.45, 8.09) billion people in the Both SDGs (0% Unmet Need) scenario, 6.87 (5.64, 8.44) billion people in the Both

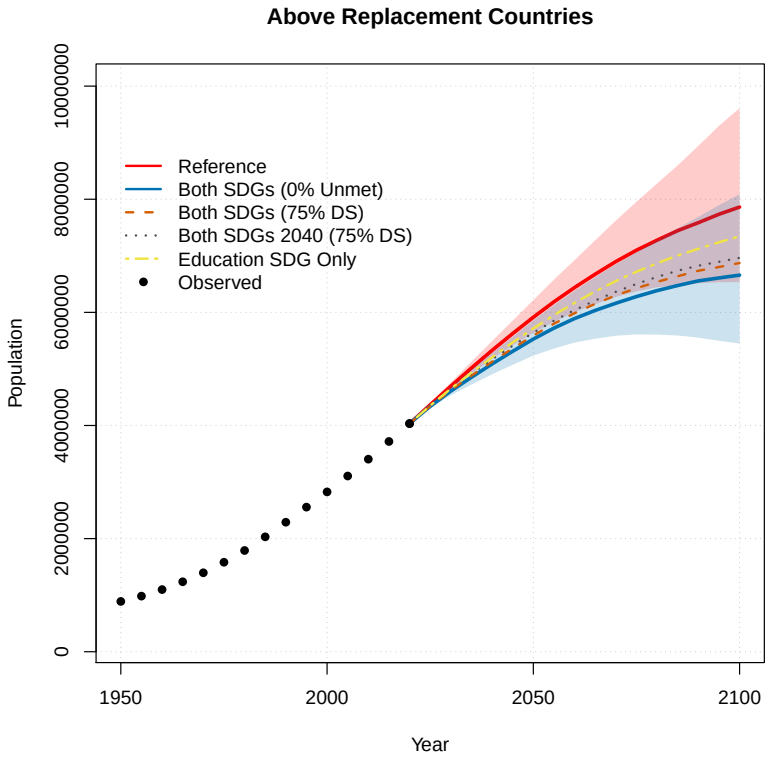


Figure B.6: Comparison of median population projections for the Above Replacement Countries aggregate from reference scenario in red, Both SDGs (0% Unmet) in dark blue, Both SDGs (75% DS) in orange dashed, Both SDGs 2040 (75% DS) in dark grey dotted, and Education SDG Only in yellow dash-dotted. 95% projection intervals for the reference and Both SDGs (0% Unmet) scenarios are also plotted.

SDGs (75% DS) scenario, 6.96 (5.76, 8.46) billion people in the Both SDGs 2040 (75% DS) scenario, and 7.34 (6.07, 8.88) billion people in the Education SDG Only scenario. The median population projection in the Both SDGs (0% Unmet) scenario is 1.20 billion people lower than in the reference scenario. If the SDGs correspond to attaining 75% Demand Satisfied instead, this difference is 989 million people. If the SDGs correspond to attaining 75% Demand Satisfied and were met in 2040 instead of 2030, this difference is 898 million people, indicating that policies focused on meeting the SDGs a decade later in countries that

Table B.6: Median projections in 2035, 2050, and 2100 for population size of the Above Replacement Countries aggregate in millions of people for all projection scenarios with 95% PIs. Row indicating differences between projection scenarios show differences between median projected population size.

	2035	2050	2100
Reference	5016 (4911, 5128)	5914 (5639, 6217)	7862 (6534, 9611)
Both SDGs (0% Unmet)	4853 (4723, 4984)	5528 (5234, 5857)	6658 (5450, 8087)
Both SDGs (75% DS)	4881 (4758, 5008)	5591 (5310, 5922)	6873 (5639, 8436)
Both SDGs 2040	4914 (4792, 5040)	5639 (5344, 5943)	6964 (5764, 8456)
Educ SDG Only	4933 (4819, 5056)	5715 (5427, 6033)	7344 (6065, 8877)
Reference − Both SDGs (0% Unmet)	163	386	1204
Reference − Both SDGs (75% DS)	135	323	989
Reference − Both SDGS 2040	102	274	898
Reference − Educ SDG Only	83	199	518
Both SDGs (75% DS) − Both SDGs (0% Unmet)	28	63	215
Both SDGs 2040 − Both SDGs (75% DS)	33	49	91
Educ SDG Only − Both SDGs (0% Unmet)	80	187	686
Educ SDG Only − Both SDGs (75% DS)	52	124	471

are not currently on track to meet the SDGs in 2030 could still have a substantial impact on population size for the Above Replacement Countries aggregate.

If only the target of universal lower secondary education is attained with no additional intervention targeting family planning, population is projected to be 518 million people lower than the reference scenario in 2100. The additional effect of attaining the SDG target corresponding to family planning is a reduction of 686 million people in 2100 if Target 3.7 is interpreted as 0% Unmet Need. If the target is instead interpreted as 75% Demand Satisfied, the additional reduction is 471 million people.

B.6 Additional Comparisons with Related Work

Additional details of the comparison of our results with those of Abel et al. (2016) and Vollset et al. (2020) are provided in this section. We compare population size projections

under reference and SDG intervention scenarios from our model, the Abel et al. model, and the Vollset et al. model for the regional aggregate of a subset of 72 out of the 83 countries in the Above Replacement Countries aggregate, where the Above Replacement Countries aggregate was defined in Section B.5. Eleven countries from the Above Replacement Countries aggregate were excluded from the comparison due to lack of available projections for the Abel et al. model. These countries were Afghanistan, Angola, Bolivia, Botswana, Côte d'Ivoire, Israel, Oman, Sri Lanka, Sudan, Togo, and Yemen.

B.6.1 Approximation for population of the regional aggregate

The conditional TFR projection model outputs posterior trajectories of projected TFR for each policy intervention scenario. These trajectories of projected TFR are then combined with trajectories of projected life expectancy and net migration to create trajectories of projected population size. For regional aggregates, each trajectory in the distribution of projected population size is constructed as the sum of the projected population trajectories for all countries within the regional aggregate. This method was used to create the population projection distributions from the proposed conditional TFR projection model for the sub-Saharan Africa aggregate and the Above Replacement Countries aggregate for each projection scenario.

We are unable to obtain the full distribution of projected population for regional aggregates from Abel et al. (2016) and Vollset et al. (2020) as Abel et al. and Vollset et al. do not publish individual projection trajectories. Thus, we cannot directly compare the median projected population size of the subset of the Above Replacement Countries aggregate from our conditional projection model, the Abel et al. model, and the Vollset et al. model. We instead compare an approximation to the median projected population size of the regional aggregate, where the approximation is constructed as the sum of the median population projections for all countries in the aggregate.

This "sum of medians" approximation is not identical to the median of the projected population distribution of the regional aggregate, but for our conditional projection results we found the sum of medians was a good approximation for most time periods. Figure B.7 shows a comparison of the sum of median population projections for all countries in the subset of the Above Replacement Countries aggregate and the median of the projected population distribution for the regional aggregate from the reference scenario for our conditional projection model. From 2025 through 2075, the sum of medians is a close approximation to the true median population projection for the aggregate. The difference between the approximation and the true median population projection gets larger over time after 2075. At 2100, the difference between the median of the projected population distribution and the sum of medians approximation is 152,746,269 people. While this is a large number of people, this difference corresponds to the sum of medians approximation being only around 2.1% lower than the median of the projected population distribution for the subset of the Above Replacement Countries aggregate.

B.6.2 Comparison for all projection years

We focus our comparisons on the projected differences between the reference scenario and the SDG intervention scenario from our model, the Abel et al. model, and the Vollset et al. model for the regional aggregate of 72 countries. These projected differences are plotted in Figure B.8 for all projection years in 2020–2100. The projected differences from our conditional projection model are shown for the Both SDGs (0% Unmet) and Both SDGs (75% DS) projection scenarios, where Both SDGs (0% Unmet) aligns most closely with the SDG assumptions used by Abel et al. and Vollset et al.. Results from Both SDGs (75% DS) are included in the comparison to illustrate a more realistic interpretation of meeting the SDGs. We note that our conditional projection model and the Abel et al. model project population in five-year increments, while the Vollset et al. model projects population on the

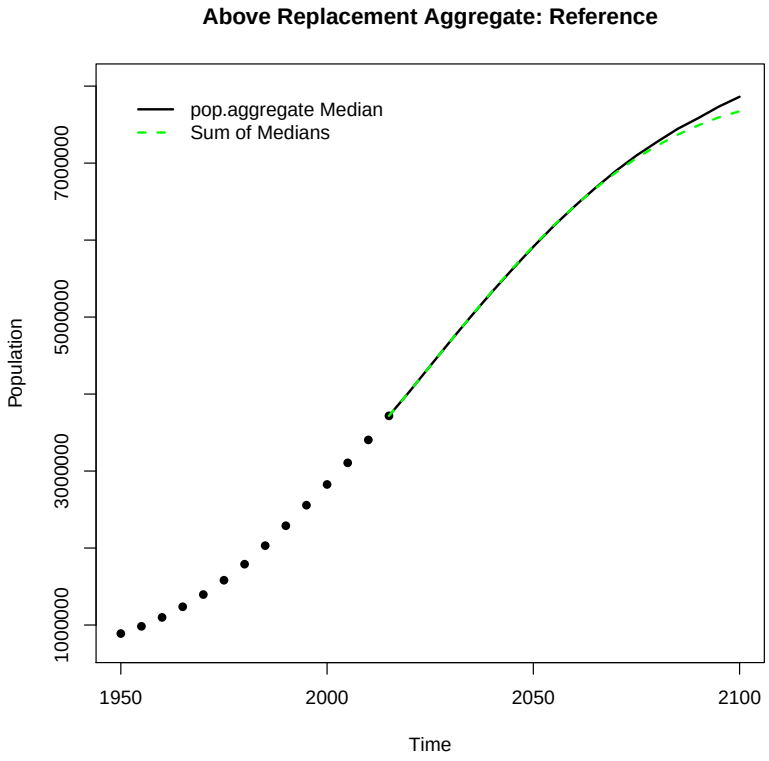


Figure B.7: Comparison of median population projections for the subset of the Above Replacement Countries aggregate from projected population distribution (solid black line) and from summing the median projections for each country in the aggregate (dashed green line) for the reference scenario from our conditional projection model

annual scale.

We first consider comparisons between our results and the Abel et al. (2016) results. Compared to Abel et al., we project larger reductions in population size in 2100 due to attaining both SDGs in 2030. Under SDG2, Abel et al. project the population of the regional aggregate to be 4.657 billion people in 2100, which is a reduction of 755 million people compared to their reference scenario projection of 5.412 billion people. In contrast, we project a reduction from 7.010 billion people to 5.984 billion people between our reference

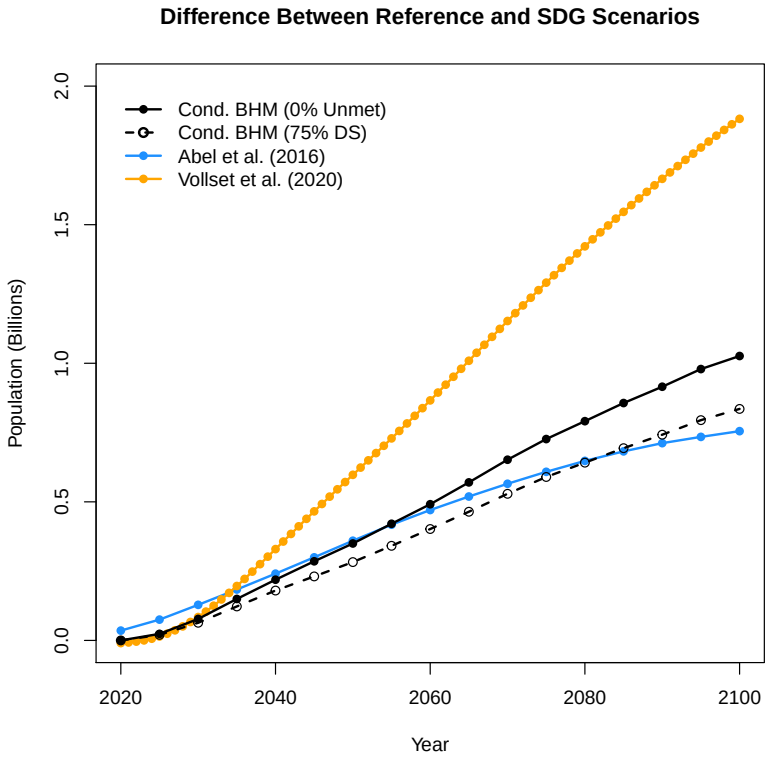


Figure B.8: Comparison of differences between reference scenario and SDG intervention scenario population projections for the regional aggregate of 72 countries from our conditional projection model (Cond. BHM), Abel et al. (2016), and Vollset et al. (2020) in billions of people. The SDG results from our conditional projection model follow the Both SDGs (0% Unmet) and Both SDGs (75% DS) scenarios. The SDG results from the Abel et al. model follow their SDG2 scenario.

and Both SDGs (0% Unmet) scenarios, which is a difference of 1.026 billion people. The differences in projection results between our model and the Abel et al. model are much smaller in 2050. Our Both SDGs (0% Unmet) scenario projects a reduction from 5.516 billion people to 5.166 billion people, which is a difference of 350 million people from meeting the SDGs. The Abel et al. model projects a very similar difference of 360 million people in 2050 from meeting the SDGs, with a reference scenario population of 5.022 billion people and an

SDG intervention scenario population of 4.662 billion people.

The projected reductions in population size from our Both SDGs (0% Unmet) scenario roughly agree with the projected reductions in population size from Abel et al. up until mid-century, at which point the projected reductions diverge. A large part of the differences in the projected effect of the SDGs after mid-century is due to the underlying differences between the fertility projection models used by the UN and the Wittgenstein Centre. Abel et al. project the population of the regional aggregate to peak slightly after mid-century in both the reference and SDG intervention scenarios, whereas our population projections continue to increase to 2100 in all scenarios.

Comparing the projected reductions in population size from our more realistic Both SDGs (75% DS) scenario with the Abel et al. SDG2 scenario, we see similar reductions for all projected time periods. The projected reductions in population in our Both SDGs (75% DS) scenario are slightly smaller than the projected reductions from Abel et al. up through 2080, at which point our Both SDGs (75% DS) scenario projects a larger reduction in population size. Differences between projected reductions from the two models are largest in magnitude in 2100. However, even in 2100, the differences projected from our Both SDGs (75% DS) scenario are fairly close to the Abel et al. results, with our model projecting a reduction of 835 million people and the Abel et al. model projecting a reduction of 755 million people.

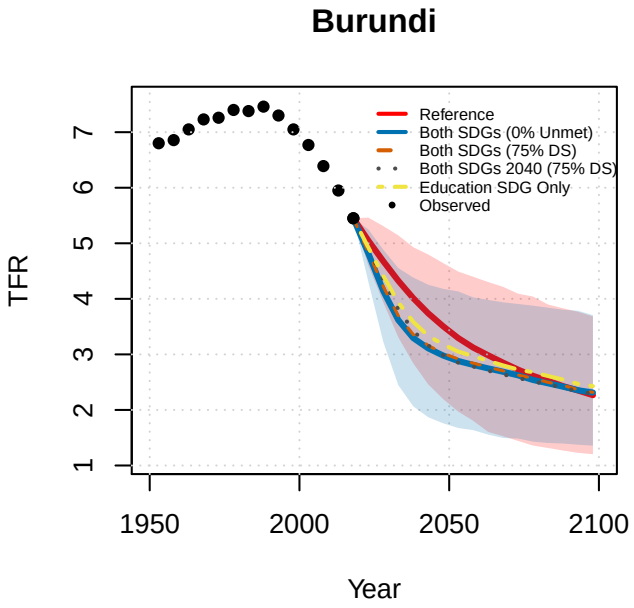
Next, we consider comparisons with Vollset et al. (2020). In 2050, Vollset et al. project population size for the regional aggregate to be 5.418 billion people in the reference scenario and 4.820 billion people in the SDG intervention scenario, which is a difference of 598 million people In our Both SDGs (0% Unmet) scenario, we project a smaller difference of 350 million people. In 2100, Vollset et al. projects population under the SDG scenario to be 3.730 billion people compared to their reference scenario projection of 5.612 billion people, which is a reduction in population of 1.882 billion people. In comparison, we project a smaller

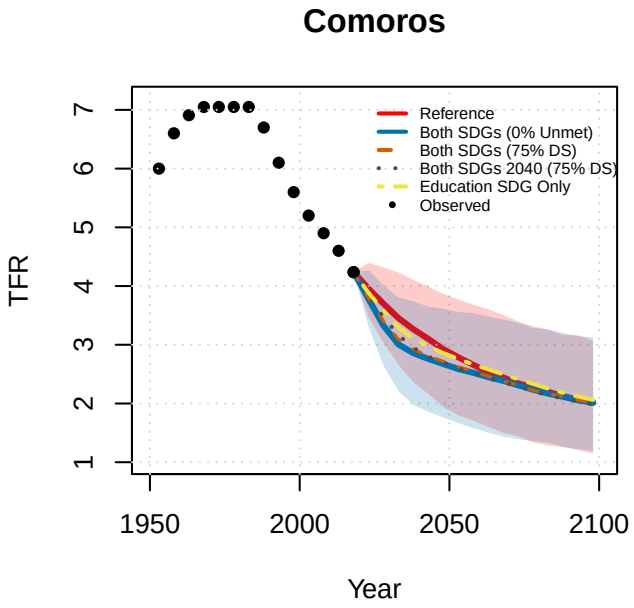
reduction of 1.026 billion people in our Both SDGs (0% Umet) scenario.

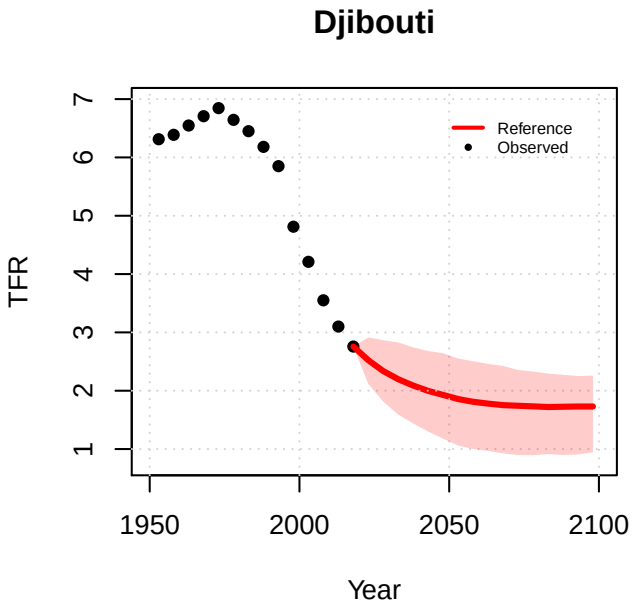
The projected reduction in population from Vollset et al. is similar to the projected reductions from both our model and the Abel et al. model from 2020 through around 2035, but very quickly begins to diverge as illustrated in Figure B.8. By 2100, the projected reduction in population from Vollset et al. is about 1.83 times as large as the projected reduction in population from our Both SDGs (0% Unmet) scenario and about 2.25 times as large as the projected reduction from our Both SDGs (75% DS) scenario. Comparing the Vollset et al. projections with the Abel et al. projections, Vollset et al. project the effect of meeting the SDGs for the regional aggregate to be about 1.66 times the effect projected by Abel et al. in 2050 and almost 2.5 times the effect projected by Abel et al. in 2100.

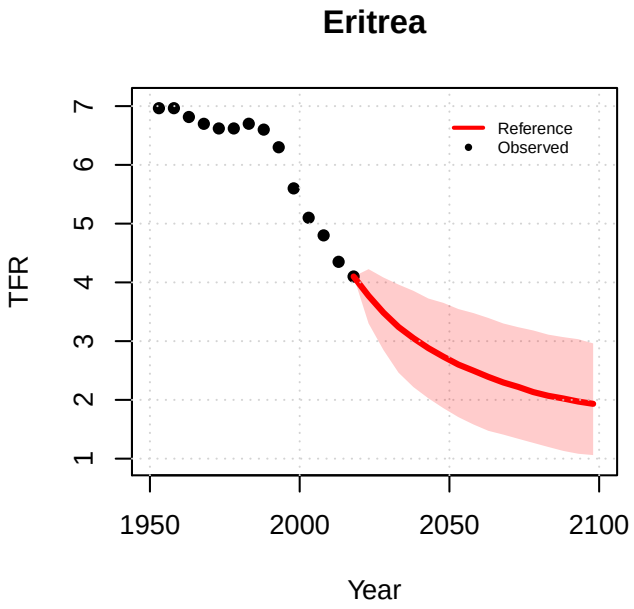
B.7 Results for All Countries

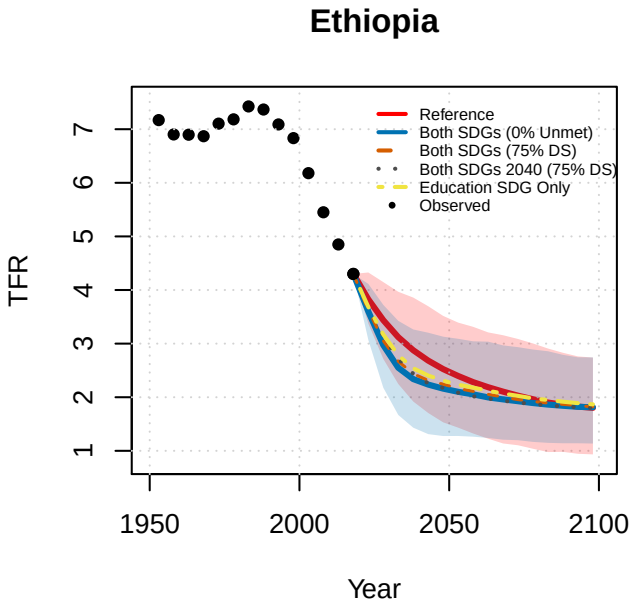
TFR projection results for 2020–2025 through 2095–2100 from all projection scenarios are illustrated for the 92 countries that are used to create the sub-Saharan Africa and Above Replacement Countries regional aggregates. For the 83 countries with TFR $\geq$ 2.1 in 2015–2020 that had available covariate data, the projection scenarios illustrated in the figures correspond to actual intervention-based projections of TFR that were created using the conditional TFR projection model. For countries with TFR less than 2.1 in 2015–2020 and countries that did not have available covariate data, TFR projections were only created for use in the sub-Saharan Africa regional aggregate. We do not consider the effect of policy interventions in these countries and instead use TFR projections from bayesTFR (version 6.4.0) for all projection scenarios. The nine countries where this applies are Djibouti, Eritrea, Mauritius, Mayotte, Réunion, the Seychelles, Somalia, South Sudan, and Mauritania.

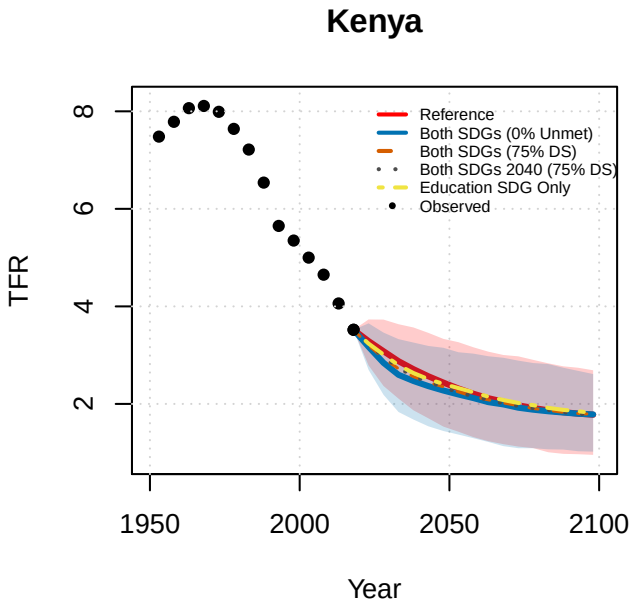

Burundi
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

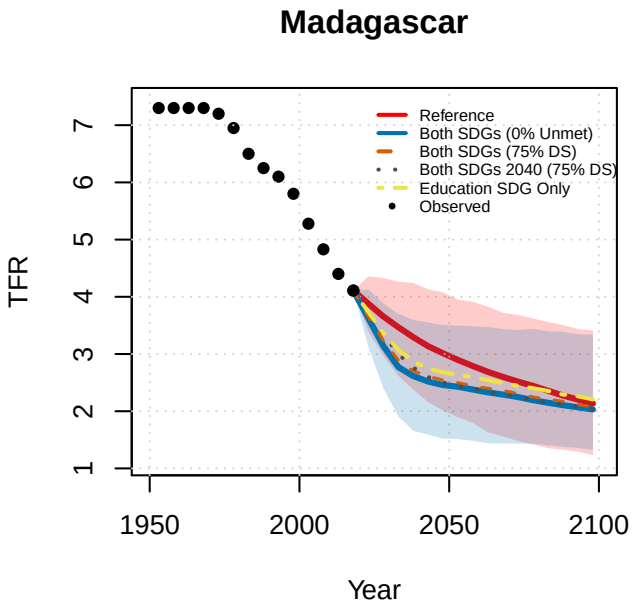

Comoros
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

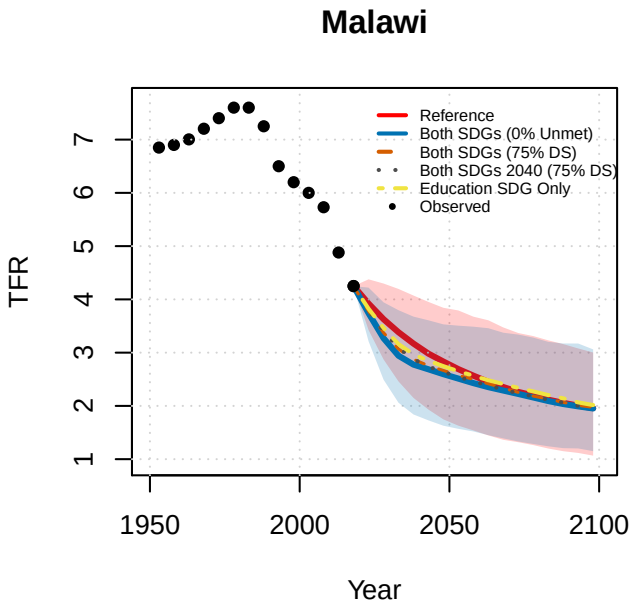

Djibouti
Reference
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

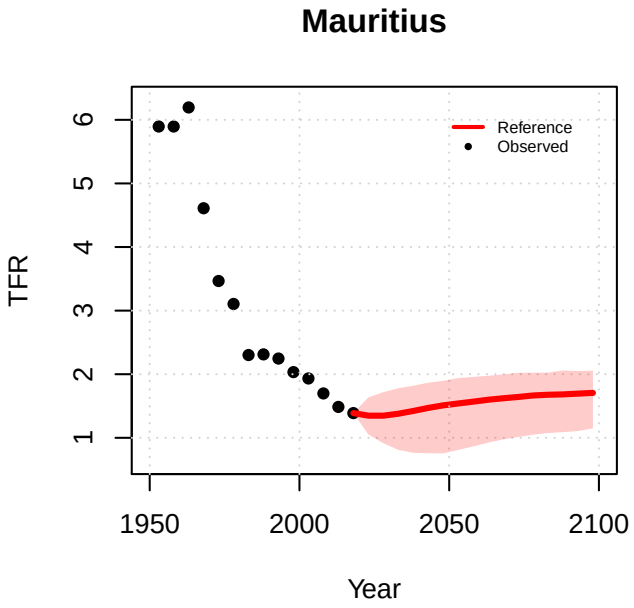

Eritrea
Reference
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

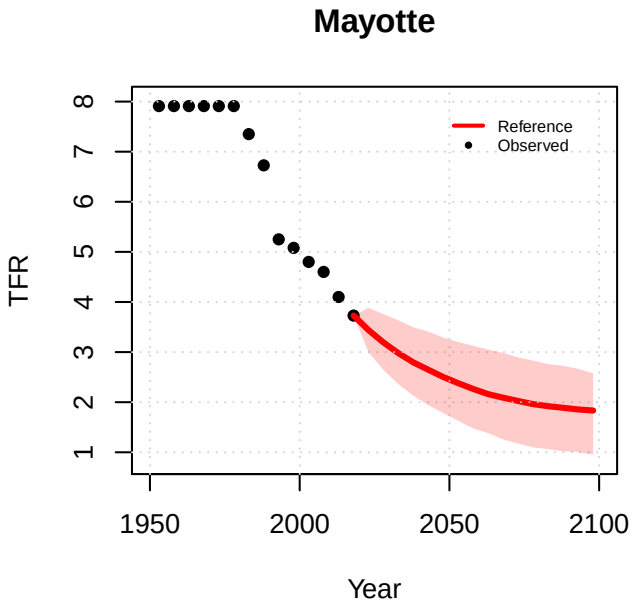

Ethiopia
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

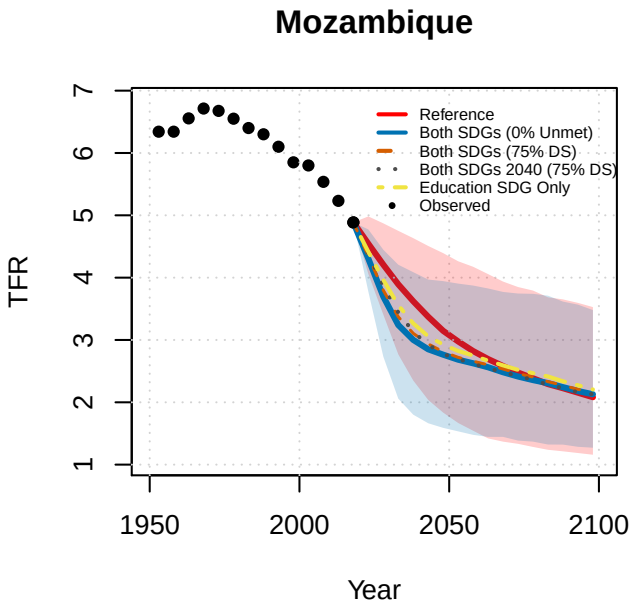

Kenya
TFR
Year
1950
2000
2050
2100
2
4
6
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

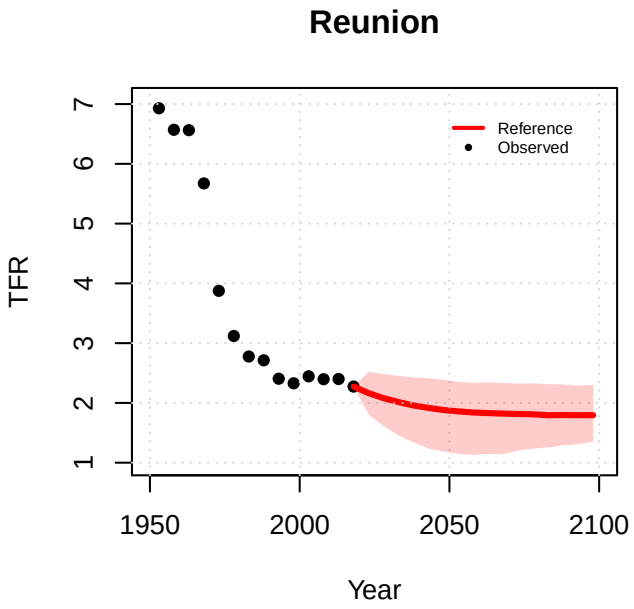

Madagascar
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

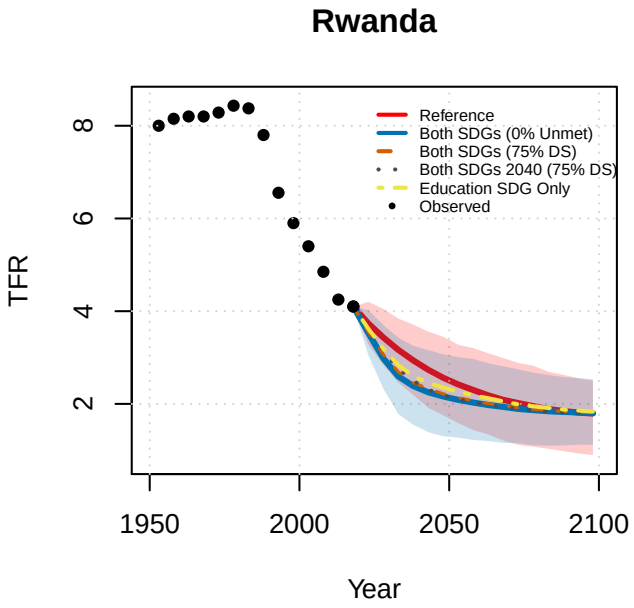

Malawi
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

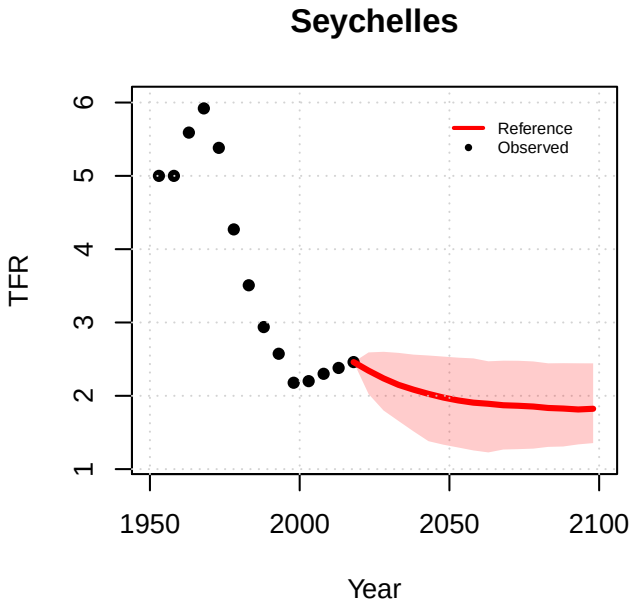

Mauritius
TFR
Year
Reference
Observed
1
2
3
4
5
6
1950
2000
2050
2100

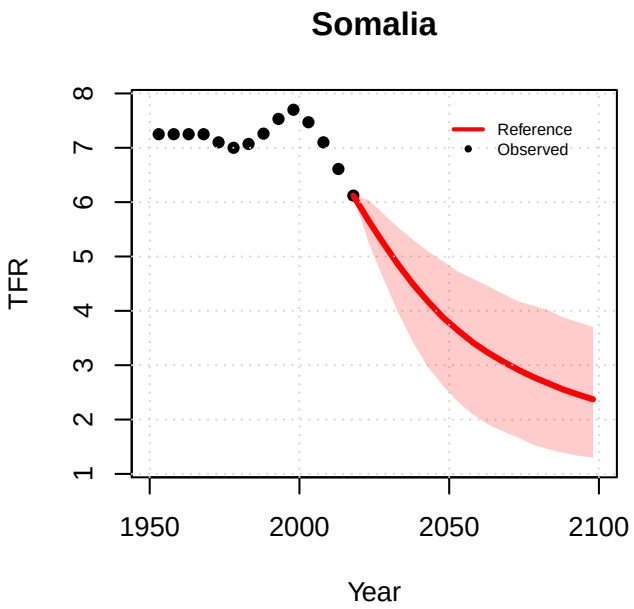

Mayotte
TFR
Year
Reference
Observed
1
2
3
4
5
6
7
8
1950
2000
2050
2100

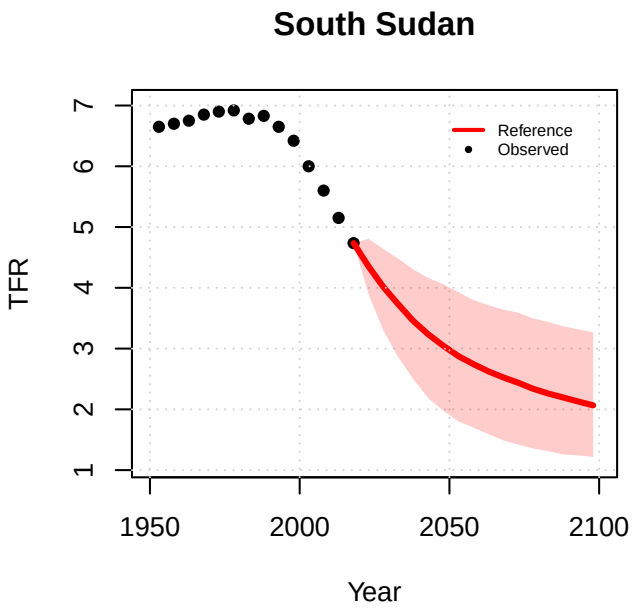

Mozambique
TFR
Year
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
1
2
3
4
5
6
7
1950
2000
2050
2100

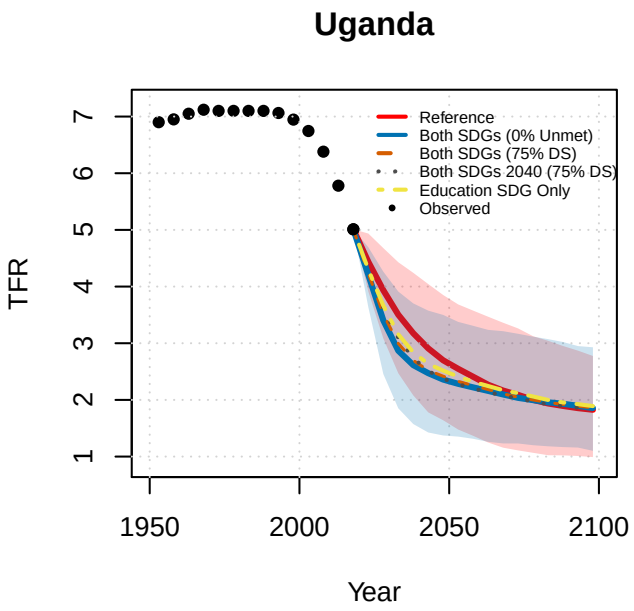

Reunion
TFR
Year
Reference
Observed
1
2
3
4
5
6
7
1950
2000
2050
2100

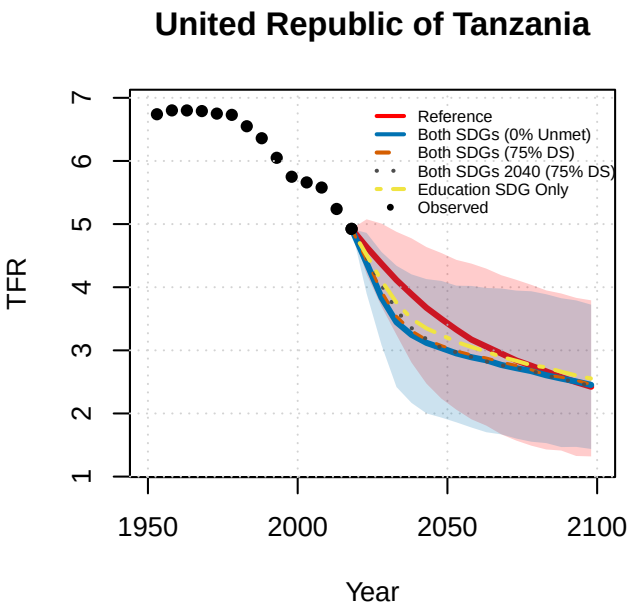

Rwanda
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
2
4
6
8

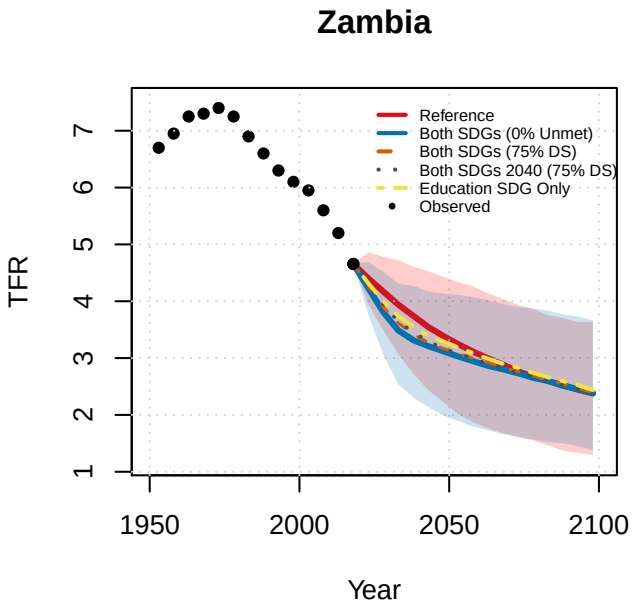

Seychelles
Reference
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

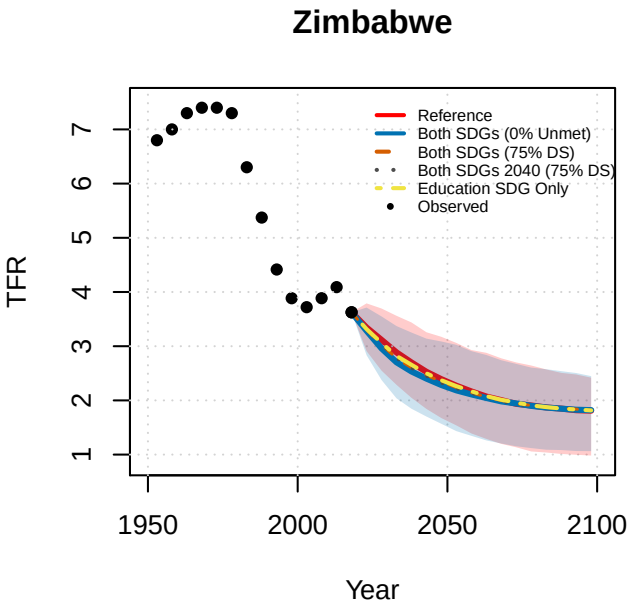

Somalia
Reference
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
8

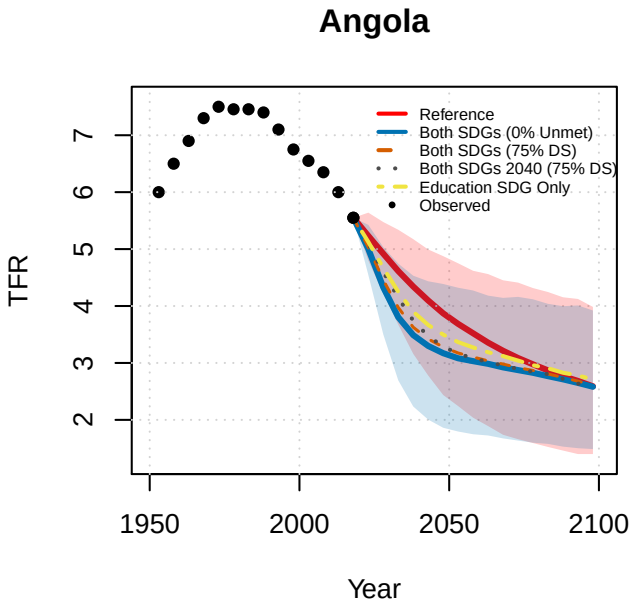

South Sudan
Reference
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

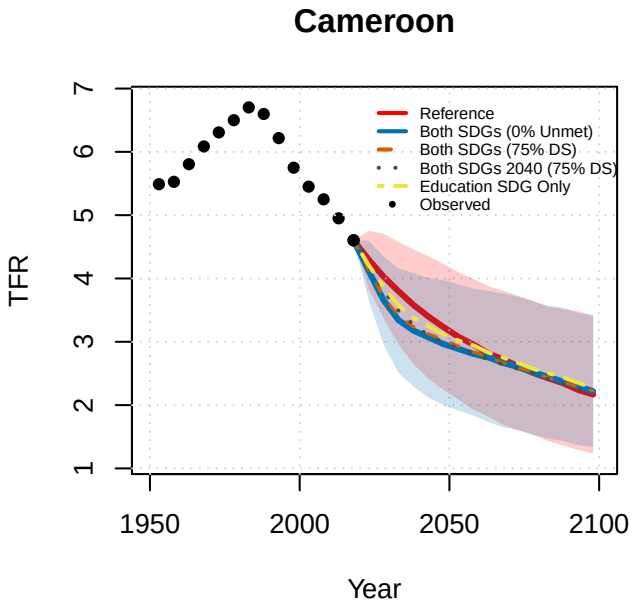

Uganda
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

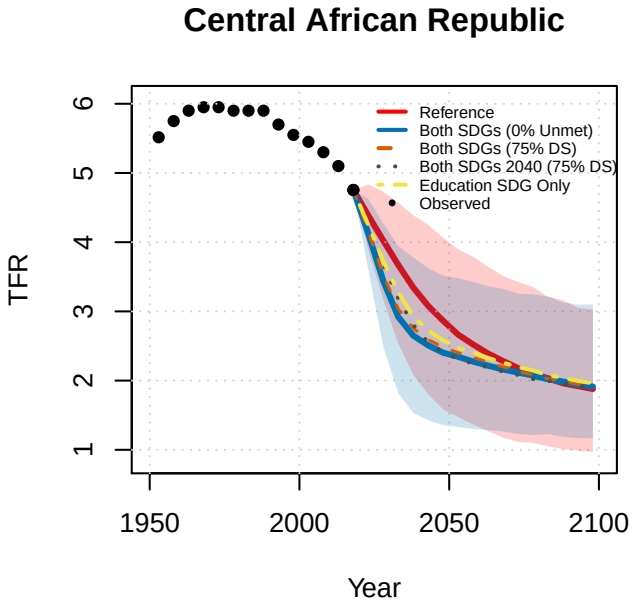

United Republic of Tanzania
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

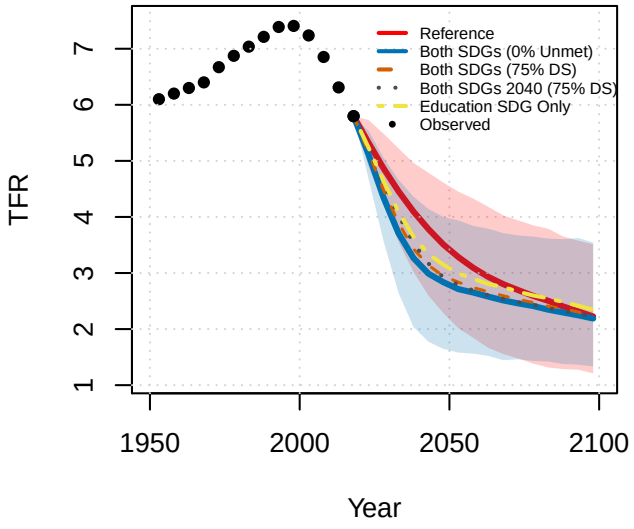

Zambia
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

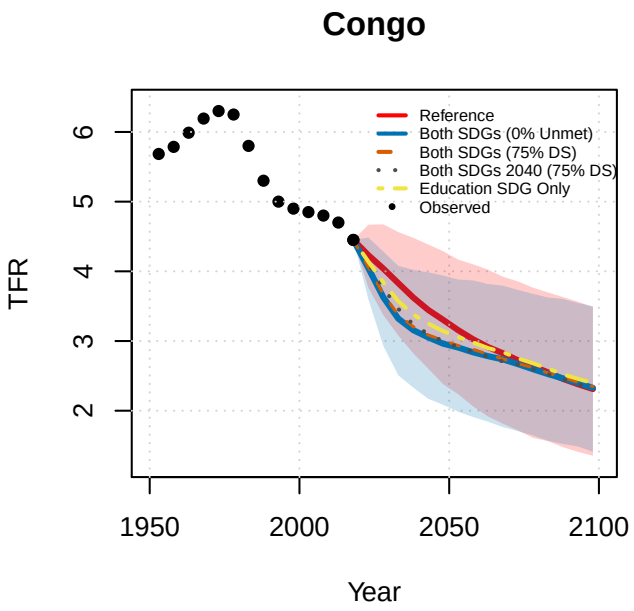

Zimbabwe
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

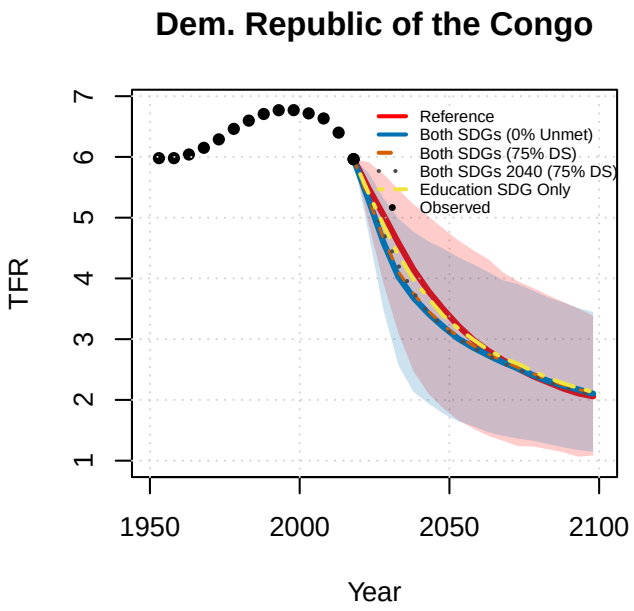

Angola
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
2
3
4
5
6
7

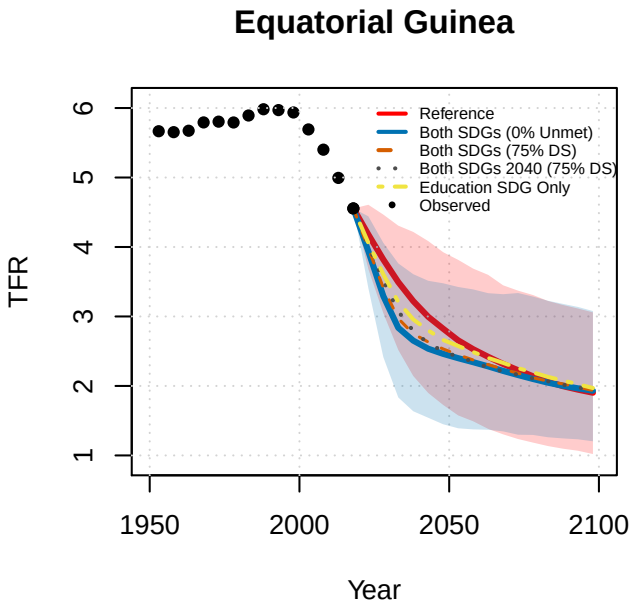

Cameroon
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

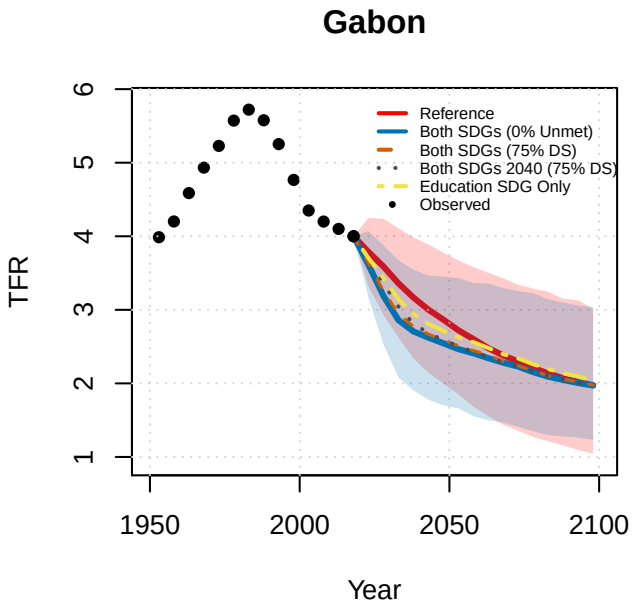

Central African Republic
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

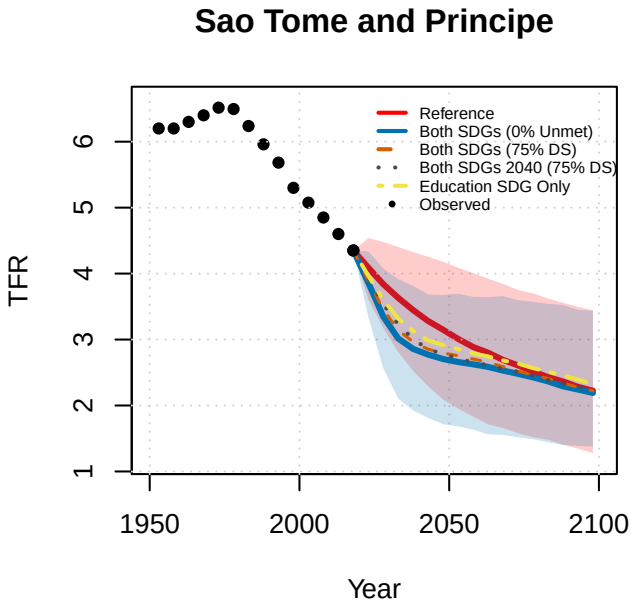

Chad
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

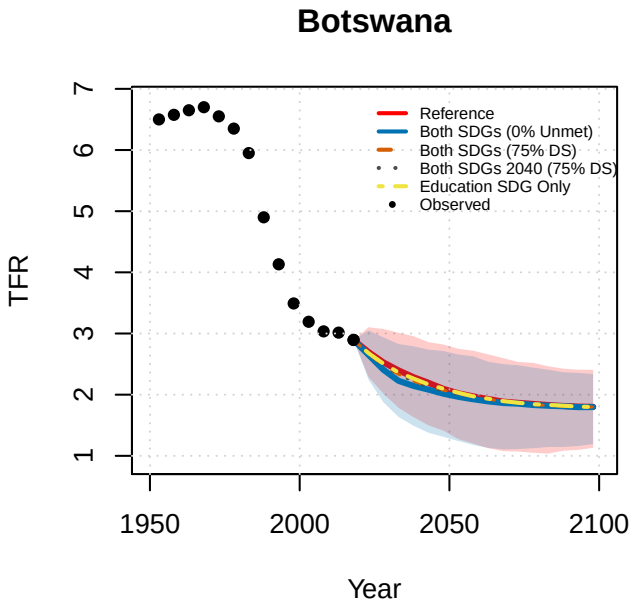

Congo
TFR
Year
1950
2000
2050
2100
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

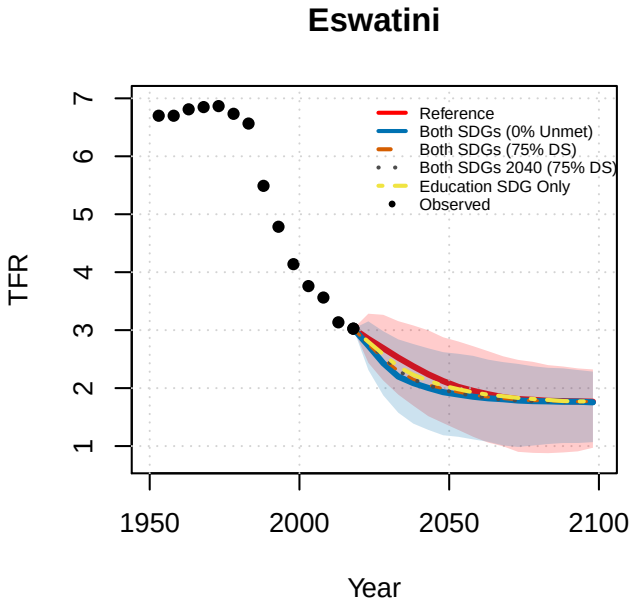

Dem. Republic of the Congo
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

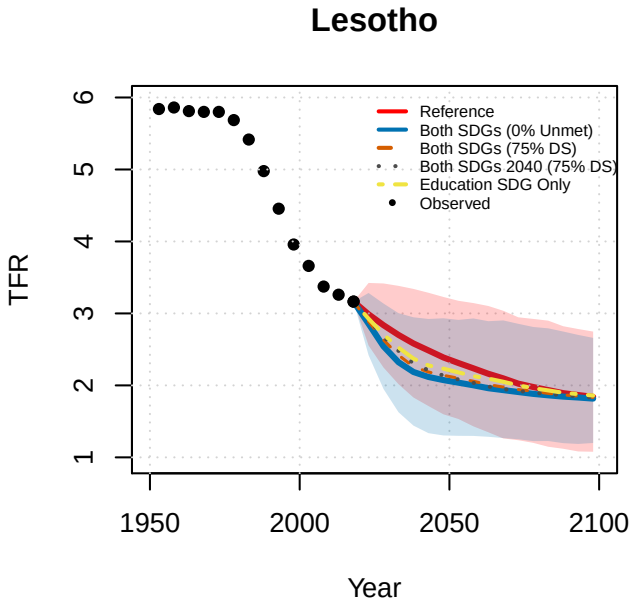

Equatorial Guinea
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

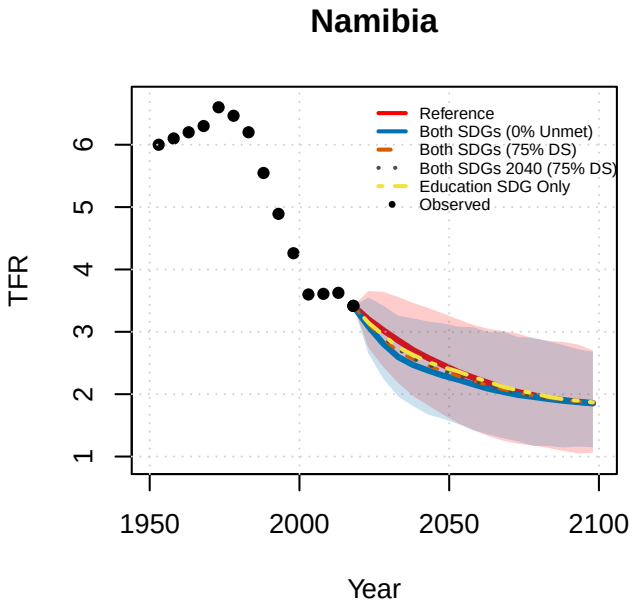

Gabon
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

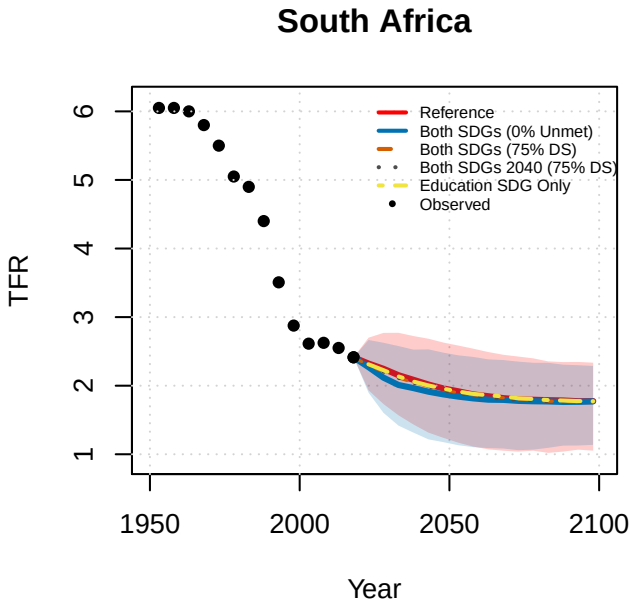

Sao Tome and Principe
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

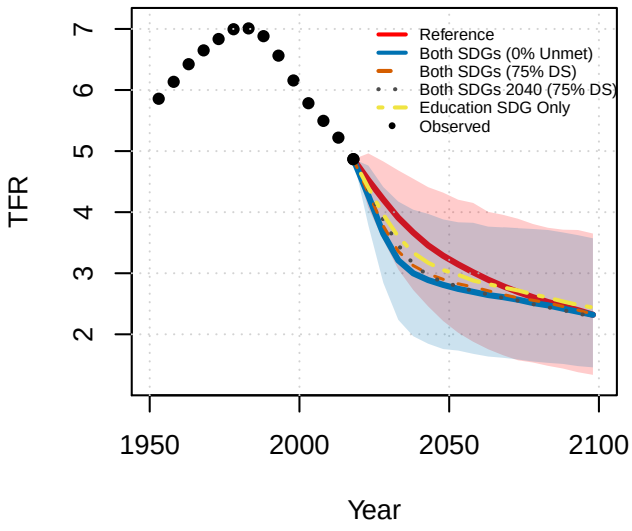

Botswana
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

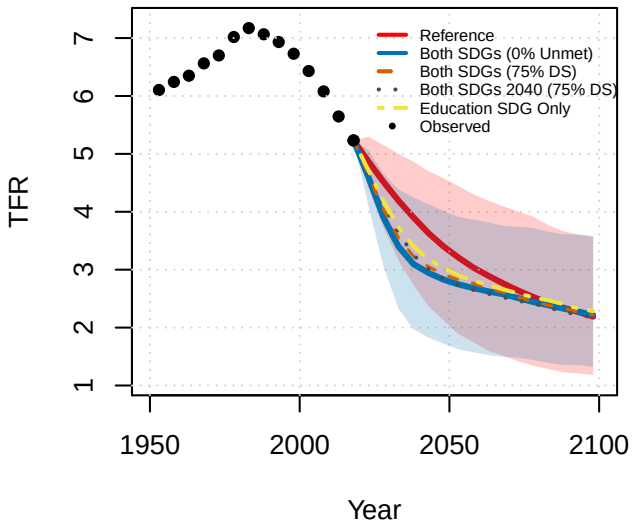

Eswatini
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

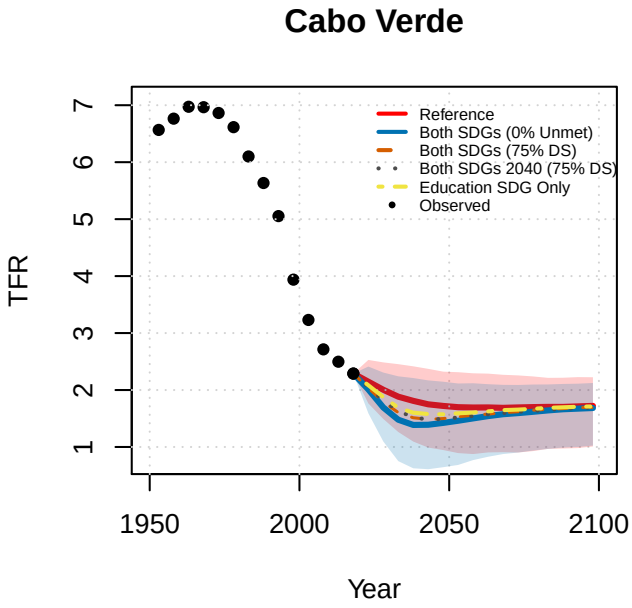

Lesotho
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

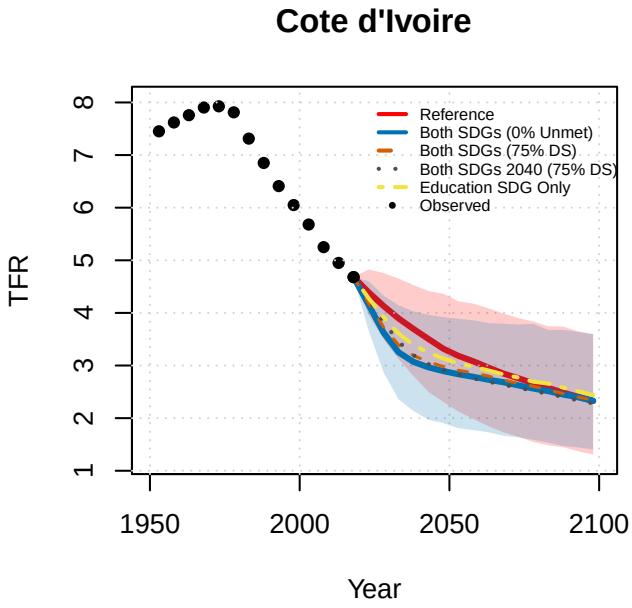

Namibia
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

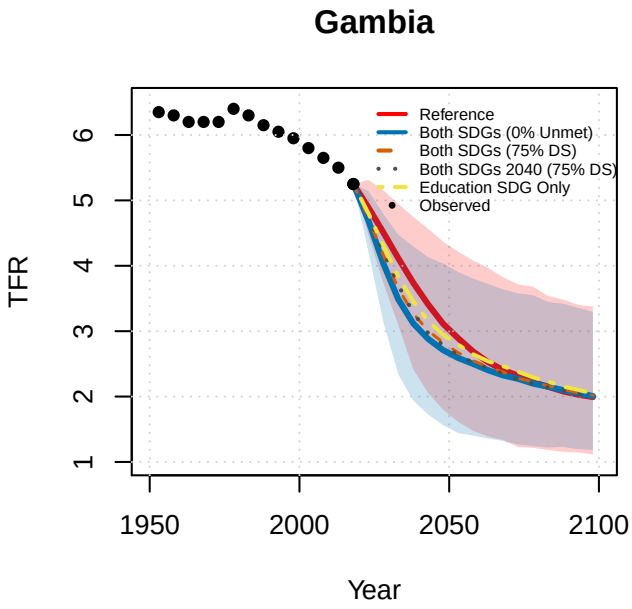

South Africa
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

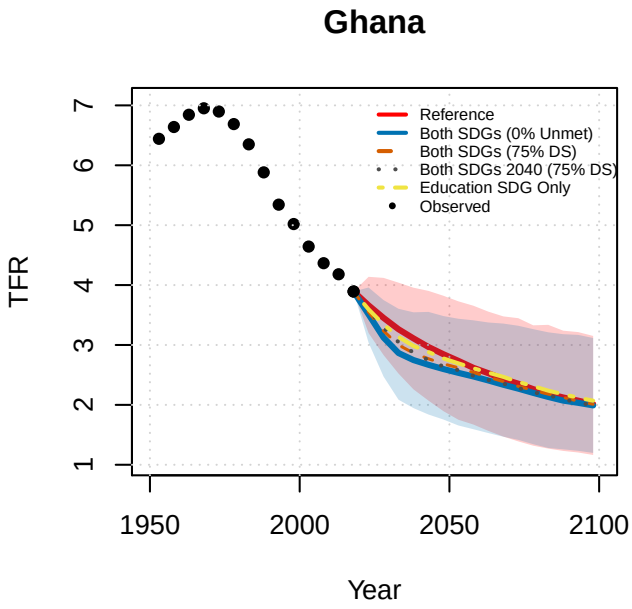

Benin
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
2
3
4
5
6
7

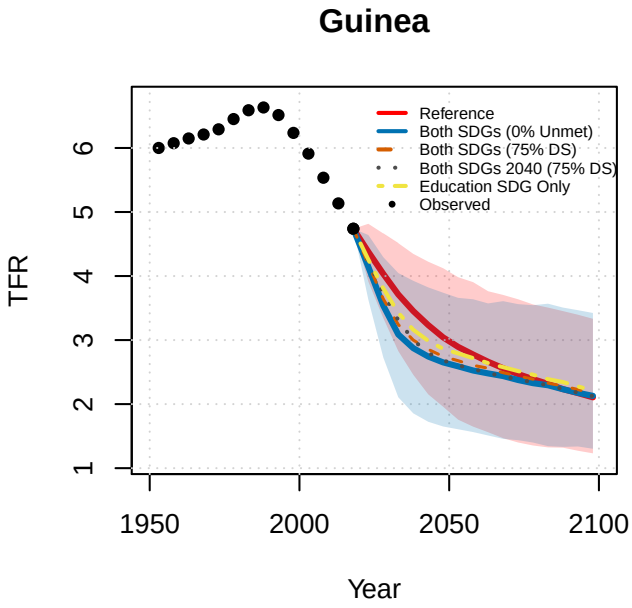

Burkina Faso
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

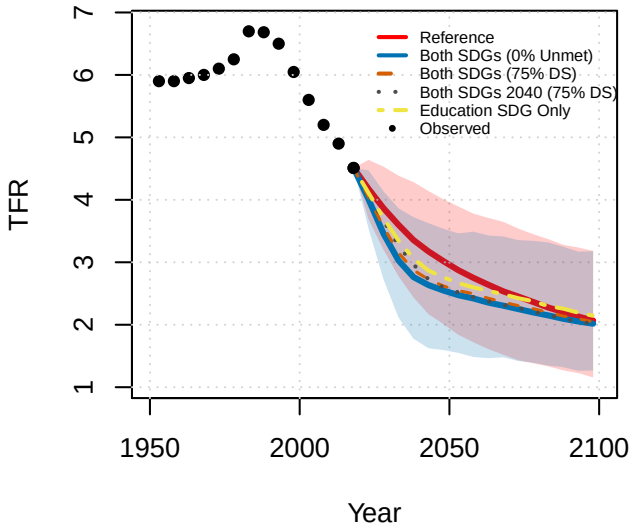

Cabo Verde
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

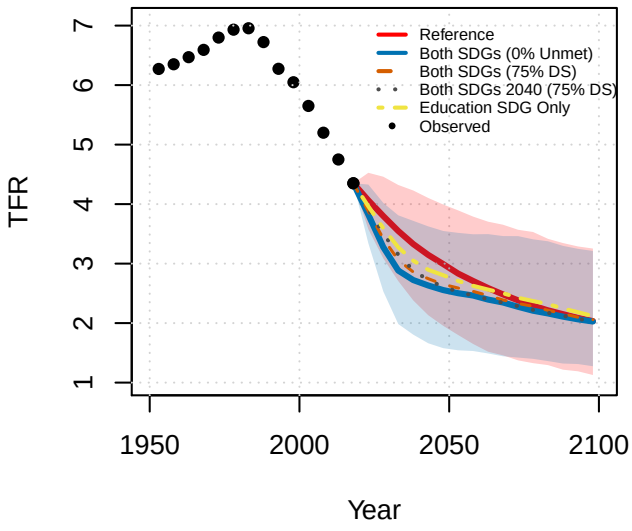

Cote d'Ivoire
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

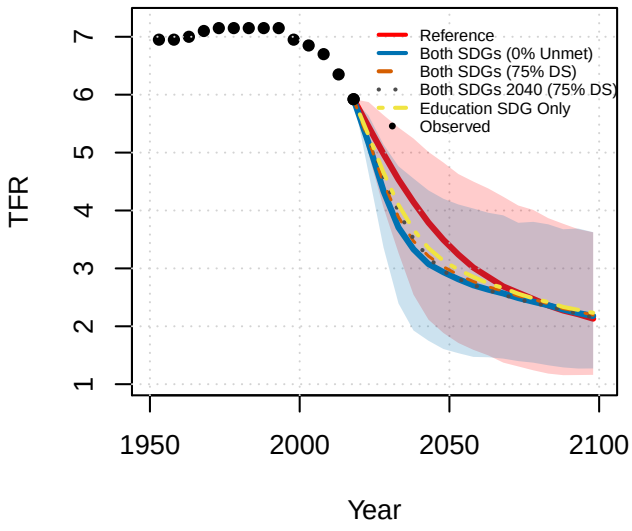

Gambia
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

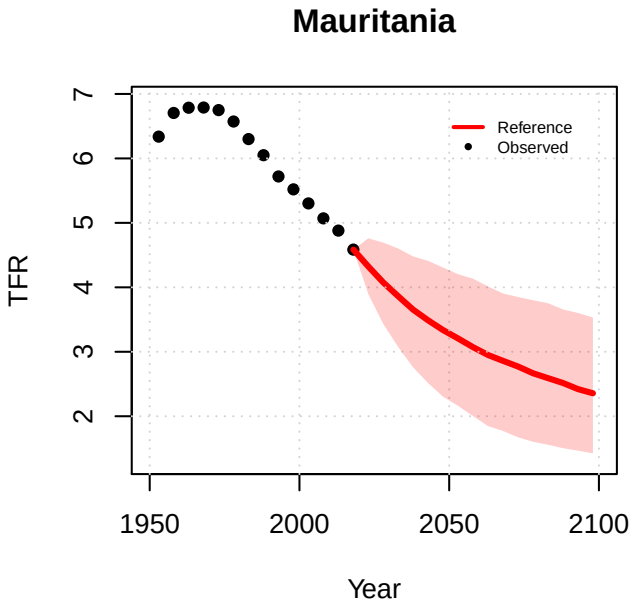

Ghana
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

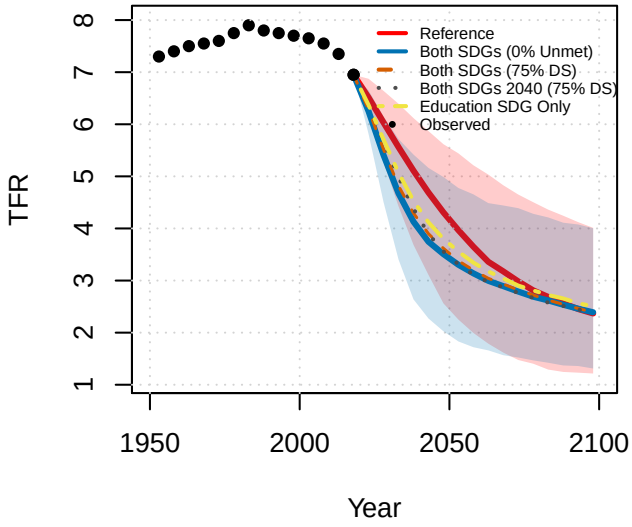

Guinea
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

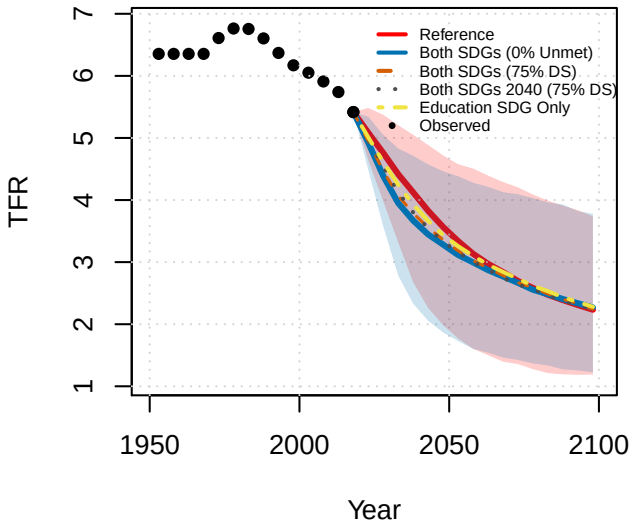

Guinea−Bissau
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

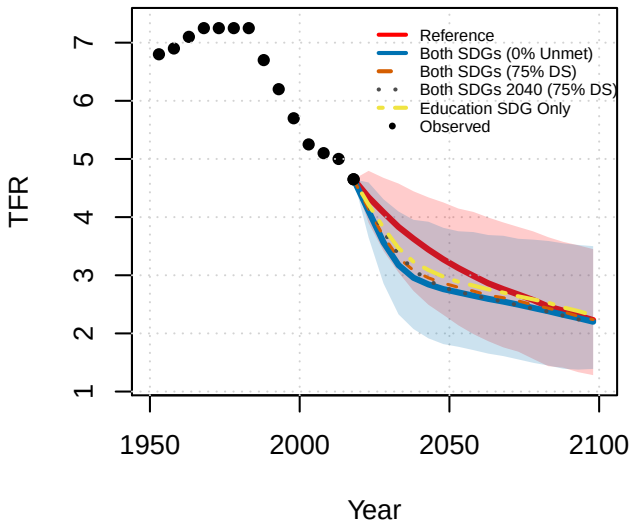

Liberia
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

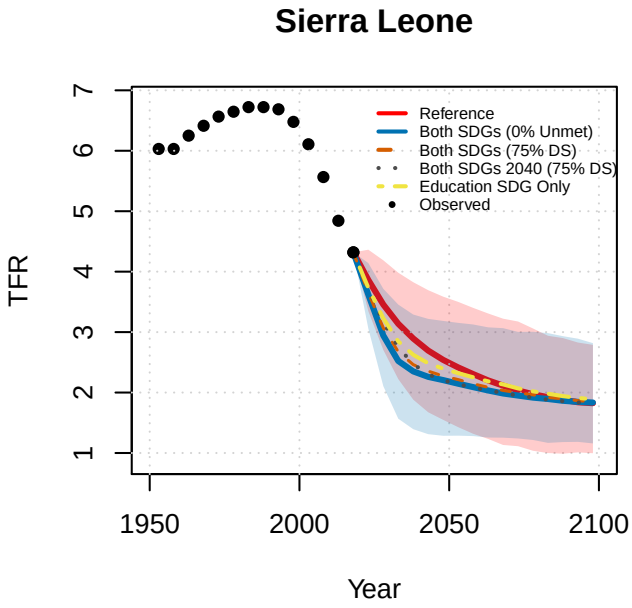

Mali
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

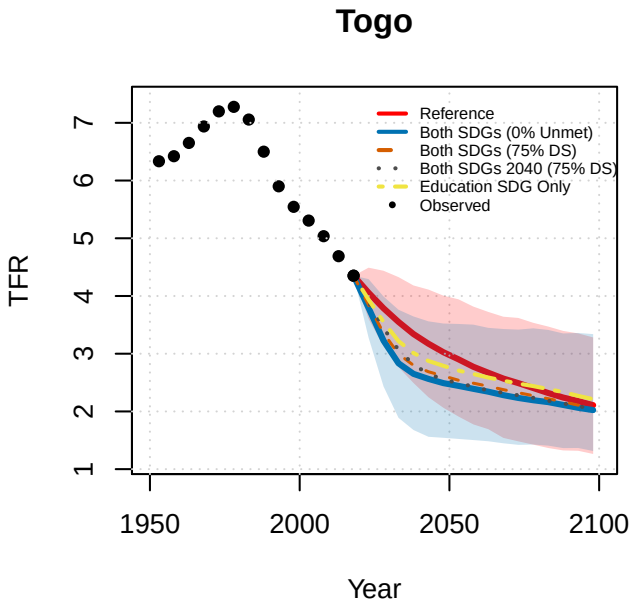

Mauritania
TFR
Year
1950
2000
2050
2100
2
3
4
5
6
7
Reference
Observed

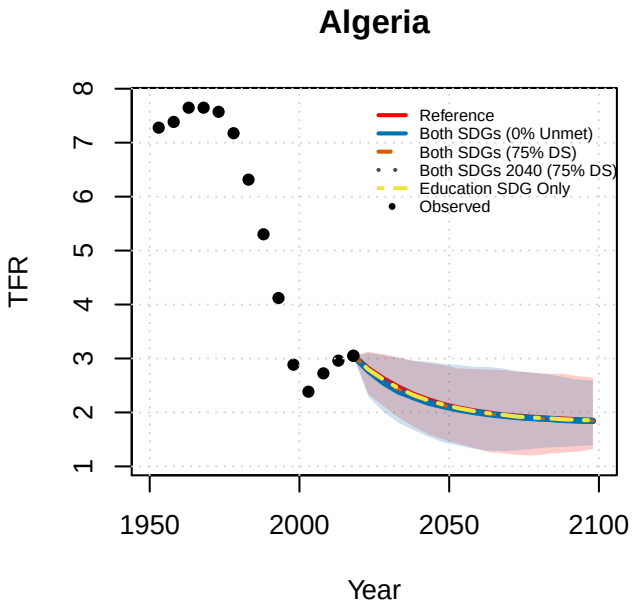

Niger
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

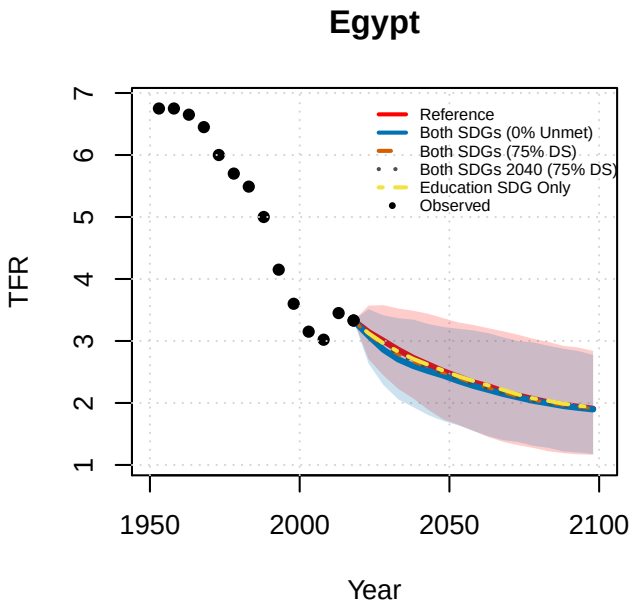

Nigeria
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

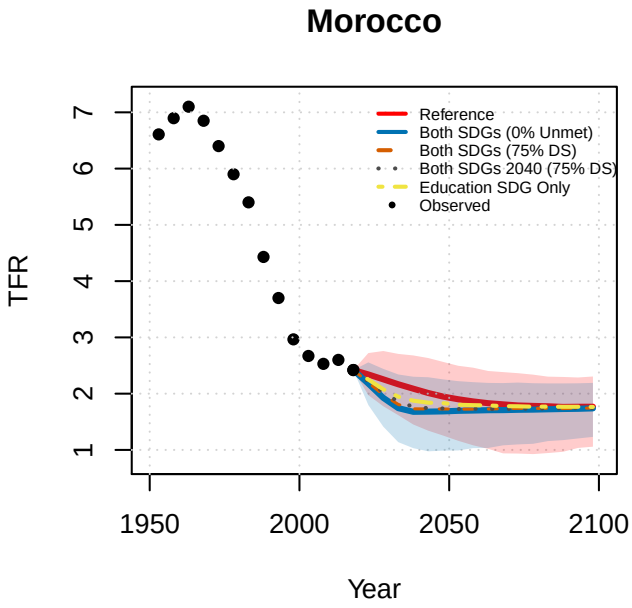

Senegal
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

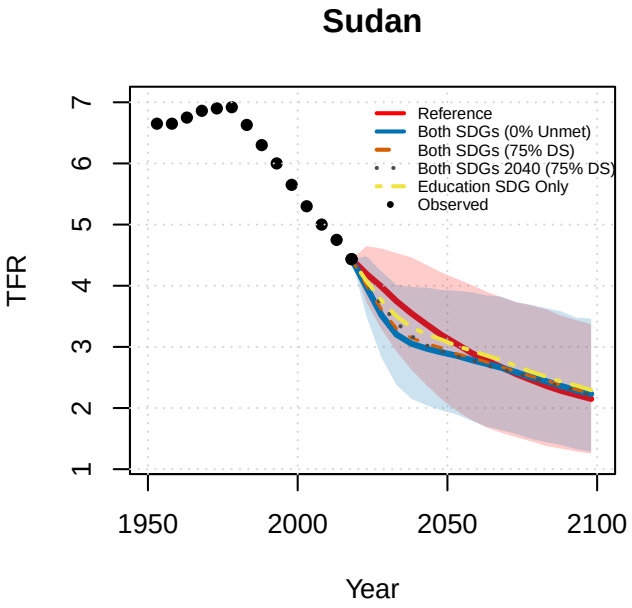

Sierra Leone
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

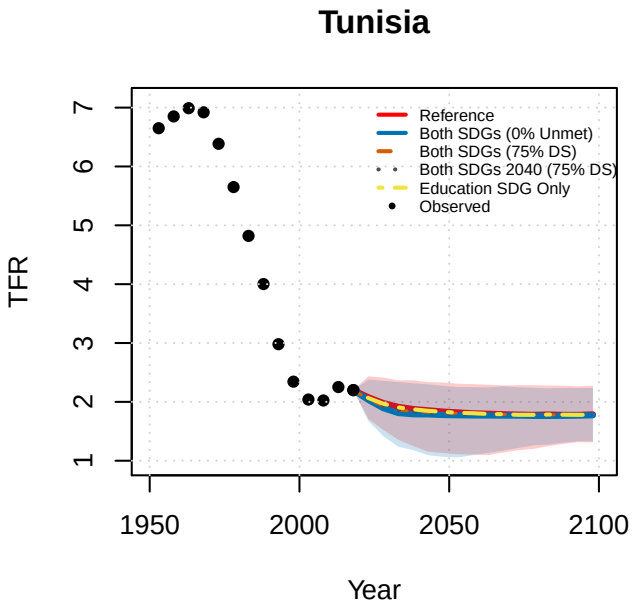

Togo
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

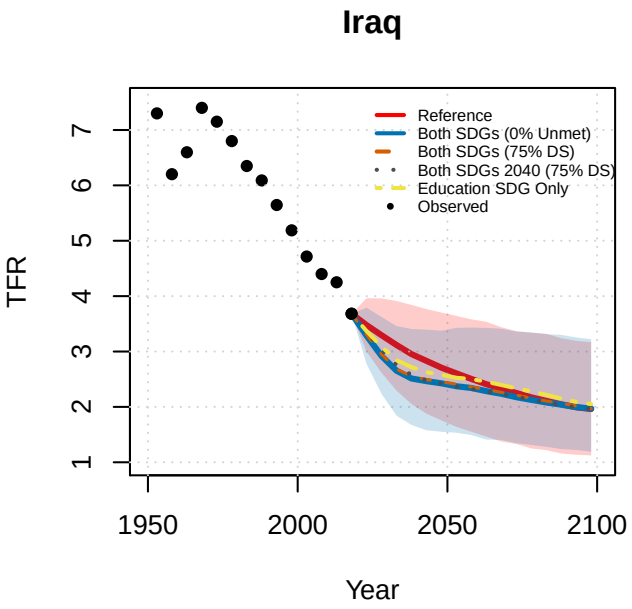

Algeria
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

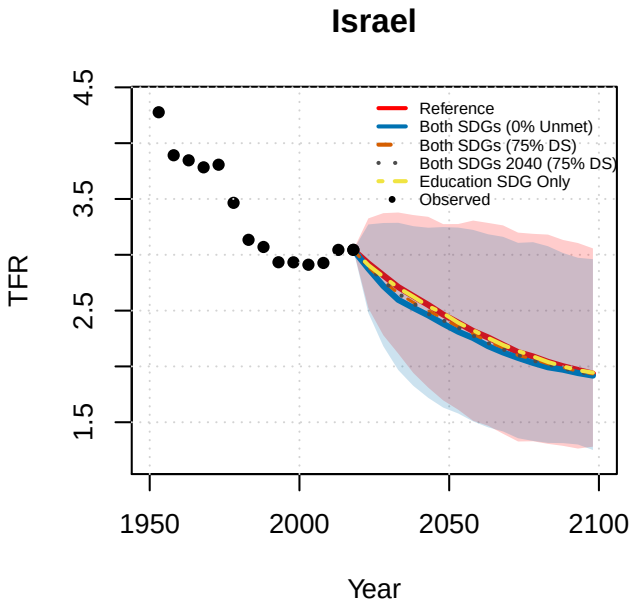

Egypt
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

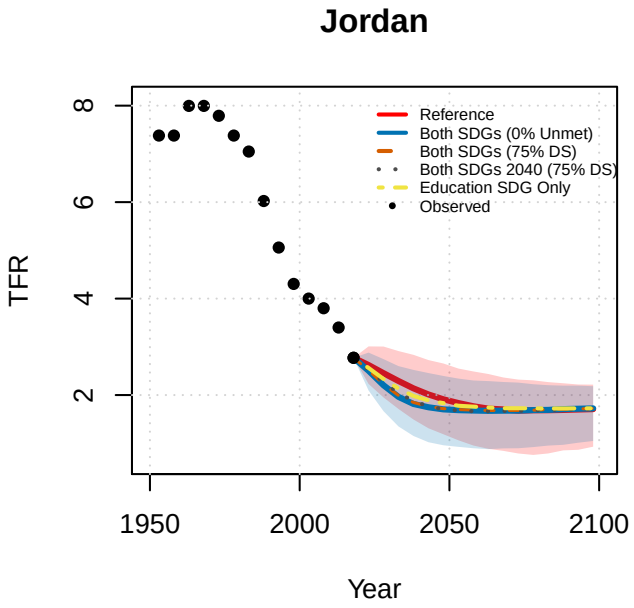

Morocco
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

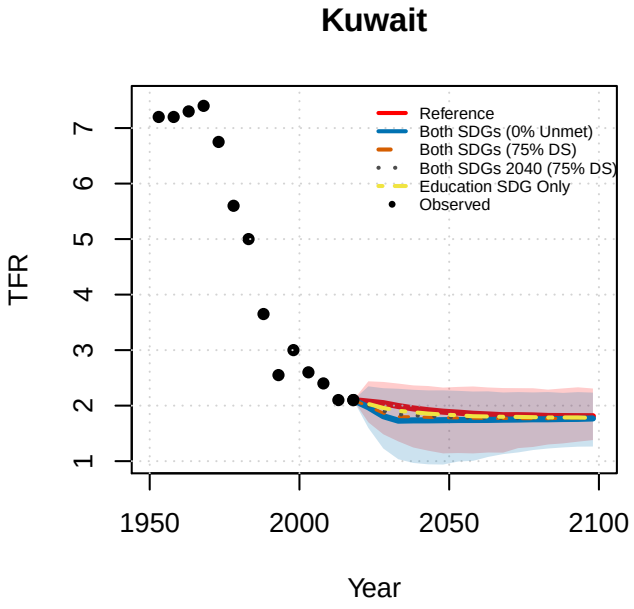

Sudan
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

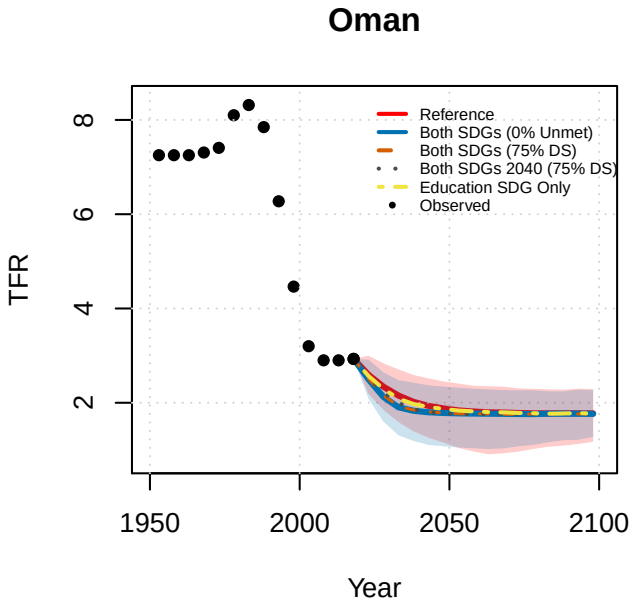

Tunisia
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

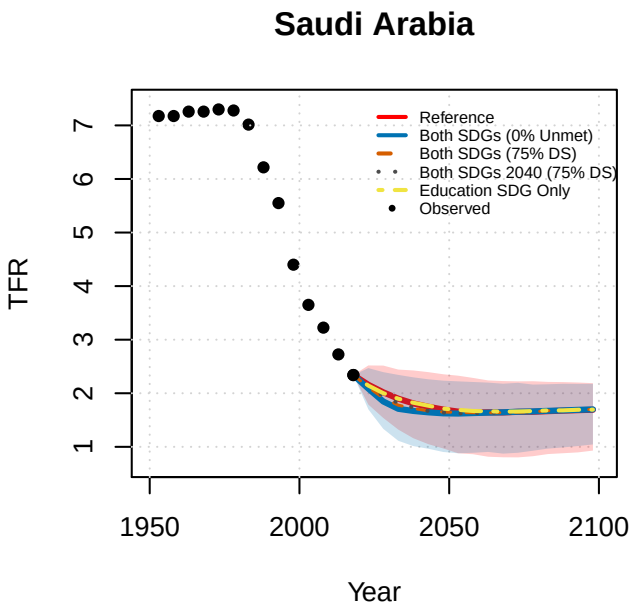

Iraq
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

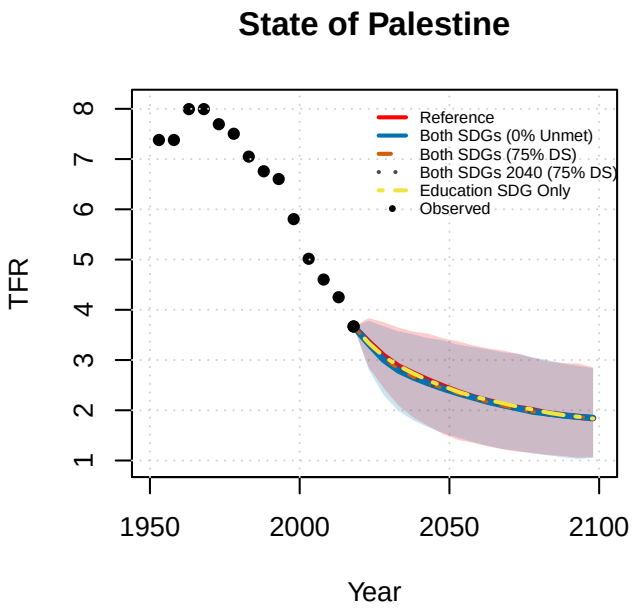

Israel
TFR
Year
1950
2000
2050
2100
1.5
2.5
3.5
4.5
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

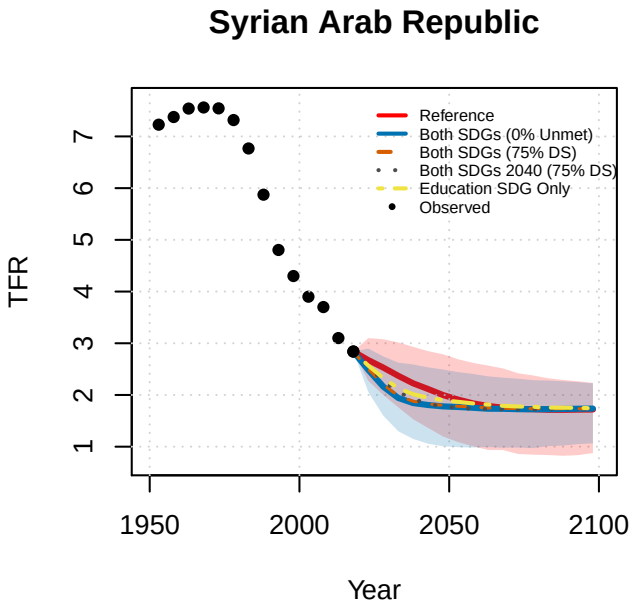

Jordan
TFR
Year
1950
2000
2050
2100
2
4
6
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

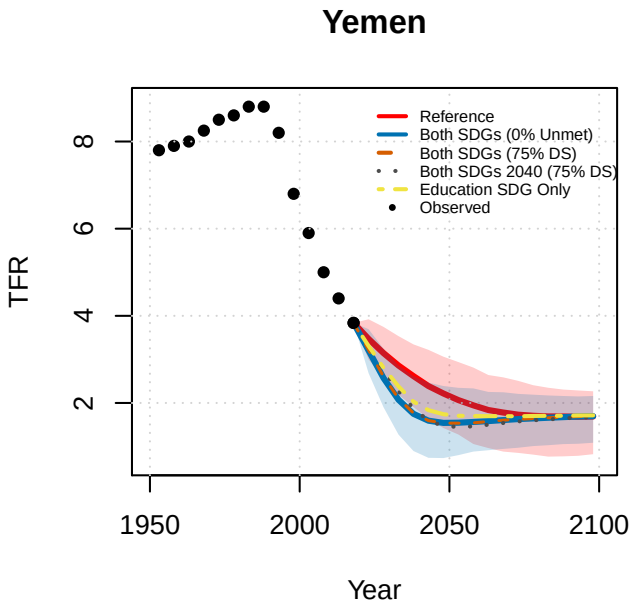

Kuwait
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

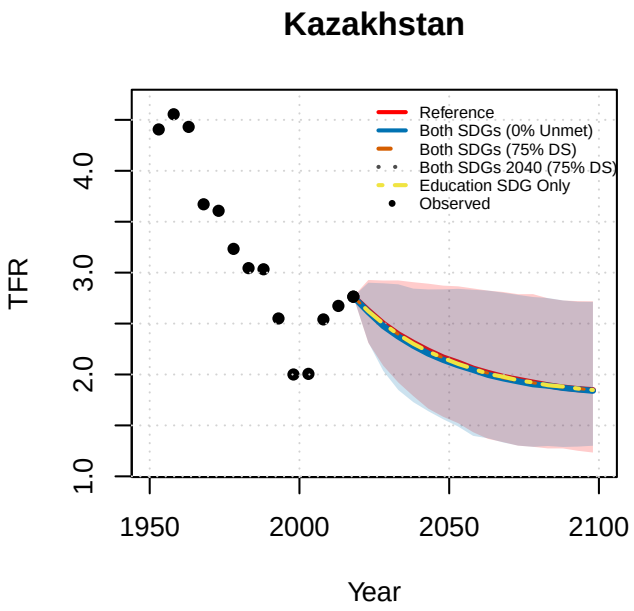

Oman
TFR
Year
1950
2000
2050
2100
2
4
6
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

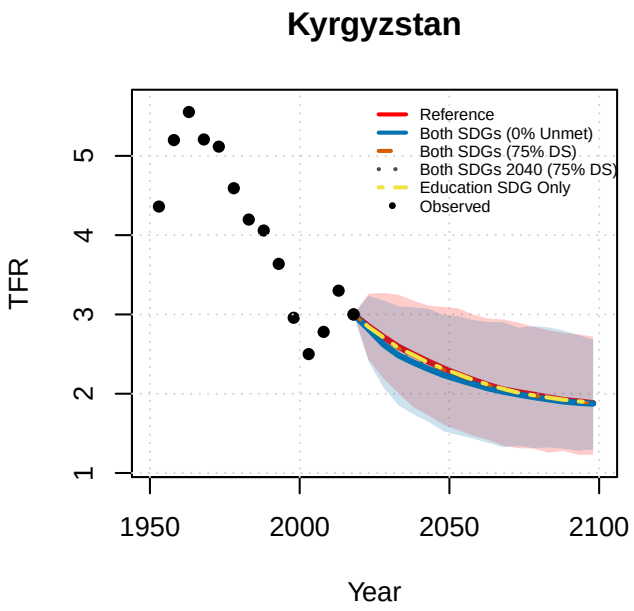

Saudi Arabia
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

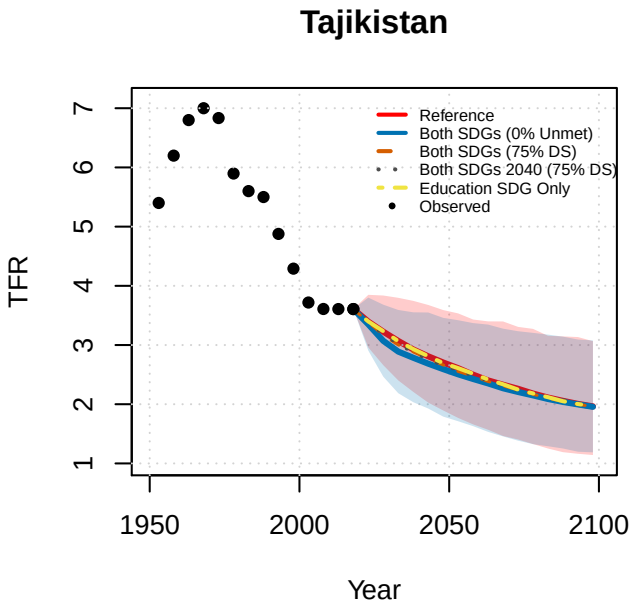

State of Palestine
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

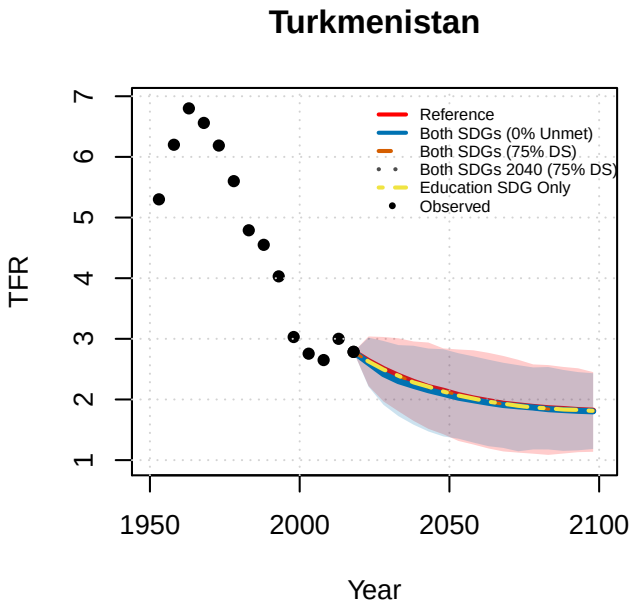

Syrian Arab Republic
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

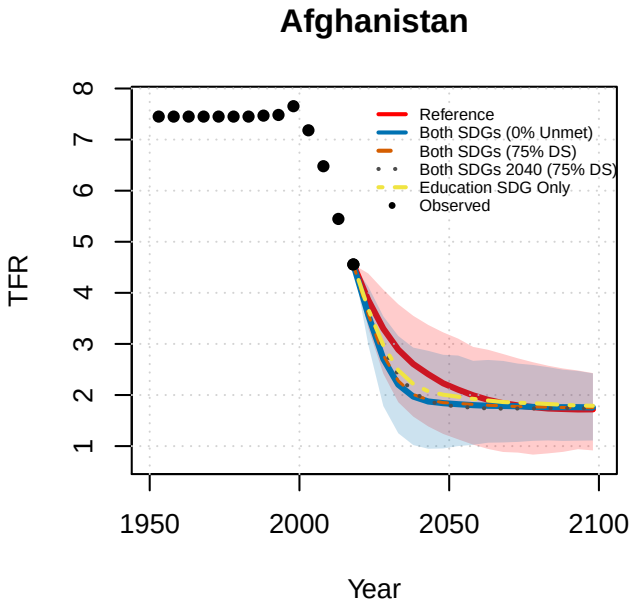

Yemen
TFR
Year
1950
2000
2050
2100
2
4
6
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

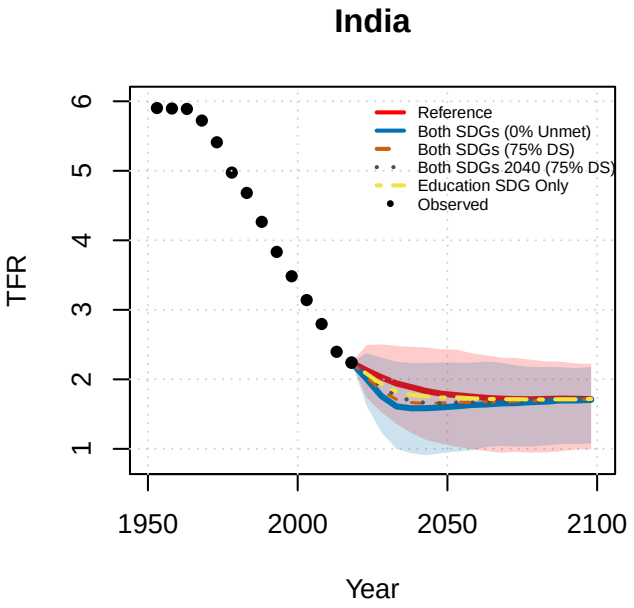

Kazakhstan
TFR
Year
1.0
2.0
3.0
4.0
1950
2000
2050
2100
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

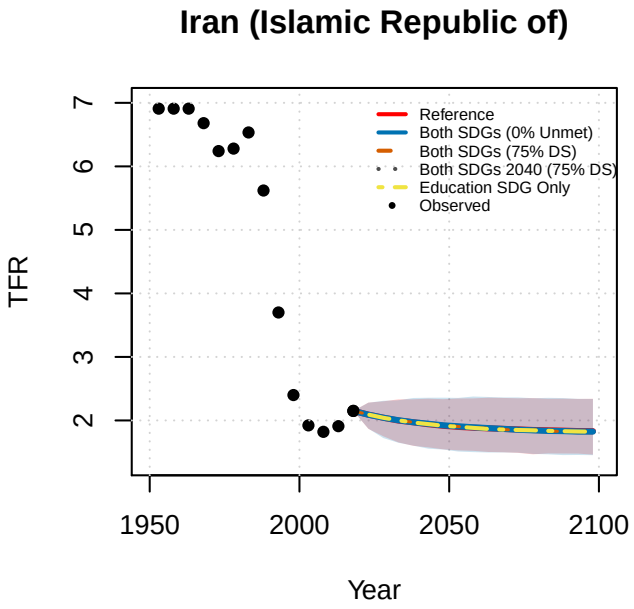

Kyrgyzstan
TFR
Year
1
2
3
4
5
1950
2000
2050
2100
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

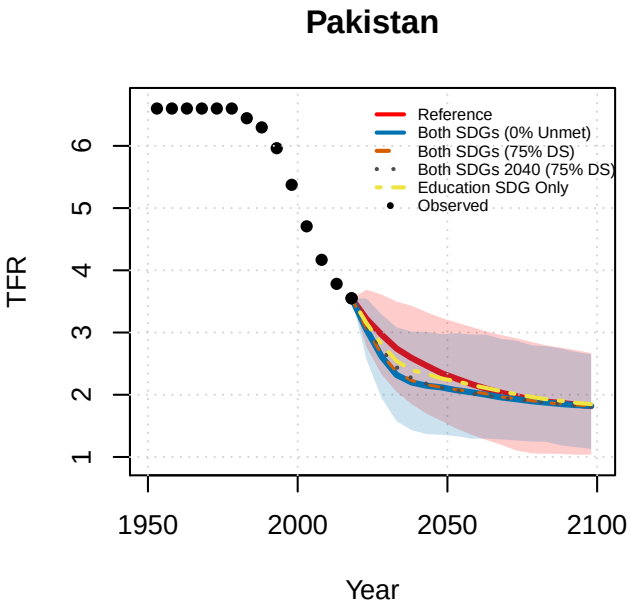

Tajikistan
TFR
Year
1
2
3
4
5
6
7
1950
2000
2050
2100
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

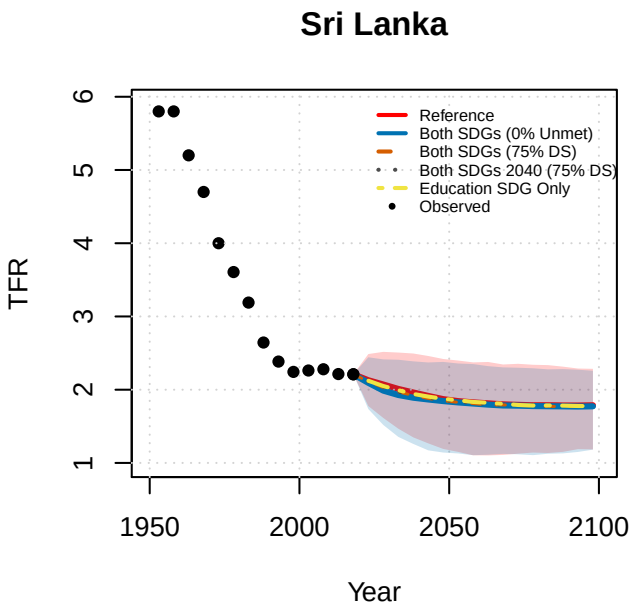

Turkmenistan
TFR
Year
1
2
3
4
5
6
7
1950
2000
2050
2100
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

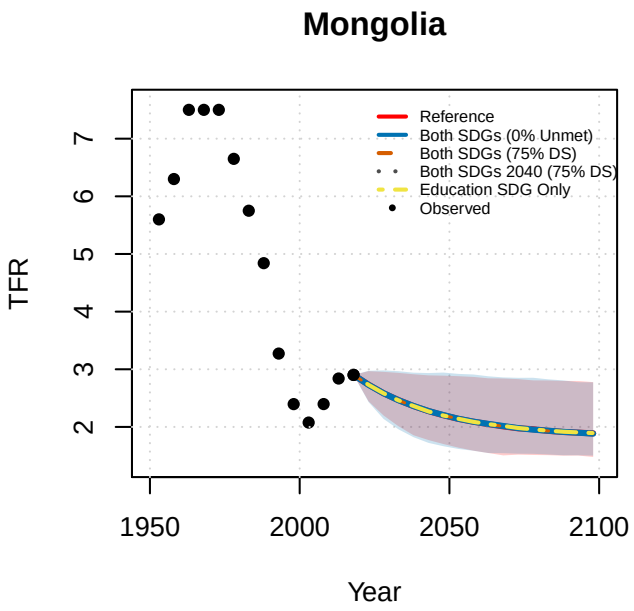

Afghanistan
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

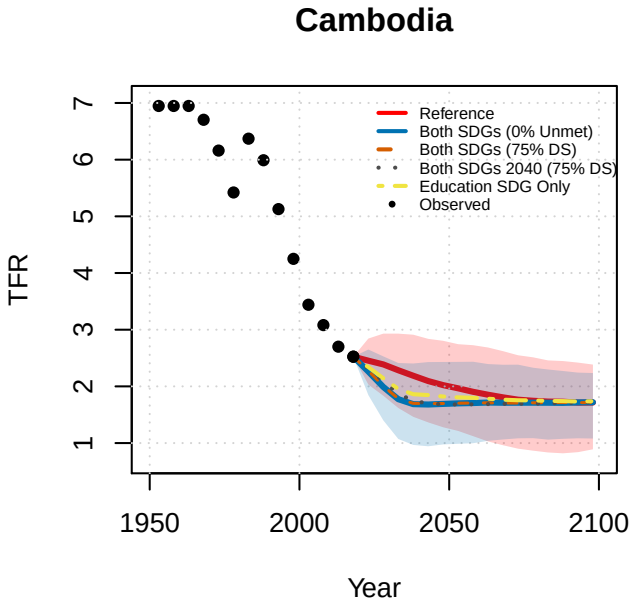

India
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

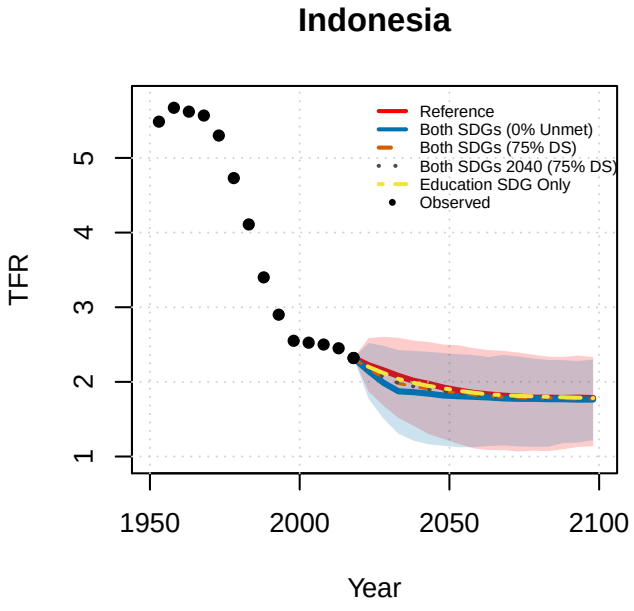

Iran (Islamic Republic of)
TFR
Year
1950
2000
2050
2100
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

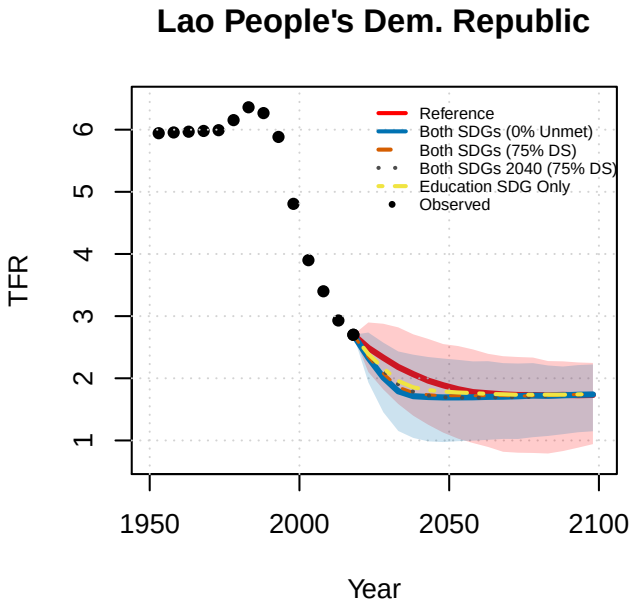

Pakistan
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

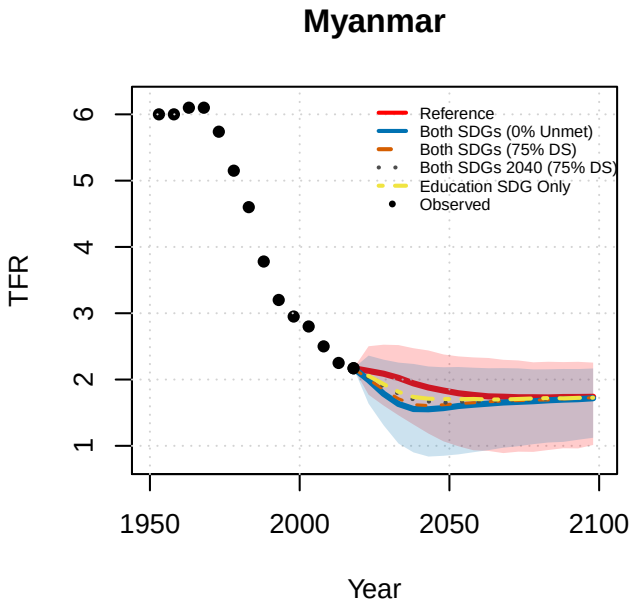

Sri Lanka
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

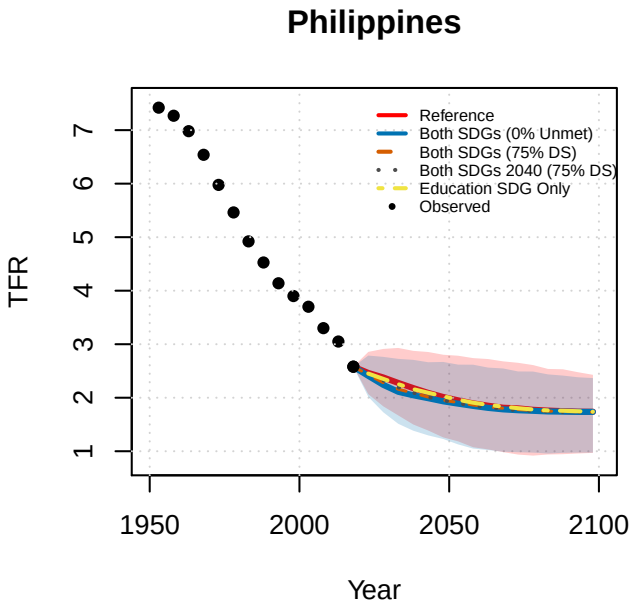

Mongolia
TFR
Year
1950
2000
2050
2100
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

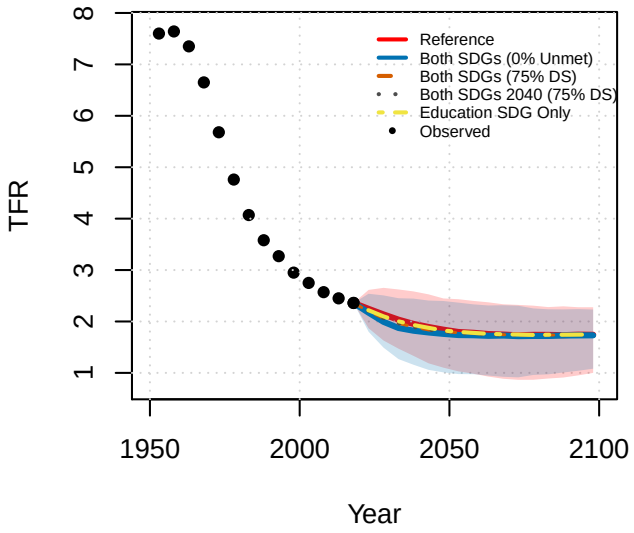

Cambodia
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

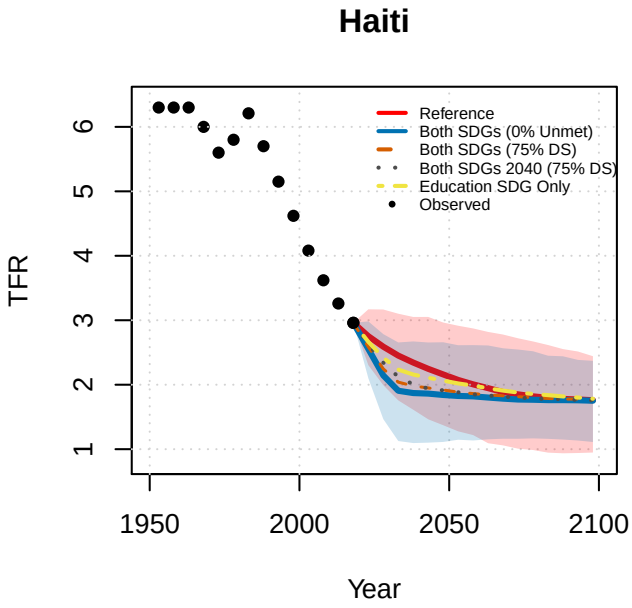

Indonesia
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

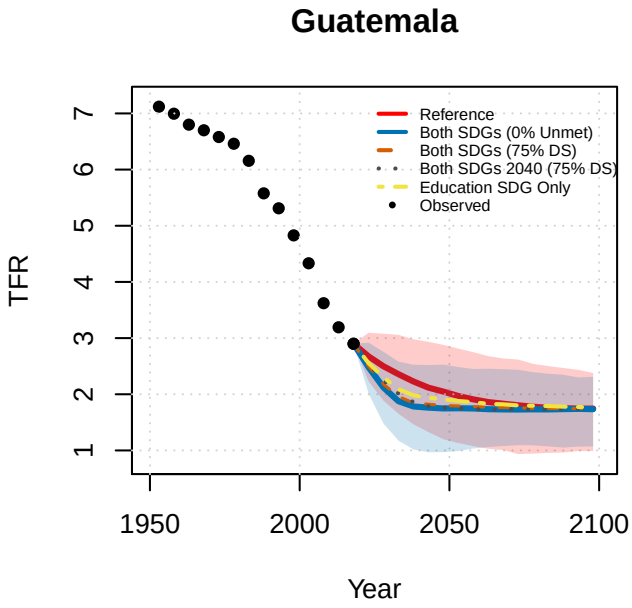

Lao People's Dem. Republic
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

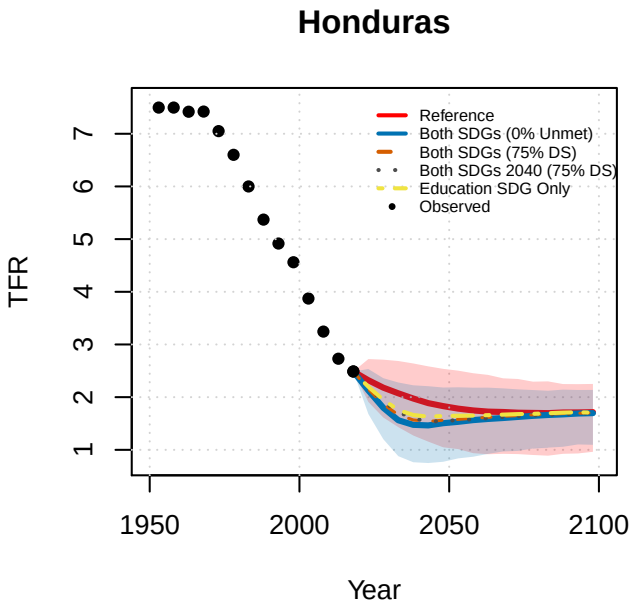

Myanmar
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

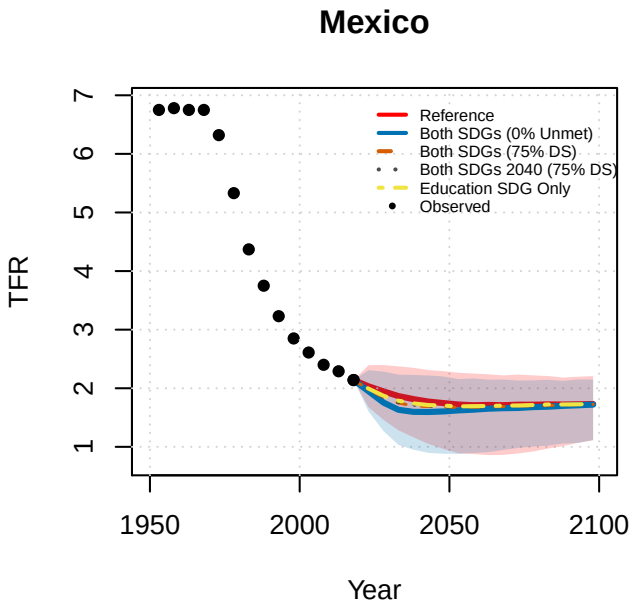

Philippines
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

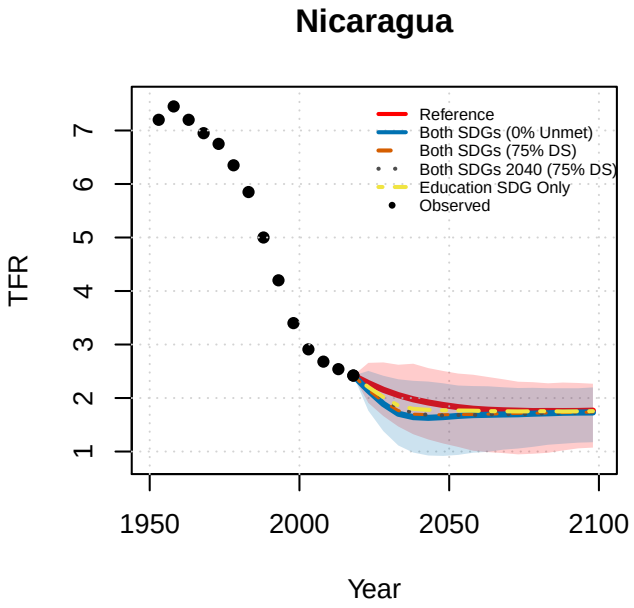

Dominican Republic
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7
8
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed

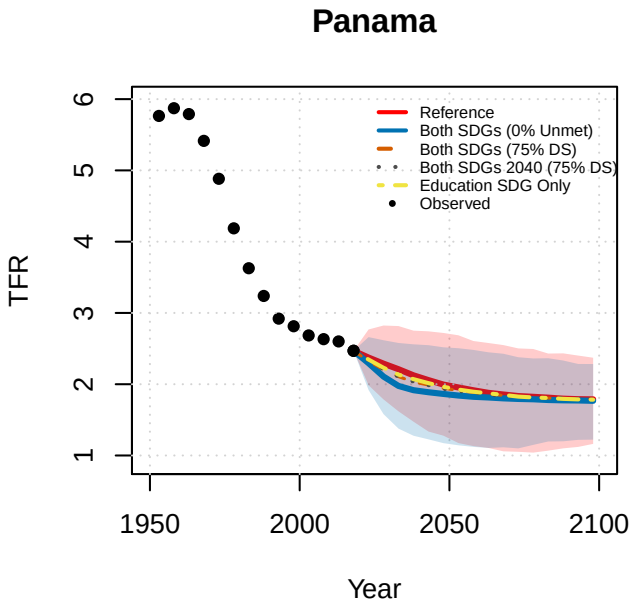

Haiti
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

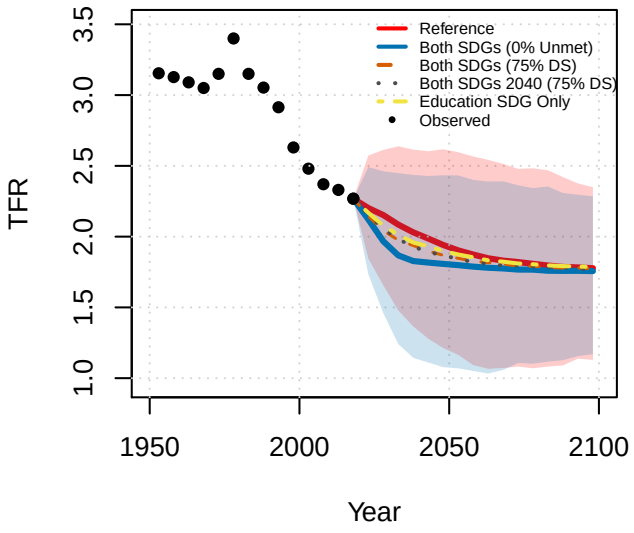

Guatemala
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

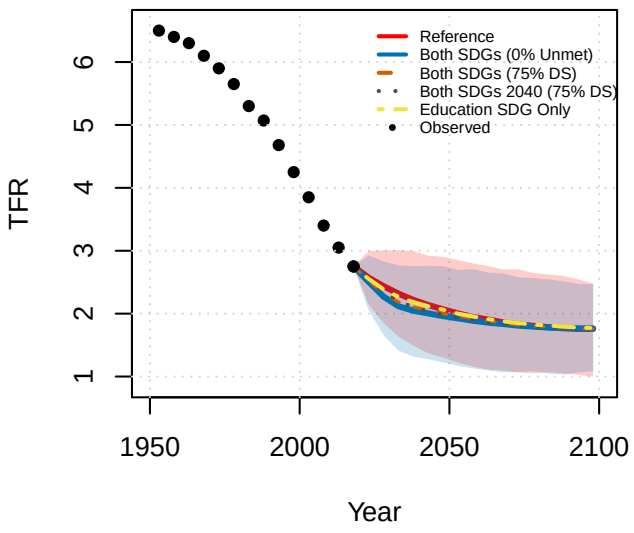

Honduras
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7


Mexico
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7

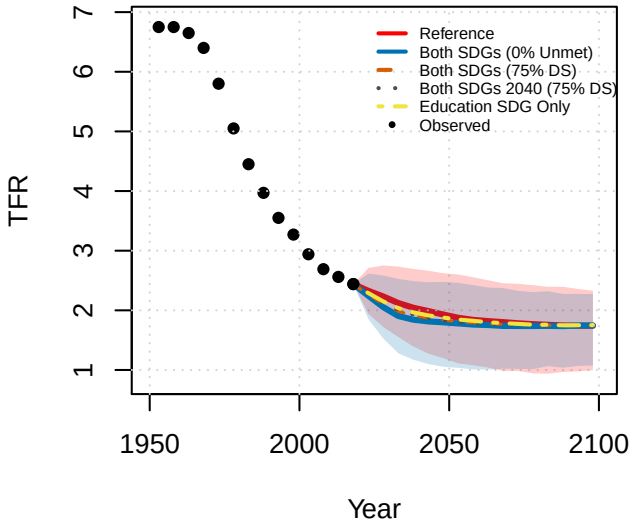

Nicaragua
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6
7


Panama
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

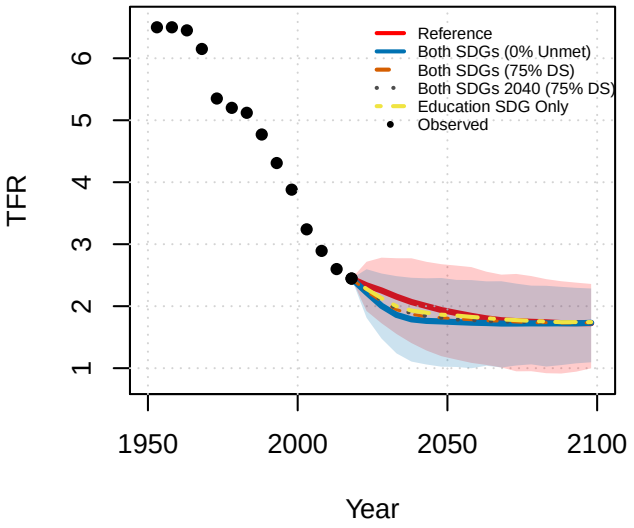

Argentina
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1.0
1.5
2.0
2.5
3.0
3.5


Bolivia (Plurinational State of)
Reference
Both SDGs (0% Unmet)
Both SDGs (75% DS)
Both SDGs 2040 (75% DS)
Education SDG Only
Observed
TFR
Year
1950
2000
2050
2100
1
2
3
4
5
6

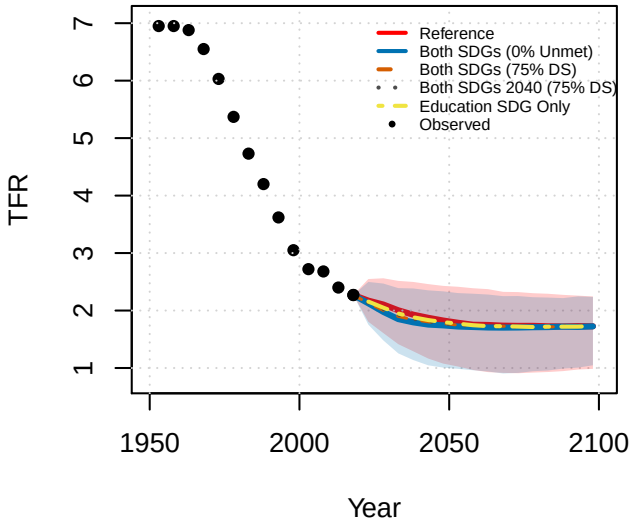
Ecuador



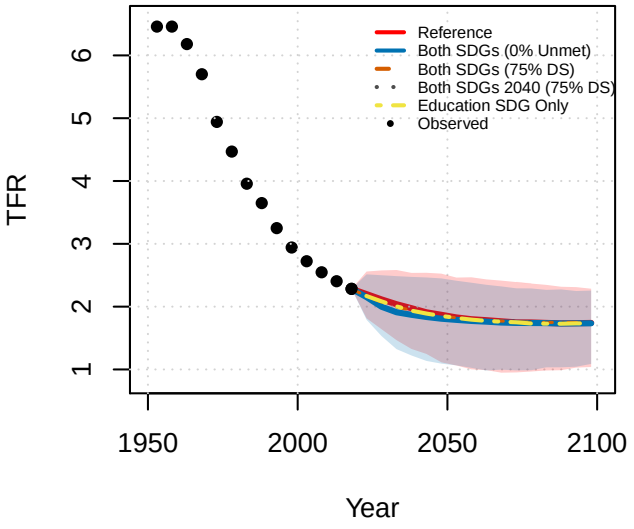
Paraguay



Peru



Venezuela (Bolivarian Republic of)



Appendix C

APPENDICES FOR CHAPTER 4

C.1 Simulation of MAR and MNAR

For MAR, we assumed observations in earlier years are more likely to be missing. For each country-time pair (c, t), we simulated a value $A_{c,t}$ based on t with a large noise term. For the nonlinear and linear simulated data, $A_{c,t} \sim N(t, 10^2)$. For the enrollment data, $A_{c,t} \sim N(t, 40^2)$. For rate of simulated missingness r, we determined which values to simulate as missing based on the rth quantile of $\mathbf{A}$ as follows:

$$A_{boundary} = r\text{th quantile of } \boldsymbol{A}$$

$$R_{c,t} = \begin{cases} 0 & \text{if } A_{c,t} < A_{boundary} \\ 1 & \text{otherwise} \end{cases}$$

where the response indicator $R_{c,t}$ is 0 if the observation for (c, t) is missing and 1 otherwise. Missingness in $\mathbf{X}$ and $\mathbf{Y}$ was simulated independently following the method described above.

For MNAR, we simulated observations as missing depending on their observed values, where smaller values are more likely to be missing. For each $Y_{c,t}$, we first simulated a value $A_{c,t}$ based on the value of $Y_{c,t}$ with a large noise term. For the nonlinear and linear simulated data, $A_{c,t} \sim N(Y_{c,t}, 40^2)$. For the enrollment data, $A_{c,t} \sim N(Y_{c,t}, 15^2)$. For rate of simulated missingness r, we determined which values to simulate as missing based on the rth quantile

of $\mathbf{A}$ as follows:

$$A_{boundary} = r\text{th quantile of } A_{c,t}$$

$$R_{c,t} = \begin{cases} 0 & \text{if } A_{c,t} < A_{boundary} \\ 1 & \text{otherwise} \end{cases}$$

where the response indicator $R_{c,t}$ is 0 if the observation for (c,t) is missing and 1 otherwise. Missingness in $\mathbf{X}$ was simulated analogously to above, where the only difference was in the simulation of $A_{c,t}$. For the nonlinear and linear simulated data, $A_{c,t}$ for $\mathbf{X}$ was simulated as $A_{c,t} \sim N(X_{c,t}, 40^2)$. For the enrollment data, $A_{c,t}$ for $\mathbf{X}$ was simulated as $A_{c,t} \sim N(X_{c,t}, 25^2)$. Missingness in $\mathbf{X}$ was simulated independently of missingness in $\mathbf{Y}$.

C.2 Data-based Control Parameters Algorithm

The control parameters for the prior distributions of the $\mathbf{X}$ model ($\delta_X, \nu_{drift}, \zeta_{drift}$, and δ_{drift}) are estimated using the observed first differences $X_{c,t} - X_{c,t-1}$ as

$$\begin{aligned} \delta_X &= \widehat{Var}\left(X_{c,t} - X_{c,t-1}\right) \\ \nu_{drift} &= \bar{X}_{diff} \\ \zeta_{drift} &= SE(\bar{X}_{diff}) \\ \delta_{drift} &= \widehat{Var}(\bar{X}_{c,diff}) \end{aligned}$$

where

$$\begin{aligned} \bar{X}_{diff} &= \frac{1}{n_{train}} \sum_{c,t} \left(X_{c,t} - X_{c,t-1}\right) \\ \bar{X}_{c,diff} &= \frac{1}{n_c} \sum_{t} \left(X_{c,t} - X_{c,t-1}\right). \end{aligned}$$

The control parameters for the prior distributions of the $\mathbf{Y}|\mathbf{X}$ model ($\delta_Y, \zeta_0,$ and δ_0) are estimated using the observed first differences $Y_{c,t} - Y_{c,t-1}$ as

$$\begin{aligned}\delta_Y =& \widehat{Var}\left(Y_{c,t} - Y_{c,t-1}\right)\\ \zeta_0 =& SE(\bar{Y}_{diff})\\ \delta_0 =& \widehat{Var}(\bar{Y}_{c,diff})\end{aligned}$$

where

$$\begin{aligned}\bar{Y}_{diff} =& \frac{1}{n_{train}} \sum_{c,t} \left(Y_{c,t} - Y_{c,t-1}\right)\\ \bar{Y}_{c,diff} =& \frac{1}{n_c} \sum_{t} \left(Y_{c,t} - Y_{c,t-1}\right).\end{aligned}$$

The control parameters for the prior distributions of $(Y_{c,0}, X_{c,0})$ are estimated for all c as

$$\begin{bmatrix} Y_{c,0} \\ X_{c,0} \end{bmatrix} \sim TN\left(\boldsymbol{\mu}_{early}, \boldsymbol{\Sigma}_{early}\right)$$

$$\boldsymbol{\mu}_{early} = \begin{bmatrix} \bar{Y}_{early} \\ \bar{X}_{early} \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{early} = \begin{bmatrix} \widehat{SD}(Y_{early})^2 & \hat{r} \times \widehat{SD}(Y_{early}) \times \widehat{SD}(X_{early}) \\ \hat{r} \times \widehat{SD}(Y_{early}) \times \widehat{SD}(X_{early}) & \widehat{SD}(X_{early})^2 \end{bmatrix}$$

The prior mean and covariance matrix are calculated using the sample means $\bar{Y}_{early}$ and $\bar{X}_{early}$, sample standard deviations $\widehat{SD}(Y_{early})$ and $\widehat{SD}(X_{early})$, and the sample Pearson correlation coefficient $\hat{r}$ for a subset of the data $(\mathbf{Y}_{early}, \mathbf{X}_{early})$ that consists of all complete cases $(Y_{c,t}, X_{c,t})$ in "early" times. The time range determined to be "early" is based on availability of the observed complete cases. For the enrollment data, $(\mathbf{Y}_{early}, \mathbf{X}_{early})$ consists of all country-time pairs in years 1970-1980 (corresponding to $t = 1, ..., 11$) where both $Y_{c,t}$ and $X_{c,t}$ were observed. The truncation was specified as $X_{c,0} \in [0, min(X_{c,obs})]$ and $Y_{c,0} \in [0, min(Y_{c,obs})]$, where $min(X_{c,obs})$ is the minimum observed value of X for country c and $min(Y_{c,obs})$ is the minimum observed value of Y for country c.

C.3 MCMC Algorithm

The MCMC algorithm for the MINTS method uses univariate Gibbs sampling steps for most imputation model parameters combined with a Metropolis-within-Gibbs step for the parameter block (β, ρ). Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_X, \boldsymbol{\theta}_Y)$ denote the vector of all parameters in the imputation model, where $\boldsymbol{\theta}_X = (\boldsymbol{\gamma}, \sigma^2_X, \mathbf{X}_0, \mu_{drift}, \sigma^2_{drift})$ and $\boldsymbol{\theta}_Y = (\boldsymbol{\alpha}, \beta, \rho, \sigma^2_Y, \mathbf{Y}_0, \mu_0, \sigma^2_0)$. Let $n = C \times T$ denote the total number of country-years. Let the hyperparameters for the joint distribution of $\mathbf{X}_0$ and $\mathbf{Y}_0$ be written as

$$\boldsymbol{\mu}_{early} = \begin{bmatrix} \mu_{Y,early} \\ \mu_{X,early} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{early} = \begin{bmatrix} \sigma^2_{Y,early} & r \times \sigma_{Y,early}\sigma_{X,early} \\ r \times \sigma_{Y,early}\sigma_{X,early} & \sigma^2_{X,early} \end{bmatrix}.$$

To simplify notation, let $\mathbf{X}^{(i)} = (\mathbf{X}_{obs}, \mathbf{X}^{(i)}_{mis})$ represent the vector of the observed portion of $\mathbf{X}$ and the imputed values for $\mathbf{X}_{mis}$ from the ith iteration. Similarly, let $\mathbf{Y}^{(i)} = (\mathbf{Y}_{obs}, \mathbf{Y}^{(i)}_{mis})$.

Each chain of the MCMC is initialized with starting values $\boldsymbol{\theta}^{(0)}_X$ and $\boldsymbol{\theta}^{(0)}_Y$ randomly drawn from diffuse distributions modeled after the prior distributions. Initial values for the missing observations are constructed using a combination of linear interpolation of the observed values, predicted values from the spline f, and constant extrapolation. In each chain, starting values $\mathbf{X}^{(0)}_{mis}$ and $\mathbf{Y}^{(0)}_{mis}$ are randomly drawn from normal distributions centered at the previously constructed initial values.

For each chain, the MCMC sampling algorithm then proceeds as follows for iterations $i = 1, ..., .n_{iter}$:

1. $\sigma^{2,(i)}_Y$ is sampled from

$$\begin{aligned}
&\sigma^2_Y | \beta^{(i-1)}, \rho^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \mathbf{X}^{(i-1)}_0, \mathbf{Y}^{(i-1)}_0, \mathbf{X}^{(i-1)}, \mathbf{Y}^{(i-1)} \\
&\sim InvGamma\Bigg(2 + \frac{n}{2}, \\
&\qquad \delta_Y + \frac{1}{2}\sum_{t=1}^{T}\left(\frac{1}{h(\mathbf{X}^{(i-1)}_t)}\left(\mathbf{Y}^{(i-1)}_t - \rho^{(i-1)}\mathbf{Y}^{(i-1)}_{t-1} - \boldsymbol{\alpha}^{(i-1)} - \beta^{(i-1)} f(\mathbf{X}^{(i-1)}_t)\right)^2\right)\Bigg).
\end{aligned}$$

2. The parameters $(\beta^{(i)}, \rho^{(i)})$ are sampled as a block using a Metropolis-Hastings step. Values β_{new} and ρ_{new} are drawn from the proposal function given by the truncated bivariate normal distribution

$$TN\left(\begin{bmatrix}\beta^{(i-1)}\\ \rho^{(i-1)}\end{bmatrix}, \phi\begin{bmatrix}0.001475944 & -0.001511349\\ -0.001511349 & 0.001577437\end{bmatrix}\right),$$

where the tuning parameter ϕ for the covariance matrix is determined following a pilot run of the MCMC to ensure the acceptance rate is around 40% and the truncation is such that $\beta \in (-\infty, \infty)$ and $\rho \in [0, 1]$.

The acceptance ratio is calculated as

$$\begin{aligned}&\frac{p(\mathbf{Y}^{(i-1)}|\mathbf{X}^{(i-1)}, \mathbf{Y}_0^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \sigma_Y^{2,(i)}, \beta_{new}, \rho_{new})}{p(\mathbf{Y}^{(i-1)}|\mathbf{X}^{(i-1)}, \mathbf{Y}_0^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \sigma_Y^{2,(i)}, \beta^{(i-1)}, \rho^{(i-1)})}\\ &= \exp\left[\sum_{t=1}^{T}\frac{-\left(\mathbf{Y}_t^{(i-1)} - \boldsymbol{\alpha}^{(i-1)} - \beta^{new}f(\mathbf{X}_t^{(i-1)}) - \rho^{new}\mathbf{Y}_{t-1}^{(i-1)}\right)^2}{2\sigma_Y^{2,(i)}h(\mathbf{X}_t^{(i-1)})}\right.\\ &\qquad\left.-\sum_{t=1}^{T}\frac{-\left(\mathbf{Y}_t^{(i-1)} - \boldsymbol{\alpha}^{(i-1)} - \beta^{(i-1)}f(\mathbf{X}_t^{(i-1)}) - \rho^{(i-1)}\mathbf{Y}_{t-1}^{(i-1)}\right)^2}{2\sigma_Y^{2,(i)}h(\mathbf{X}_t^{(i-1)})}\right].\end{aligned}$$

3. $\sigma_X^{2,(i)}$ is sampled from

$$\begin{aligned}&\sigma_X^2|\boldsymbol{\gamma}^{(i-1)}, \mathbf{X}_0^{(i-1)}, \mathbf{X}^{(i-1)}\\ &\sim InvGamma\left(2 + \frac{n}{2}, \delta_X + \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{X}_t^{(i-1)} - \mathbf{X}_{t-1}^{(i-1)} - \boldsymbol{\gamma}^{(i-1)}\right)^2\right).\end{aligned}$$

4. To sample the country-specific drift terms $\gamma_c^{(i)}$, the hyperparameters $\mu_{drift}^{(i)}$ and $\sigma_{drift}^{2,(i)}$ are first sampled from

$$\mu_{drift}^{(i)}|\sigma_{drift}^{2,(i-1)}, \boldsymbol{\gamma}^{(i-1)} \sim N\left(\left(\frac{1}{\zeta_{drift}^2} + \frac{C}{\sigma_{drift}^{2,(i-1)}}\right)^{-1}\left(\frac{\nu_{drift}}{\zeta_{drift}^2} + \sum_{c=1}^{C}\frac{\gamma_c^{(i-1)}}{\sigma_{drift}^{2,(i-1)}}\right), \left(\frac{1}{\zeta_{drift}^2} + \frac{C}{\sigma_{drift}^{2,(i-1)}}\right)^{-1}\right)$$

$$\sigma_{drift}^{2,(i)}|\mu_{drift}^{(i)}, \boldsymbol{\gamma}^{(i-1)} \sim InvGamma\left(2 + \frac{C}{2}, \delta_{drift} + \frac{1}{2}\sum_{c=1}^{C}\left(\gamma_c^{(i-1)} - \mu_{drift}^{(i)}\right)^2\right).$$

For each c, $\gamma_c^{(i)}$ is sampled from

$$\gamma_c^{(i)}|\mu_{drift}^{(i)}, \sigma_{drift}^{2,(i)}, \sigma_X^{2,(i)}, \mathbf{X}_0^{(i-1)}, \mathbf{X}^{(i-1)}$$
$$\sim N\left(\left(\frac{1}{\sigma_{drift}^{2,(i)}} + \frac{T}{\sigma_X^{2,(i)}}\right)^{-1}\left(\frac{\mu_{drift}^{(i)}}{\sigma_{drift}^{2,(i)}} + \sum_{t=1}^{T}\frac{X_{c,t}^{(i-1)} - X_{c,t-1}^{(i-1)}}{\sigma_X^{2,(i)}}\right), \left(\frac{1}{\sigma_{drift}^{2,(i)}} + \frac{T}{\sigma_X^{2,(i)}}\right)^{-1}\right).$$

5. To sample the country-specific random intercept terms $\alpha_c^{(i)}$, the hyperparameters $\mu_0^{(i)}$ and $\sigma_0^{2,(i)}$ are first sampled from

$$\mu_0^{(i)}|\sigma_0^{2,(i-1)}, \boldsymbol{\alpha}^{(i-1)} \sim N\left(\left(\frac{1}{\zeta_0^2} + \frac{C}{\sigma_0^{2,(i-1)}}\right)^{-1}\left(\frac{\nu_0}{\zeta_0^2} + \sum_{c=1}^{C}\frac{\alpha_c^{(i-1)}}{\sigma_0^{2,(i-1)}}\right), \left(\frac{1}{\zeta_0^2} + \frac{C}{\sigma_0^{2,(i-1)}}\right)^{-1}\right)$$
$$\sigma_0^{2,(i)}|\mu_0^{(i)}, \boldsymbol{\alpha}^{(i-1)} \sim InvGamma\left(2 + \frac{C}{2}, \delta_0 + \frac{1}{2}\sum_{c=1}^{C}\left(\alpha_c^{(i-1)} - \mu_0^{(i)}\right)^2\right).$$

For each c, $\alpha_c^{(i)}$ is sampled from

$$\alpha_c^{(i)}|\mu_0^{(i)}, \sigma_0^{2,(i)}, \sigma_Y^{2,(i)}, \mathbf{Y}_0^{(i-1)}, \mathbf{Y}^{(i-1)}, \mathbf{X}^{(i-1)}$$
$$\sim N\left(\left(\frac{1}{\sigma_0^{2,(i)}} + \sum_{t=1}^{T}\frac{1}{\sigma_Y^{2,(i)}h(X_{c,t}^{(i-1)})}\right)^{-1}\left(\frac{\mu_0^{(i)}}{\sigma_0^{2,(i)}} + \sum_{t=1}^{T}\frac{Y_{c,t}^{(i-1)} - \rho^{(i)}Y_{c,t-1}^{(i-1)} - \beta^{(i)}f(X_{c,t}^{(i-1)})}{\sigma_Y^{2,(i)}h(X_{c,t}^{(i-1)})}\right),\right.$$
$$\left.\left(\frac{1}{\sigma_0^{2,(i)}} + \sum_{t=1}^{T}\frac{1}{\sigma_Y^{2,(i)}h(X_{c,t}^{(i-1)})}\right)^{-1}\right).$$

6. Imputed values for each missing value in $\mathbf{X}_{mis}^{(i)}$ are drawn for each country in descending order by year. For all $X_{c,t}$ that are observed, we set $X_{c,t}^{(i)} = X_{c,t}$ for all iterations i. For $t = T$, the imputed value for $X_{c,t}^{(i)}$ is drawn from

$$X_{c,T}^{(i)}|\gamma_c^{(i)}, \sigma_X^{2,(i)}, X_{c,T-1}^{(i-1)} \sim TN_{[X_{low}, X_{up}]}\left(X_{c,T-1}^{(i-1)} + \gamma_c^{(i)}, \sigma_X^{2,(i)}\right).$$

For $t < T$, the imputed value for $X_{c,t}^{(i)}$ is drawn in descending order from largest t to smallest t from

$$X_{c,t}^{(i)}|\gamma_c^{(i)}, \sigma_X^{2,(i)}, X_{c,t+1}^{(i)}, X_{c,t-1}^{(i-1)} \sim TN_{[X_{low}, X_{up}]}\left(\frac{X_{c,t+1}^{(i)} + X_{c,t-1}^{(i-1)}}{2}, \frac{\sigma_X^{2,(i)}}{2}\right).$$

7. $X_{c,0}^{(i)}$ is sampled from

$$X_{c,0}^{(i)}|\gamma_c^{(i)}, \sigma_X^{2,(i)}, \boldsymbol{\mu}_{early}, \boldsymbol{\Sigma}_{early}, X_{c,1}^{(i)}$$
$$\sim TN_{[X_{0,low}, X_{0,up}]}\left(\left(\frac{1}{\sigma_X^{2,(i)}} + \frac{1}{\sigma_{X,early}^2}\right)^{-1}\left(\frac{X_{c,1}^{(i)} - \gamma_c^{(i)}}{\sigma_X^{2,(i)}} + \frac{\mu_{X,early}}{\sigma_{X,early}^2}\right), \left(\frac{1}{\sigma_X^{2,(i)}} + \frac{1}{\sigma_{X,early}^2}\right)^{-1}\right).$$

8. Imputed values for each missing value in $\mathbf{Y}_{mis}^{(i)}$ are drawn for each country in descending order by year. For all $Y_{c,t}$ that are observed, we set $Y_{c,t}^{(i)} = Y_{c,t}$ for all iterations i. For $t = T$, the imputed value for $Y_{c,t}^{(i)}$ is drawn from

$$Y_{c,T}^{(i)}|\alpha_c^{(i)}, \rho^{(i)}, \beta^{(i)}, \sigma_Y^{2,(i)}, Y_{c,T-1}^{(i-1)}, X_{c,T}^{(i)} \sim TN_{[Y_{low}, Y_{up}]}\left(\alpha_c^{(i)} + \rho^{(i)} Y_{c,T-1}^{(i-1)} + \beta^{(i)} f(X_{c,T}^{(i)}), \sigma_Y^{2,(i)} h(X_{c,T}^{(i)})\right).$$

For $t < T$, the imputed value for $Y_{c,t}^{(i)}$ is drawn in descending order from largest t to smallest t from

$$Y_{c,t}^{(i)}|\alpha_c^{(i)}, \rho^{(i)}, \beta^{(i)}, \sigma_Y^{2,(i)}, Y_{c,t+1}^{(i)}, Y_{c,t-1}^{(i-1)}, X_{c,t}^{(i)}, X_{c,t+1}^{(i)}$$
$$\sim TN_{[Y_{low}, Y_{up}]}\left(\left(\frac{(\rho^{(i)})^2}{\sigma_Y^{2,(i)} h(X_{c,t+1}^{(i)})} + \frac{1}{\sigma_Y^{2,(i)} h(X_{c,t}^{(i)})}\right)^{-1}\left(\frac{\rho^{(i)}(Y_{c,t+1}^{(i)} - \alpha_c^{(i)} - \beta^{(i)} f(X_{c,t+1}^{(i)}))}{\sigma_Y^{2,(i)} h(X_{c,t+1}^{(i)})} + \frac{\rho^{(i)} Y_{c,t-1}^{(i-1)} + \alpha_c^{(i)} + \beta^{(i)} f(X_{c,t}^{(i)})}{\sigma_Y^{2,(i)} h(X_{c,t}^{(i)})}\right), \left(\frac{(\rho^{(i)})^2}{\sigma_Y^{2,(i)} h(X_{c,t+1}^{(i)})} + \frac{1}{\sigma_Y^{2,(i)} h(X_{c,t}^{(i)})}\right)^{-1}\right).$$

9. $Y_{c,0}^{(i)}$ is sampled from

$$Y_{c,0}^{(i)}|\alpha_c^{(i)}, \rho^{(i)}, \beta^{(i)}, \sigma_Y^{2,(i)}, X_{c,0}^{(i)}, \boldsymbol{\mu}_{early}, \boldsymbol{\Sigma}_{early}, Y_{c,1}^{(i)}, X_{c,1}^{(i)}$$
$$\sim TN_{[Y_{0,low}, Y_{0,up}]}\left(\left(\frac{(\rho^{(i)})^2}{\sigma_Y^{2,(i)} h(X_{c,1}^{(i)})} + \frac{1}{(1-r^2)\sigma_{Y,early}^2}\right)^{-1}\left(\frac{\rho^{(i)}(Y_{c,1}^{(i)} - \alpha_c^{(i)} - \beta^{(i)} f(X_{c,1}^{(i)}))}{\sigma_Y^{2,(i)} h(X_{c,1}^{(i)})} + \frac{\mu_{Y,early} + \frac{r\sigma_{Y,early}}{\sigma_{X,early}}(X_{c,0}^{(i)} - \mu_{X,early})}{(1-r^2)\sigma_{Y,early}^2}\right), \left(\frac{(\rho^{(i)})^2}{\sigma_Y^{2,(i)} h(X_{c,1}^{(i)})} + \frac{1}{(1-r^2)\sigma_{Y,early}^2}\right)^{-1}\right).$$

C.4 Analysis Model Validation Results for Random Intercept Models

C.4.1 Nonlinear Data

For the nonlinear simulated data, the second model used for analysis model validation is the random intercept model given by

$$\begin{aligned} Z_{c,t} &= \lambda_0 + \lambda_c + \lambda_1 Y_{c,t} + \varepsilon^{\lambda}_{c,t}, \\ \lambda_c &\sim N(0, \sigma^2_{\lambda}), \\ \varepsilon^{\lambda}_{c,t} &\sim N(0, \sigma^2_{\varepsilon_{\lambda}}). \end{aligned}$$

The parameter of interest in the fixed component of the model is the coefficient λ_1, while the parameter of interest in the random component of the model is the variance of the random intercepts σ^2_{λ}.

We first evaluated how well each method performs for estimation of $Q = \lambda_1$, the fixed effect coefficient on $\mathbf{Y}$. Table C.1 summarizes the results of this validation for the nonlinear simulated data. Overall, the validation results for estimation of the fixed effect coefficient are similar to the results for the linear regression analysis model. MINTS has the smallest MAE in all experiments except MAR 10%, where MICE PMM has slightly smaller MAE. MINTS also has generally good coverage and small FMI across experiments. While coverage is poor for all methods at the 80% rate, MINTS has the closest to nominal coverage. Out of the existing methods, MICE PMM has good performance for the experiments at the lower rates of missingness but performs poorly at the 80% rate. Pan Random Effects and Amelia TSCS have the best overall performance out of the existing methods that are specifically designed for hierarchical time series data, particularly in the experiments where the missing data mechanism is MCAR. However, both pan Random Effects and Amelia TSCS have worse performance than MINTS.

Table C.2 summarizes the MAE for this analysis model validation for the nonlinear simulated data. We find MINTS has the smallest MAE in all experiments except one. In the

Table C.1: Summary of analysis model validation for nonlinear simulated data for $Q = \lambda_1$, the fixed effect coefficient coefficient on Y in the random intercept model of Z on Y. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 1000 replications of each experiment. The true value of Q is 2.028.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	1.23	100.0	8.2	**0.51**	100.0	4.3	0.78	100.0	5.8
	pan Fixed	1.94	100.0	6.4	6.23	92.6	7.6	3.49	100.0	5.9
	pan Random	1.27	100.0	4.2	5.71	97.3	5.9	2.58	100.0	4.0
	Amelia TS	2.46	100.0	8.2	6.00	99.3	9.5	3.86	100.0	7.4
	Amelia CS	2.59	100.0	9.0	6.49	98.4	10.2	3.93	100.0	7.6
	Amelia TSCS	1.34	100.0	4.9	5.96	**95.2**	6.6	2.65	100.0	4.3
	MINTS	**0.30**	100.0	**0.6**	0.59	100.0	**0.7**	**0.31**	100.0	**0.5**
40%	MICE PMM	18.44	2.9	43.4	5.34	**99.8**	27.0	12.53	28.6	38.9
	pan Fixed	11.33	44.5	26.2	24.39	0.0	23.6	17.12	0.1	23.6
	pan Random	5.18	96.4	18.1	22.16	0.0	18.5	10.51	19.7	17.4
	Amelia TS	17.37	1.7	26.8	26.20	0.0	25.4	20.60	0.0	25.7
	Amelia CS	22.81	0.0	32.2	25.50	0.0	24.4	22.48	0.0	26.4
	Amelia TSCS	5.35	**95.9**	23.1	23.04	0.0	26.5	10.48	22.4	19.9
	MINTS	**1.22**	100.0	**4.2**	**3.08**	**99.8**	**5.9**	**1.48**	**100.0**	**3.4**
80%	MICE PMM	65.74	0.0	59.5	38.43	0.0	58.5	66.21	0.0	55.2
	pan Fixed	39.27	1.7	50.1	46.31	0.0	38.8	48.62	0.0	45.2
	pan Random	12.16	64.8	53.0	38.72	0.0	44.8	24.48	2.8	53.0
	Amelia TS	50.51	1.2	51.5	62.36	0.0	**34.9**	54.86	0.0	**34.4**
	Amelia CS	99.78	0.0	56.5	76.22	0.0	66.9	75.86	0.0	48.9
	Amelia TSCS	19.10	71.4	77.9	41.71	0.0	64.5	29.87	10.7	74.6
	MINTS	**7.53**	**82.0**	**39.9**	**10.74**	**59.8**	45.7	**9.54**	**66.4**	37.5

MAR 80% experiment, pan Fixed Effects has the smallest MAE and MINTS has the second smallest MAE.

Table C.2: Mean absolute error for the analysis model validation for nonlinear simulated data for $Q = \sigma^2_\lambda$, the variance of the random intercepts in the random intercept model of Z on Y. Results are averaged over the 1000 replications of each experiment. The true value of Q is 13.615.

Simulated Missingness Rate	Method	**MCAR**	**MAR**	**MNAR**
10%	MICE PMM	0.74	0.66	0.70
	pan Fixed	0.74	0.78	0.72
	pan Random	0.62	0.97	0.62
	Amelia TS	0.79	0.62	0.67
	Amelia CS	1.08	1.09	0.86
	Amelia TSCS	0.68	1.13	0.67
	MINTS	**0.20**	**0.23**	**0.21**
40%	MICE PMM	9.09	1.43	3.65
	pan Fixed	2.87	1.83	1.71
	pan Random	2.10	5.64	2.08
	Amelia TS	10.62	4.71	6.06
	Amelia CS	8.73	2.54	3.06
	Amelia TSCS	2.24	6.76	2.22
	MINTS	**0.55**	**0.57**	**0.56**
80%	MICE PMM	87.57	46.12	66.55
	pan Fixed	21.94	**7.44**	8.83
	pan Random	15.00	19.42	12.00
	Amelia TS	74.74	54.60	52.99
	Amelia CS	73.50	24.74	36.57
	Amelia TSCS	30.03	52.60	36.35
	MINTS	**5.44**	7.91	**4.24**

C.4.2 Enrollment Data

For the secondary school enrollment data, the second model used for analysis model validation is the random intercept model of TFR on NER given by

$$\begin{aligned}\text{TFR}_{c,t} =& \psi_0 + \psi_c + \psi_1 \text{NER}_{c,t} + \varepsilon^{\psi}_{c,t},\\ \psi_c \sim& N(0, \sigma^2_{\psi}),\\ \varepsilon^{\psi}_{c,t} \sim& N(0, \sigma^2_{\varepsilon_{\psi}}).\end{aligned}$$

The parameter of interest in the fixed component of the model is the coefficient ψ_1, while the parameter of interest in the random component of the model is the variance of the random intercepts σ^2_{ψ}. For each replication of each experiment, the analysis model is estimated using only the country-years that were simulated as missing and where TFR > 2.5.

Table C.3 summarizes the results of the analysis model validation for estimation of $Q = \psi_1$, the fixed effect coefficient on NER in the random intercept model. MINTS has the best performance for estimation of ψ_1 at the 10% rate, with the smallest MAE, smallest FMI, and closest to nominal coverage. For most of the experiments at the 40% and 80% rates, pan Random Effects has the best performance for estimation of ψ_1 with the smallest MAE and closest to nominal coverage. The one exception is MNAR 40%, where the pan Fixed Effects method outperforms pan Random Effects. MINTS has the second or third smallest MAE in the 40% and 80% experiments for estimation of ψ_1, but suffers from undercoverage.

Table C.4 summarizes the results of the analysis model validation for point estimation of $Q = \sigma^2_{\psi}$, the variance of the random intercepts. We find MINTS results in the smallest MAE across all experiments.

C.5 Simulation Study for Linear Data

We conducted a simulation study to evaluate how well the multiply imputed data sets from the proposed imputation method perform for estimation of quantities of interest Q,

Table C.3: Summary of analysis model validation for enrollment data for $Q = \psi_1$, the fixed effect coefficient on NER in the random intercept model of TFR on NER. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 100 replications of each experiment. The value of Q estimated using the observed country-years from the full enrollment data set is -0.058.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	1.66	5	36.2	1.98	0	42.6	7.01	0	26.4
	pan Fixed	0.23	100	17.4	0.18	**100**	17.7	2.63	0	34.8
	pan Random	0.18	100	14.8	0.20	**100**	16.6	1.96	0	48.5
	Amelia TS	1.42	19	34.9	1.74	3	38.5	6.96	0	27.6
	Amelia CS	0.66	87	20.5	0.89	70	23.0	5.07	0	42.3
	Amelia TSCS	0.20	**99**	22.6	0.23	**100**	23.8	4.23	0	86.0
	MINTS	**0.13**	100	**8.2**	**0.14**	**100**	**7.1**	**0.66**	**67**	**18.6**
40%	MICE PMM	4.09	0	40.0	4.23	0	45.0	5.32	0	44.9
	pan Fixed	0.56	50	28.6	0.28	**93**	29.1	**0.13**	**100**	**38.6**
	pan Random	**0.17**	**98**	36.2	**0.17**	100	41.1	1.20	43	72.1
	Amelia TS	3.83	0	38.0	4.00	0	41.2	5.19	0	44.3
	Amelia CS	2.27	0	35.9	2.44	0	34.2	5.19	0	44.3
	Amelia TSCS	0.22	100	59.6	0.25	100	66.5	3.31	10	91.8
	MINTS	0.30	83	**17.2**	0.27	90	**18.6**	0.74	2	49.8
80%	MICE PMM	4.87	0	61.4	5.03	0	63.1			
	pan Fixed	1.36	0	51.3	0.82	15	54.4			
	pan Random	**0.36**	**93**	69.6	**0.23**	**100**	72.8			
	Amelia TS	4.70	0	50.8	4.85	0	56.4			
	Amelia CS	4.09	0	47.2	3.91	0	**44.7**			
	Amelia TSCS	1.48	30	86.9	2.34	8	88.3			
	MINTS	0.55	17	**44.5**	0.50	22	45.0			

where the quantities considered are parameters from analysis models that are uncongenial to the imputation model. We refer to this validation exercise as "analysis model validation."

Analysis model validation was conducted for two simulated data sets: one with a nonlinear relationship between $\mathbf{X}$ and $\mathbf{Y}$ and one with a linear relationship. For each simulated data

Table C.4: Mean absolute error for the analysis model validation for enrollment data for $Q = \sigma^2_\psi$, the variance of the random intercepts in the random intercept model of TFR on NER. Results are averaged over the 100 replications of each experiment. The value of Q estimated using the observed country-years from the full enrollment data set is 1.004.

Simulated Missingness Rate	Method	**MCAR**	**MAR**	**MNAR**
10%	MICE PMM	0.12	0.12	0.51
	pan Fixed	0.09	0.09	0.20
	pan Random	0.05	0.06	0.60
	Amelia TS	0.13	0.12	0.50
	Amelia CS	0.11	0.13	0.50
	Amelia TSCS	0.08	0.13	0.67
	MINTS	**0.03**	**0.03**	**0.04**
40%	MICE PMM	0.20	0.20	0.52
	pan Fixed	0.18	0.15	0.37
	pan Random	0.07	0.11	2.23
	Amelia TS	0.17	0.19	0.49
	Amelia CS	0.15	0.15	0.53
	Amelia TSCS	0.17	0.32	0.99
	MINTS	**0.06**	**0.04**	**0.12**
80%	MICE PMM	0.41	0.41	
	pan Fixed	0.21	0.21	
	pan Random	1.05	1.36	
	Amelia TS	0.38	0.39	
	Amelia CS	0.12	0.12	
	Amelia TSCS	0.28	0.32	
	MINTS	**0.11**	**0.09**	

set, we simulated missingness following three rates of simulated missingness (10%, 40%, 80%) and three missing data mechanisms for a total of nine experiments. Each experiment was replicated $N_{rep} = 1000$ times and the average performance of the proposed imputation method was compared with the performance of existing multiple imputation methods for hierarchical time series data. This section details the simulation study using linear simulated data.

C.5.1 Data Generation

Variables $\mathbf{X}$ and $\mathbf{Y}$ are simulated following a hierarchical time series structure where the grouping variable is country and the time variable is year. We simulated data for 20 countries and 30 years for a total sample size of 600 country-year pairs. We assume $\mathbf{Y}$ is the variable of interest for substantive analyses, while $\mathbf{X}$ is an auxiliary variable that is only of interest for imputation of $\mathbf{Y}$.

Values of $\mathbf{X}$ were simulated independently for each country as follows. $\mathbf{X}$ is assumed to be generally increasing over time and is truncated to be bounded in $[0, 100]$. For country c,

$$\begin{aligned} X_{c,1} &\sim U(X_{1,low}, X_{1,up}) \\ X_{c,t+1} &\sim TN_{[0,100]}(X_{c,t} + \gamma_c, \sigma^2) \text{ for } t = 1, ..., 29 \end{aligned}$$

where $TN_{[0,100]}$ refers to the truncated normal distribution with support $[0, 100]$ and γ_c is a country-specific drift term. For the linear simulated data, $\mathbf{X}$ is simulated using $X_{1,low} = 2$, $X_{1,up} = 20$, $\sigma^2 = 2$, and $\gamma_c \sim N(2, 0.5^2)$.

$\mathbf{Y}$ was simulated to have a linear relationship with $\mathbf{X}$ and to be bounded as $Y_{c,t} \in [0, \min(X_{c,t}, 100)]$. For country c,

$$\begin{aligned} Y_{c,t} &\sim TN_{[0,\min(X_{c,t},100)]}\left(\alpha_c + 0.75X_{c,t}, 1^2\right) \\ \alpha_c &\sim U(0, 5) \end{aligned}$$

where $TN_{[0,\min(X_{c,t},100)]}$ refers to the truncated normal distribution with support $[0, \min(X_{c,t}, 100)]$.

C.5.2 Analysis Models

For the analysis model validation exercises, we focus on the setting where the analysis model is uncongenial to the imputation model. The same methodology is used to construct the outcome variable and the analysis models for the nonlinear and linear simulated data

experiments. We simulated a variable $\mathbf{Z}$ that acts as the outcome variable for the analysis models. $\mathbf{Z}$ is simulated to have a linear relationship with $\mathbf{Y}$ as follows:

$$\begin{aligned} Z_{c,t} &\sim N(\eta_c + 2Y_{c,t}, 10^2) \\ \eta_c &\sim U(0, 15). \end{aligned}$$

We considered two analysis models to model the relationship between $\mathbf{Z}$ and $\mathbf{Y}$. The first analysis model is the linear regression of $\mathbf{Z}$ on $\mathbf{Y}$, where the parameter of interest for estimation is ω_1, the coefficient on $\mathbf{Y}$ in the regression

$$\begin{aligned} Z_{c,t} &= \omega_0 + \omega_1 Y_{c,t} + \varepsilon_{c,t} \\ \varepsilon_{c,t} &\sim N(0, \sigma_\varepsilon^2). \end{aligned}$$

The second analysis model is the random intercept model given by

$$\begin{aligned} Z_{c,t} &= \lambda_0 + \lambda_c + \lambda_1 Y_{c,t} + \varepsilon_{c,t} \\ \lambda_c &\sim N(0, \sigma_\lambda^2) \\ \varepsilon_{c,t} &\sim N(0, \sigma_\varepsilon^2), \end{aligned}$$

where we have two parameters of interest. The parameter of interest in the fixed component of the model is the coefficient λ_1. The parameter of interest in the random component of the model is the variance of the random intercepts σ_λ^2.

C.5.3 Analysis Model Validation Procedure and Model Implementation

The same analysis model validation procedure and model implementations were used for the linear simulated data as were used for the nonlinear simulated data. The only differences occur with the bounds of the different imputation models. For the MINTS model, the bounds for the $\mathbf{Y}|\mathbf{X}$ model were set as $Y_{low} = 0$ and $Y_{up} = \min(X_{c,t}, 100)$ for the linear simulated data. For the models using the "mice" and "Amelia" R packages, the bounds were set as $\mathbf{X} \in [0, 100]$ and $\mathbf{Y} \in [0, 100]$.

C.5.4 Analysis Model Validation Results

For the linear regression analysis model validation, we evaluate how well each multiple imputation method performs for estimation of $Q = \omega_1$, the regression coefficient on $\mathbf{Y}$ in the linear regression of $\mathbf{Z}$ on $\mathbf{Y}$.

Table C.5: Summary of uncongenial analysis model validation for linear simulated data for $Q = \omega_1$, the regression coefficient on Y in the linear regression of Z on Y. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 1000 replications of each experiment. The true value of Q is 2.019.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	3.32	100.0	48.8	1.68	100.0	14.1	2.95	100.0	41.9
	pan Fixed	0.67	100.0	6.6	2.32	100.0	22.0	1.46	100.0	16.1
	pan Random	0.61	100.0	2.9	1.87	100.0	3.9	0.88	100.0	2.8
	Amelia TS	0.74	100.0	4.7	1.06	100.0	5.8	0.74	100.0	4.6
	Amelia CS	1.44	100.0	11.4	4.34	**99.9**	26.3	2.16	100.0	16.5
	Amelia TSCS	0.19	100.0	0.7	0.58	100.0	1.3	0.23	100.0	0.7
	MINTS	**0.16**	100.0	**0.6**	**0.22**	100.0	**0.8**	**0.16**	100.0	**0.6**
40%	MICE PMM	13.18	55.9	64.6	7.28	**93.7**	51.3	14.32	63.9	73.7
	pan Fixed	5.18	**95.9**	33.5	16.68	0.0	54.8	9.94	23.9	48.7
	pan Random	1.80	100.0	16.2	5.95	58.9	20.8	2.74	**100.0**	14.3
	Amelia TS	8.74	45.1	27.2	11.22	0.7	27.5	10.88	3.9	27.3
	Amelia CS	22.35	0.0	40.7	39.73	0.0	43.3	28.50	0.0	43.0
	Amelia TSCS	1.02	100.0	7.6	5.05	78.8	15.0	1.86	**100.0**	8.7
	MINTS	**0.68**	100.0	**6.4**	**1.11**	99.7	**11.8**	**0.71**	**100.0**	**6.1**
80%	MICE PMM	39.93	0.0	56.5	22.61	**30.3**	54.6	42.14	0.0	50.0
	pan Fixed	13.56	44.2	57.7	30.62	0.0	56.0	26.19	0.0	59.2
	pan Random	**5.37**	**95.1**	54.6	16.35	6.4	61.4	**9.44**	**53.9**	55.4
	Amelia TS	34.41	0.1	**50.1**	45.48	0.0	46.1	46.29	0.0	**44.5**
	Amelia CS	103.69	0.0	54.9	97.73	0.0	**38.6**	104.37	0.0	44.8
	Amelia TSCS	6.04	95.5	57.0	25.40	2.6	74.2	16.74	21.2	60.0
	MINTS	7.36	74.0	51.9	**15.82**	28.6	65.4	9.92	53.7	54.9

Table C.5 summarizes the analysis model validation results for the linear simulated data for estimation of $Q = \omega_1$. As the pan and Amelia methods assume linearity between $\mathbf{X}$ and $\mathbf{Y}$, we expect the existing methods to perform better for imputation of $\mathbf{Y}$ in the linear setting than in the nonlinear setting. Correspondingly, we see the existing methods do generally have better performance for estimation of ω_1 with the linear simulated data, with smaller MAE and better coverage compared to the results for the nonlinear simulated data. At the 10% and 40% rate of simulated missingness, the proposed MINTS method tends to perform the best in terms of MAE, coverage, and FMI, however the Amelia TSCS method is a close second best for many experiments. At the 80% rate of simulated missingness, all methods tend to perform poorly. This is unsurprising as average FMI is estimated to be above 50% for most methods in these experiments. The pan Random Effects and Amelia TSCS methods both perform well in the MCAR 80% experiment, with similar MAE and close to nominal coverage. However, all methods have substantial undercoverage in the MAR 80% and MNAR 80% experiments.

For the random intercept analysis model validation, we evaluate how well each multiple imputation method performs for estimation of two parameters. First, we consider estimation of $Q = \lambda_1$, the fixed effect coefficient on $\mathbf{Y}$. Table C.6 summarizes the analysis model validation results for estimation of the fixed effect coefficient λ_1 using the linear simulated data. At the 10% and 40% rates of simulated missingness, MINTS consistently has the smallest MAE, the smallest FMI, and has good coverage. The Amelia TSCS method also performs well for at the 10% and 40% rates of simulated missingness, with the second smallest MAE and second smallest FMI across experiments. Amelia TSCS also generally has good coverage at these lower rates of simulated missingness, with the notable exception of the MAR 40% experiment. At the 80% rate of simulated missingness, performance of all methods deteriorates. Pan Random Effects, Amelia TSCS, and MINTS have the best performance overall in terms of MAE and coverage, however no method performs consistently well across

missing data mechanisms at this highest rate of simulated missingness.

Table C.6: Summary of uncongenial analysis model validation for linear simulated data for $Q = \lambda_1$, the fixed effect coefficient on Y in the random intercept model of Z on Y. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 1000 replications of each experiment. The true value of Q is 2.012.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	5.58	100.0	68.4	3.24	100.0	27.2	4.96	**99.9**	61.5
	pan Fixed	0.78	100.0	7.5	1.60	100.0	14.1	1.35	100.0	14.1
	pan Random	0.67	100.0	3.3	2.18	100.0	4.9	1.02	100.0	3.2
	Amelia TS	1.25	100.0	6.9	1.45	100.0	8.7	1.23	100.0	6.8
	Amelia CS	1.65	100.0	13.2	4.25	100.0	27.9	2.36	100.0	18.7
	Amelia TSCS	0.21	100.0	**0.7**	0.71	100.0	1.8	0.27	100.0	0.8
	MINTS	**0.16**	100.0	**0.7**	**0.25**	100.0	**1.1**	**0.18**	100.0	**0.7**
40%	MICE PMM	20.61	3.0	73.2	12.99	42.0	66.4	21.48	15.7	81.6
	pan Fixed	6.10	**91.7**	35.9	10.75	27.1	45.2	9.14	55.1	48.6
	pan Random	1.97	100.0	17.5	6.64	46.7	24.1	3.08	**100.0**	17.0
	Amelia TS	14.97	0.4	34.0	16.14	0.0	34.8	16.51	0.0	34.4
	Amelia CS	25.04	0.0	43.5	40.36	0.0	44.7	30.48	0.0	44.9
	Amelia TSCS	1.14	100.0	8.1	5.83	63.5	18.3	2.24	**100.0**	9.4
	MINTS	**0.71**	100.0	**6.8**	**1.14**	**99.9**	**16.1**	**0.75**	**100.0**	**7.2**
80%	MICE PMM	47.93	0.0	58.7	33.99	1.3	55.9	49.51	0.0	51.9
	pan Fixed	16.73	27.9	60.0	17.34	38.5	71.2	24.42	1.1	62.5
	pan Random	4.73	**96.4**	56.5	14.97	16.1	64.3	**8.98**	61.5	58.6
	Amelia TS	43.81	0.0	52.3	54.77	0.0	47.4	54.59	0.0	**46.5**
	Amelia CS	111.81	0.0	**50.3**	107.29	0.0	**46.9**	113.59	0.0	47.2
	Amelia TSCS	**4.29**	99.2	63.9	12.22	**67.6**	69.8	11.40	**69.1**	64.3
	MINTS	7.21	77.1	52.7	**12.02**	59.3	74.5	9.35	60.6	57.0

Finally, we also evaluate how well each multiple imputation method performs for point estimation of $Q = \sigma^2_\lambda$, the variance of the random intercepts. Table C.7 summarizes the MAE for this analysis model validation for the linear simulated data. MINTS also results in

the smallest MAE for the experiments at the 10% and 40% rates of simulated missingness. At the 80% rate of simulated missingness, MINTS has the smallest MAE for the MCAR 80% experiment, while pan Random Effects has the smallest MAE for the MAR 80% and MNAR 80% experiments.

Table C.7: Mean absolute error for the uncongenial analysis model validation for linear simulated data for $Q = \sigma^2_\lambda$, the variance of the random intercepts in the random intercept model of Z on Y. Results are averaged over the 1000 replications of each experiment. The true value of Q is 13.900.

Simulated Missingness Rate	Method	**MCAR**	**MAR**	**MNAR**
10%	MICE PMM	2.76	2.88	3.10
	pan Fixed	0.36	1.09	0.58
	pan Random	0.53	1.01	0.63
	Amelia TS	0.55	0.58	0.47
	Amelia CS	0.91	1.88	1.48
	Amelia TSCS	0.23	0.40	0.27
	MINTS	**0.20**	**0.23**	**0.20**
40%	MICE PMM	24.28	18.93	29.93
	pan Fixed	1.95	12.56	4.21
	pan Random	1.37	3.71	1.78
	Amelia TS	13.63	14.93	15.22
	Amelia CS	7.10	6.08	11.22
	Amelia TSCS	0.88	3.05	1.20
	MINTS	**0.61**	**1.46**	**0.71**
80%	MICE PMM	144.37	103.81	133.39
	pan Fixed	17.82	39.18	20.49
	pan Random	11.75	**22.69**	**16.19**
	Amelia TS	132.74	117.30	130.4
	Amelia CS	104.89	37.02	73.09
	Amelia TSCS	32.85	90.02	79.78
	MINTS	**4.56**	33.93	17.69

C.6 Congenial Analysis Model Validation for Linear Data

Although our primary focus is on the setting of uncongeniality, we also conducted the analysis model validation for two congenial analysis models for the linear simulated data for the purposes of comparing the performance of the MINTS method with the existing imputation methods in the "ideal" setting for imputation. The first congenial analysis model is the linear regression of Y on X, where the parameter we are interested in estimating is χ_1, the coefficient on X in the regression

$$\begin{aligned} Y_{c,t} =& \chi_0 + \chi_1 X_{c,t} + \varepsilon_{c,t}, \\ \varepsilon_{c,t} \sim& N(0, \sigma^2_\varepsilon). \end{aligned}$$

The second analysis model is the random intercept model of Y on X given by

$$\begin{aligned} Y_{c,t} =& \phi_0 + \phi_c + \phi_1 X_{c,t} + \varepsilon_{c,t}, \\ \phi_c \sim& N(0, \sigma^2_\phi), \\ \varepsilon_{c,t} \sim& N(0, \sigma^2_\varepsilon), \end{aligned}$$

where we have two parameters of interest. The parameter of interest in the fixed component of the model is the coefficient ϕ_1. The parameter of interest in the random component of the model is the variance of the random intercepts σ^2_ϕ. For the regression coefficient parameters of interest χ_1 and ϕ_1, we calculate MAE, mean coverage of 95% intervals, and mean FMI. For the variance of random intercepts σ^2_ϕ, we only calculate MAE.

Table C.8 summarizes the validation results for Q = regression coefficient χ_1 for the congenial linear regression of Y on X. Overall, we found no clear standout method across experiments. MINTS performs well at the 10% and 40% rates of simulated missingness, but does not have the best performance by any metric. At the 80% rate of simulated missingness, MINTS has middling performance, especially in terms of coverage. MINTS outperforms several existing imputation methods by MAE and coverage in the MCAR 80%

and MNAR 80% experiments, but has the second worst MAE and worst coverage in the MAR 80% experiment. Out of the existing imputation methods, the Amelia TSCS and Amelia CS methods tend to perform the best by MAE and FMI at the 10% and 40% rates of simulated missingness, but perform substantially worse at the 80% rate of simulated missingness. The pan Random Effects and Amelia TS methods tend to have the best performance in terms of MAE and coverage at the 80% rate of simulated missingness but do not stand out in terms of performance at the lower rates of missingness. Based on this validation exercise, we find MINTS is a competitive choice at low and moderate rates of missing data. However, some of the existing imputation methods are likely preferable at high rates of missingness when the linearity assumption is satisfied and the analysis model of interest is congenial to the imputation model.

Tables C.9 and C.10 summarize the validation results for $Q =$ the fixed effect coefficient ϕ_1 and $Q =$ the variance of the random intercepts σ^2_ϕ, respectively, from the random intercept model of Y on X. Overall, we found no clear standout model across experiments for either the fixed effect coefficient or the variance of the random intercepts. MINTS does well for estimation of both parameters at the 10% and 40% rates of simulated missingness, but has markedly worse performance at the 80% rate of simulated missingness for the MAR and MNAR experiments. For the fixed effect coefficient, MINTS actually has the worst MAE and coverage in the MAR 80% experiment out of all methods compared. Out of the existing imputation methods, the Amelia TSCS and Amelia CS models tend to perform the best at the 10% and 40% rates of simulated missingness. At the 80% rate of simulated missingness, the pan Random Effects model tends to perform the best for estimation of fixed effect coefficient while the MICE PMM model tends to perform the best for estimation of the variance of the random intercepts. Based on this validation exercise, we find MINTS is a competitive choice at low and moderate rates of missing data. However, unless we are in the MCAR setting, most of the existing imputation models are preferable over MINTS at high rates of

Table C.8: Summary of congenial analysis model validation for linear simulated data for $Q = \chi_1$, the regression coefficient on X in the linear regression of Y on X. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 1000 replications of each experiment. The true value of Q is 0.741.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	2.103	89.2	64.0	0.640	**98.0**	38.6	1.432	90.1	43.8
	pan Fixed	0.602	**97.3**	30.6	0.851	91.0	53.1	0.995	83.4	42.3
	pan Random	0.284	99.9	19.6	0.263	99.8	22.1	0.318	**99.5**	17.9
	Amelia TS	0.135	100.0	18.2	0.145	100.0	20.2	0.119	100.0	16.0
	Amelia CS	0.072	100.0	6.0	0.086	100.0	**6.7**	**0.071**	100.0	5.4
	Amelia TSCS	**0.069**	100.0	**5.5**	**0.072**	100.0	7.0	**0.071**	100.0	**5.2**
	MINTS	0.074	100.0	6.9	0.093	100.0	7.7	0.078	100.0	6.1
40%	MICE PMM	3.104	100.0	93.2	2.845	97.9	86.6	4.231	97.6	89.4
	pan Fixed	2.117	79.3	65.8	0.899	98.9	67.5	1.697	91.6	69.8
	pan Random	0.550	99.2	51.3	0.684	**94.2**	60.9	0.715	**96.7**	51.0
	Amelia TS	0.349	**98.8**	62.4	0.360	98.8	67.0	0.295	99.4	58.2
	Amelia CS	**0.200**	99.9	36.3	**0.206**	100.0	**34.4**	**0.183**	100.0	33.9
	Amelia TSCS	0.220	100.0	43.8	0.215	100.0	50.0	0.220	100.0	48.9
	MINTS	0.223	100.0	**33.9**	0.450	92.4	39.7	0.237	99.4	**31.2**
80%	MICE PMM	5.765	83.7	92.4	3.660	86.1	92.3	2.498	91.7	89.5
	pan Fixed	2.147	97.6	93.6	2.100	92.1	92.5	3.121	84.6	90.4
	pan Random	**1.258**	97.6	87.6	**1.463**	96.7	91.8	1.488	**95.3**	87.3
	Amelia TS	1.435	**94.4**	96.2	1.557	**95.6**	97.3	**1.258**	95.4	96.0
	Amelia CS	10.221	80.7	93.6	3.668	97.8	93.8	4.837	93.7	90.8
	Amelia TSCS	5.829	86.2	90.7	6.372	77.9	88.9	10.454	26.7	85.4
	MINTS	1.561	88.3	**75.9**	4.284	54.5	**76.3**	2.373	77.3	**69.6**

missingness when the linearity assumption is satisfied and the analysis model of interest is congenial to the imputation model.

Table C.9: Summary of congenial analysis model validation for linear simulated data for $Q = \phi_1$, the fixed effect coefficient on X in the random intercept model of Y on X. MAE denotes mean absolute error, Cvg denotes the average coverage of 95% intervals as a percentage, and FMI denotes the fraction of missing information as a percentage. MAE is multiplied by 100 before reporting. Results are averaged over the 1000 replications of each experiment. The true value of Q is 0.755.

Simulated Missingness Rate	Method	**MCAR**			**MAR**			**MNAR**		
		MAE	Cvg	FMI	MAE	Cvg	FMI	MAE	Cvg	FMI
10%	MICE PMM	2.948	82.5	78.9	0.862	97.2	68.9	2.033	**94.8**	77.8
	pan Fixed	0.873	85.5	39.9	0.739	91.8	47.6	1.061	75.0	42.8
	pan Random	0.402	**97.5**	31.3	0.400	**94.8**	38.3	0.462	94.7	30.7
	Amelia TS	0.297	**97.5**	31.9	0.276	98.4	41.4	0.291	98.0	31.9
	Amelia CS	**0.081**	100.0	**18.4**	0.117	99.9	**24.4**	0.083	100.0	**18.4**
	Amelia TSCS	**0.081**	100.0	**18.4**	**0.097**	99.9	27.2	**0.082**	100.0	18.8
	MINTS	0.093	100.0	19.5	0.132	100.0	27.0	0.102	100.0	19.2
40%	MICE PMM	4.420	99.6	92.7	4.297	91.8	89.5	5.430	**93.6**	91.1
	pan Fixed	2.792	67.8	72.0	1.904	83.8	72.4	2.097	87.9	76.4
	pan Random	0.802	**95.6**	62.3	0.985	83.9	72.2	1.034	83.8	62.4
	Amelia TS	0.978	65.6	65.0	0.820	86.3	73.8	0.933	69.8	64.8
	Amelia CS	**0.224**	98.0	65.7	0.282	**97.1**	70.6	**0.236**	96.8	64.6
	Amelia TSCS	0.257	99.0	74.7	**0.261**	99.4	80.3	0.246	99.5	75.1
	MINTS	0.273	99.1	**51.4**	0.562	90.7	**65.5**	0.305	97.7	**52.3**
80%	MICE PMM	6.989	70.9	92.6	4.437	79.9	92.5	3.397	82.0	90.0
	pan Fixed	2.638	97.6	95.6	3.235	92.5	96.7	3.200	88.2	94.3
	pan Random	**1.537**	**95.9**	93.0	**1.757**	**95.5**	95.3	1.848	**93.2**	92.7
	Amelia TS	1.756	86.8	96.3	2.130	87.3	97.3	1.779	86.1	96.1
	Amelia CS	1.829	99.2	98.3	1.976	97.6	98.3	**1.458**	98.1	97.4
	Amelia TSCS	5.101	93.0	93.4	4.144	93.2	93.1	6.040	86.8	90.0
	MINTS	1.666	89.7	**80.8**	9.839	28.5	**89.9**	3.752	59.3	**79.1**

C.7 Model Implementation for Out-of-Sample Validation Exercise

In the out-of-sample validation exercise for the enrollment data, we evaluated the predictive performance of each multiple imputation method for imputing the individual values of NER that were simulated as missing.

Table C.10: Mean absolute error for the congenial analysis model validation for linear simulated data for $Q = \sigma^2_\phi$, the variance of the random intercepts in the random intercept model of Y on X. Results are averaged over the 1000 replications of each experiment. The true value of Q is 2.725.

Simulated Missingness Rate	Method	**MCAR**	**MAR**	**MNAR**
10%	MICE PMM	0.478	0.538	0.373
	pan Fixed	0.216	0.308	0.237
	pan Random	0.136	0.161	0.143
	Amelia TS	0.915	0.890	0.859
	Amelia CS	0.057	**0.051**	0.057
	Amelia TSCS	**0.055**	0.060	**0.053**
	MINTS	0.060	0.070	0.064
40%	MICE PMM	1.772	1.392	0.659
	pan Fixed	0.490	0.347	0.525
	pan Random	0.259	0.437	0.315
	Amelia TS	2.341	2.211	2.219
	Amelia CS	0.203	**0.168**	0.231
	Amelia TSCS	0.188	0.267	0.243
	MINTS	**0.187**	0.234	**0.206**
80%	MICE PMM	1.590	**2.310**	**1.969**
	pan Fixed	1.858	2.393	5.260
	pan Random	3.713	2.399	3.613
	Amelia TS	2.694	2.688	2.680
	Amelia CS	77.320	19.167	48.992
	Amelia TSCS	14.058	30.787	41.399
	MINTS	**1.374**	10.042	4.635

Samples from the posterior predictive distribution of the missing values from the MINTS method were obtained using a modification of the imputation procedure. The estimation phase was run with five chains for sufficient iterations to achieve convergence, where the total number of iterations per chain differed across experiments and ranged from 5000 to 35000 iterations after burn-in. Model bounds were set as $X_{low} = 0$, $X_{up} = \infty$, $Y_{low} = 0$, and $Y_{up} = \min(X_{c,t}, 100)$. In the imputation phase, an additional 5000 iterations was run for

each chain. Instead of selecting a fixed number of iterations to act as the imputed values, all iterations from the imputation phase were pooled together and taken as samples from the posterior predictive distribution of the imputed values given the observed values.

For the MICE models, we specified bounds of $\mathbf{Y} \in [0, 100]$ and $\mathbf{X} \in [0, \infty)$ and the forms of the imputation models as described in the main text, but otherwise used the default settings in the mice package. The MICE algorithm was used to create five multiply imputed data sets of 5000 iterations each. Iterations from all multiple imputations were pooled together and taken as samples from the posterior predictive distribution of the imputed values given the observed values.

For the Amelia models, we specified bounds of $\mathbf{Y} \in [0, 100]$ and $\mathbf{X} \in [0, \infty)$ and specified the forms of the imputation models as described in the main text, but otherwise used the default settings in the Amelia package. To obtain estimates of the posterior predictive distributions of the imputed values, we used a modification of the method used within the Amelia package to construct confidence intervals for individual imputed values. The Amelia algorithm bootstraps the data M times and uses an EM algorithm to estimate the parameters of the complete data likelihood for each bootstrapped data set. By default, $d = 1$ draw of the imputed values is taken from the estimated complete data likelihood for each of the bootstrapped samples. Confidence intervals for imputed values are created by taking $d > 1$ draws of the imputed values and using quantiles of the distribution of draws as the interval bounds. We used a combination of $M = 100$ bootstrapped samples and $d = 50$ draws per bootstrapped sample, where the large number of M was used to ensure uncertainty about the imputation model parameters was captured. The draws were pooled together and taken as samples from the predictive distribution of the imputed values given the observed values.

For all models, medians from the estimated posterior predictive distributions were used as point estimates for the imputed values and the 0.025 and 0.975 quantiles of the estimated posterior predictive distributions were used as 95% interval estimates.

C.8 Comparison of MAE Across Validation Exercises

We summarize the average MAE across experiments for each imputation model within each validation exercise in Table C.11. All validation exercises except the congenial analysis model validation exercises for the linear simulated data are summarized in Table C.11. We note that this summary should not be interpreted as inferential and is only of interest as an exploratory comparison tool to enable a simple comparison of one metric from the full evaluation results.

Table C.11: Average mean absolute error (MAE) across all experiments for each multiple imputation method within each validation exercise. For the enrollment data, OOS denotes to the out-of-sample validation, β_1 is the parameter from the linear regression analysis model validation, and ψ_1 and σ^2_ψ are the parameters from the random intercept analysis model validation. For the nonlinear and linear simulated data, ω_1 is the parameter from the linear regression analysis model validation and λ_1 and σ^2_λ are the parameters from the random intercept analysis model validation. MAE for the β_1, ψ_1, ω_1, and λ_1 columns are multiplied by 100 before reporting.

Method	Enrollment Data				Nonlinear Data			Linear Data		
	OOS	β_1	ψ_1	σ^2_ψ	ω_1	λ_1	σ^2_λ	ω_1	λ_1	σ^2_λ
MICE PMM	14.68	2.97	4.28	0.31	20.16	23.25	24.06	16.38	22.26	51.49
pan Fixed	13.90	0.52	0.77	0.19	23.18	22.08	5.21	11.85	9.80	10.92
pan Random	5.36	1.48	0.56	0.69	14.30	13.64	6.50	5.00	4.91	**6.63**
Amelia TS	15.70	2.59	4.08	0.30	22.76	27.14	22.87	17.73	22.75	47.32
Amelia CS	9.89	2.15	3.07	0.23	37.55	37.30	16.91	44.92	48.54	27.07
Amelia TSCS	7.49	2.04	1.53	0.37	17.89	15.50	14.74	6.35	4.26	23.19
MINTS	**2.77**	**0.18**	**0.41**	**0.07**	**4.15**	**3.87**	**2.21**	**4.02**	**3.53**	**6.62**

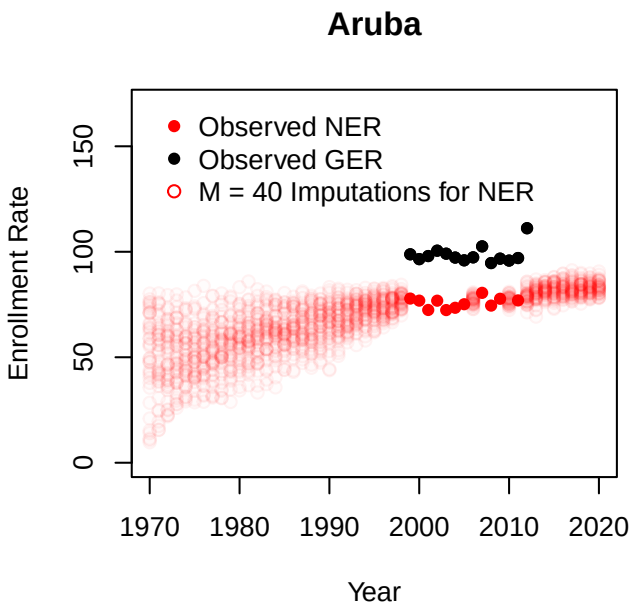
MINTS consistently results in the smallest average MAE across experiments in each validation exercise, with the largest gains in performance compared to the existing methods occurring for the enrollment data and the nonlinear simulated data experiments. Out of the existing imputation methods, we found the models using the pan methodology tend to have the smallest average MAE. For the validation exercises using enrollment data, the pan

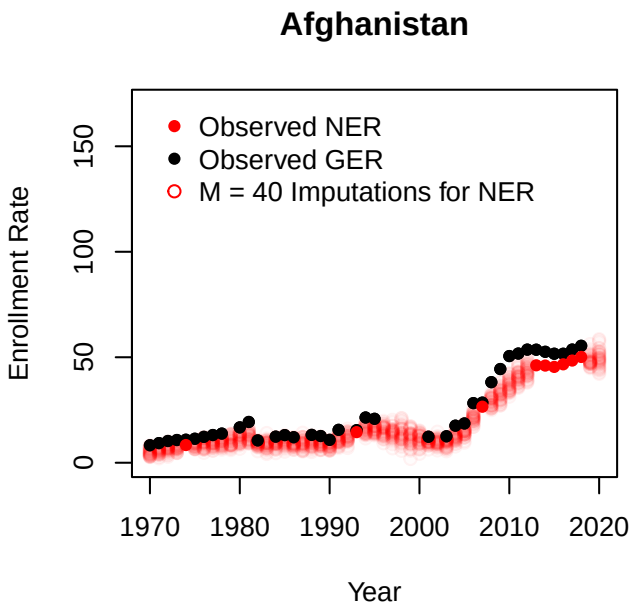
Random Effects method tends to perform the best out of the existing methods, with the second smallest average MAE for the out-of-sample validation exercise and the second or third smallest average MAE for the analysis model validation estimation of β_1 and ψ_1. The pan Fixed Effects method also performs well in terms of average MAE for analysis model validation with the enrollment data, but has much larger average MAE for the out-of-sample validation exercise. For the validation exercises using the nonlinear simulated data, the pan Random Effects method has the smallest average MAE out of the existing imputation methods for estimation of ω_1 and λ_1, while the pan Fixed Effects method has the smallest average MAE out of the existing methods for estimation of σ^2_λ.

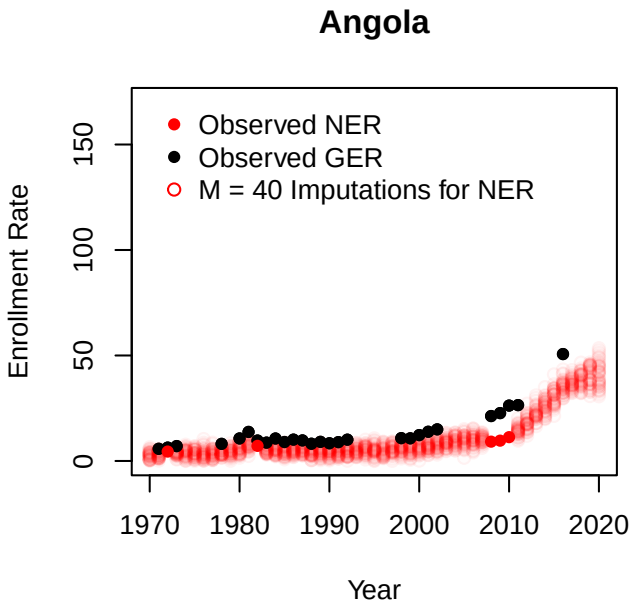
In the setting of the linear simulated data, we expect the existing methods to fare better due to the assumption of linearity inherent in the existing imputation models. Although MINTS still results in the smallest average MAE for the linear simulated data experiments, pan Random Effects follows closely behind. This is especially noticeable for estimation of σ^2_λ, the variance of the random effects, where the average MAE for pan Random Effects is only 0.006 larger than the average MAE for MINTS. The Amelia TSCS method also performs well for estimation of regression coefficients in the linear simulated data experiments, with the second smallest average MAE for the estimation of λ_1.

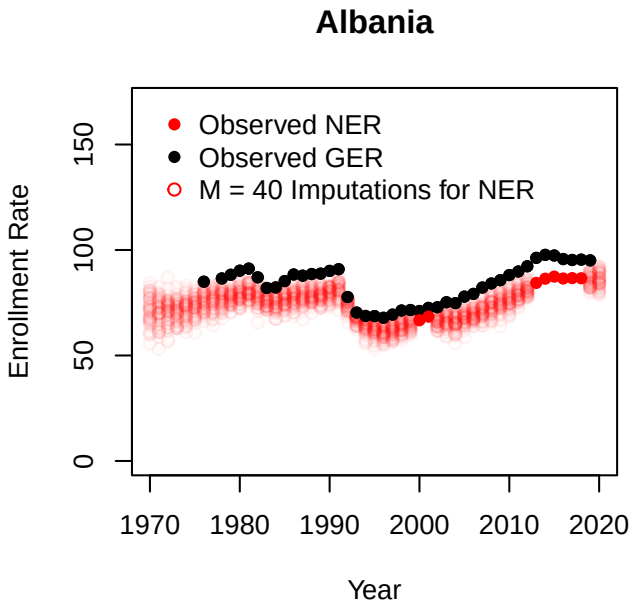
C.9 Multiple Imputations for NER

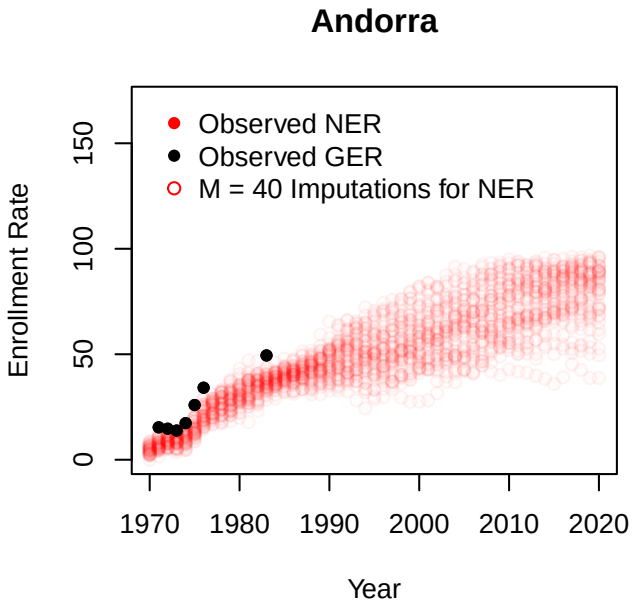
We used MINTS to create 40 multiple imputations for the missing country-years in the original enrollment data set without simulating any additional country-years as missing. The figures below show the multiple imputations for NER as translucent red circles along with the observed values of NER and GER from the original data set in solid red and black circles, respectively.

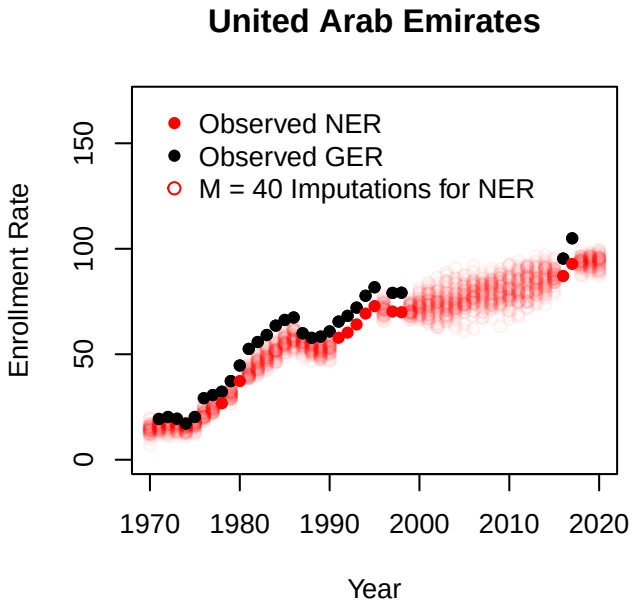

Aruba
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

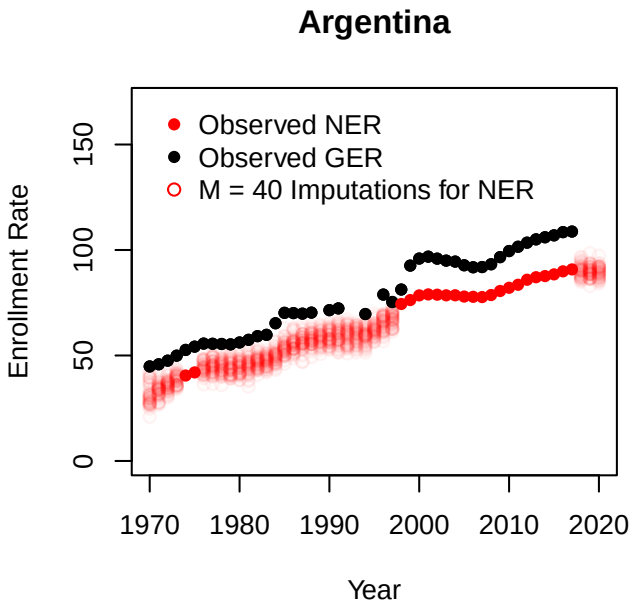

Afghanistan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

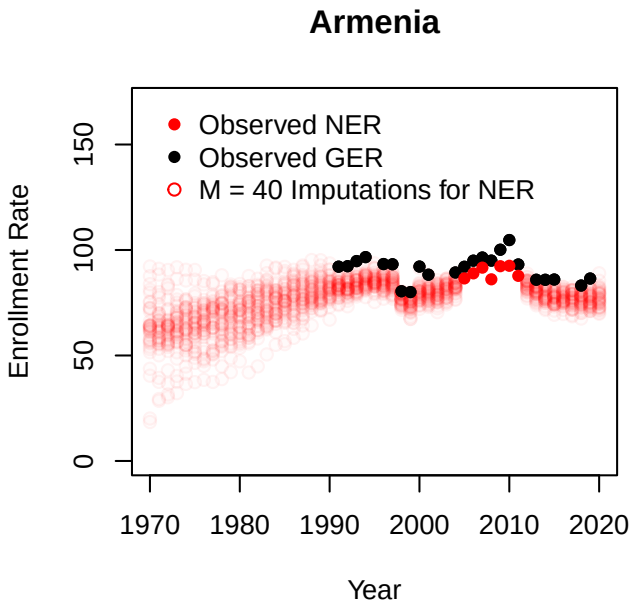

Angola
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

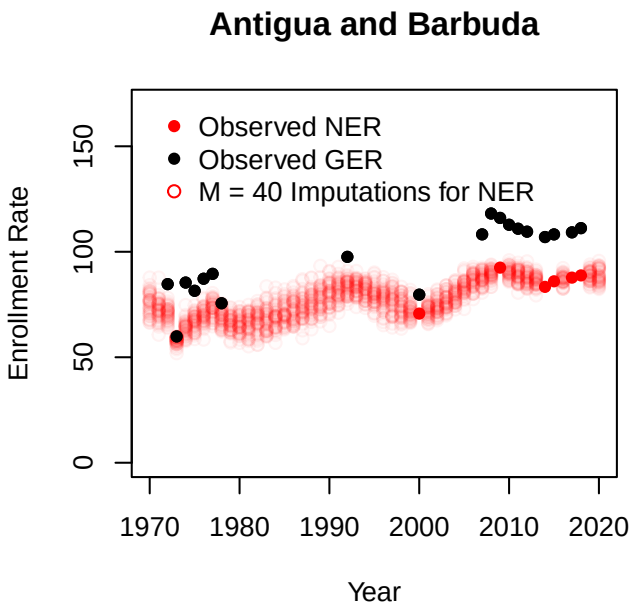

Albania
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

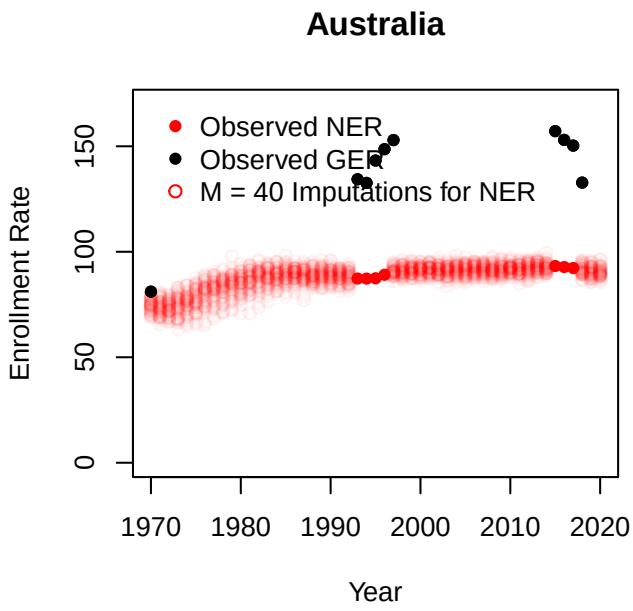

Andorra
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

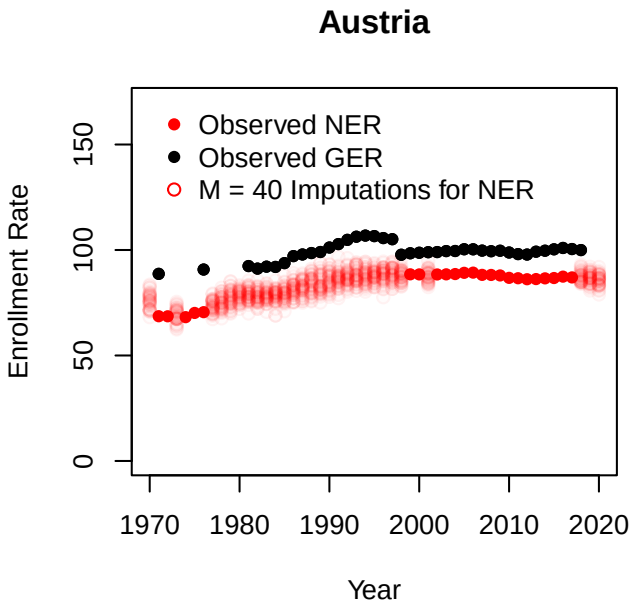

United Arab Emirates
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

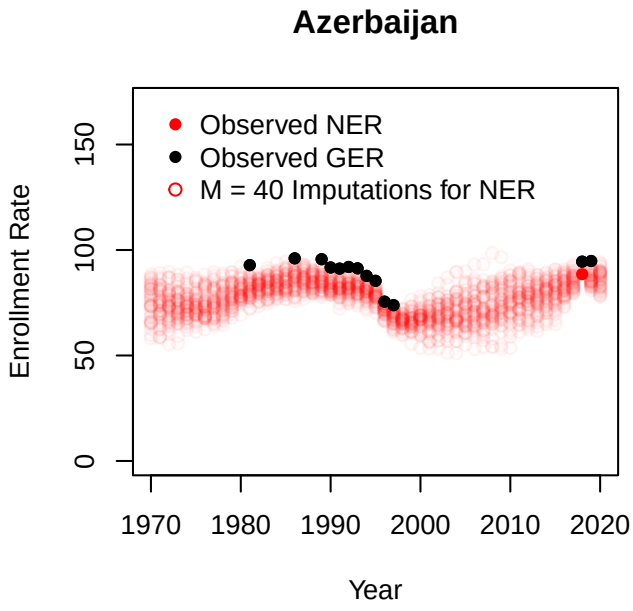

Argentina
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

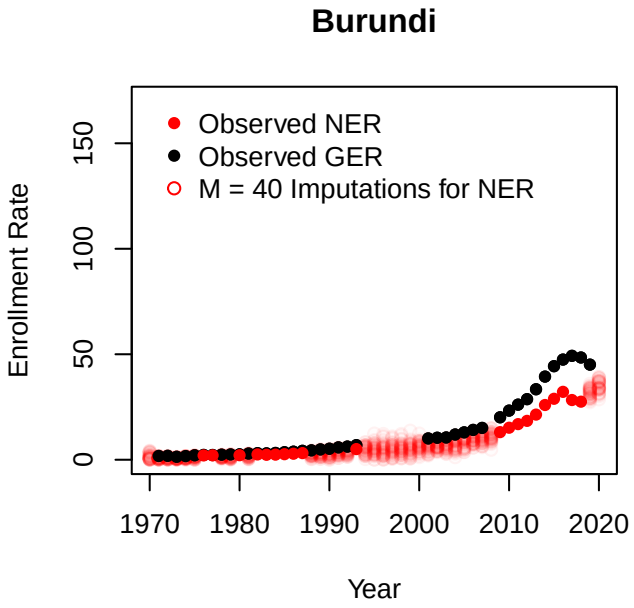

Armenia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

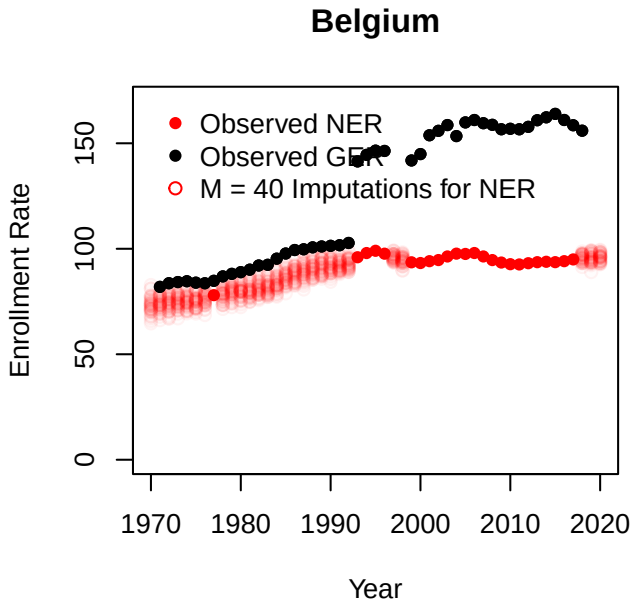

Antigua and Barbuda
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

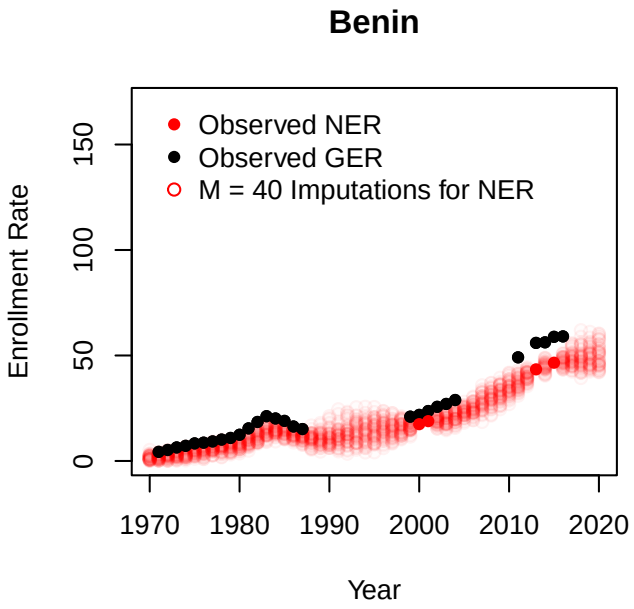

Australia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

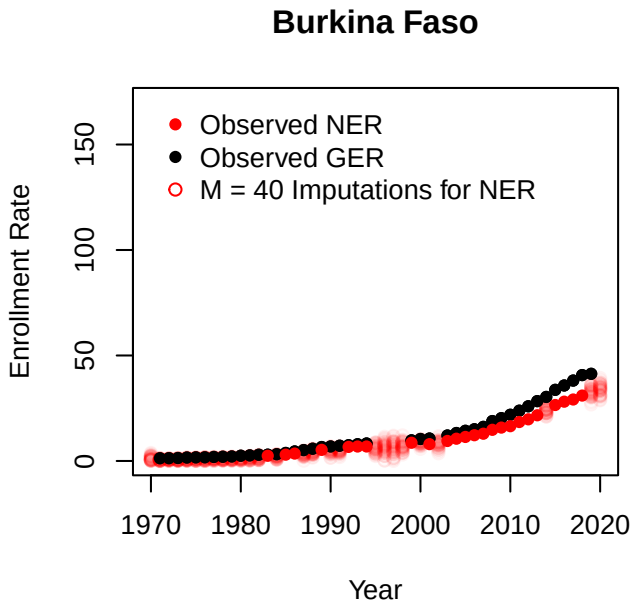

Austria
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

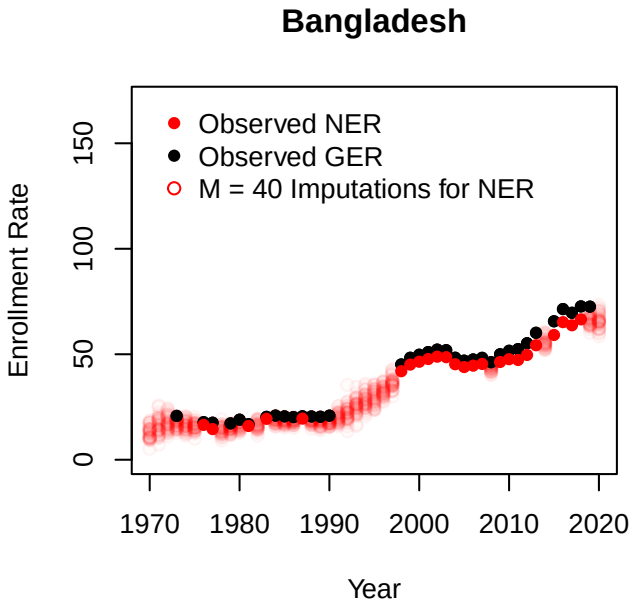

Azerbaijan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

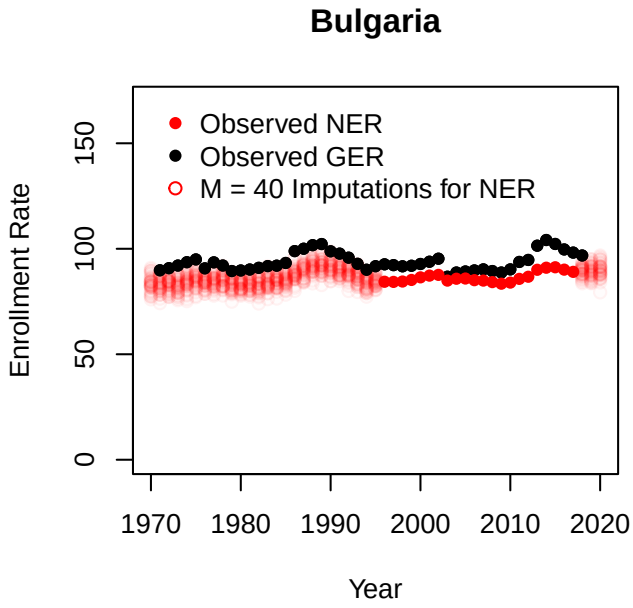

Burundi
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

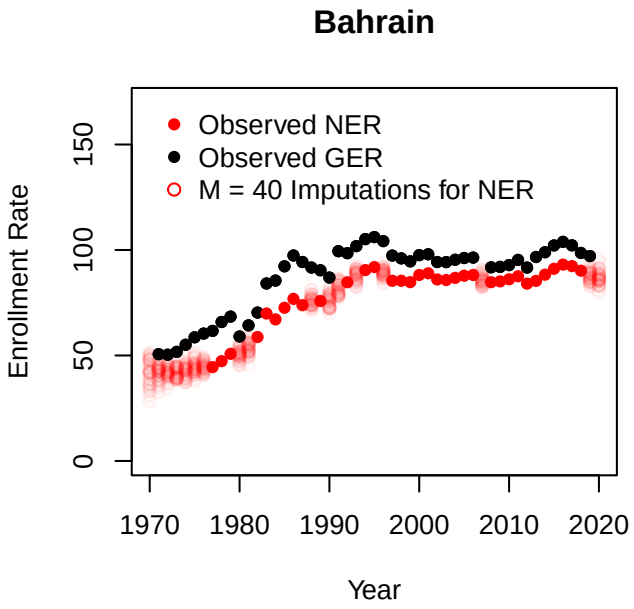

Belgium
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

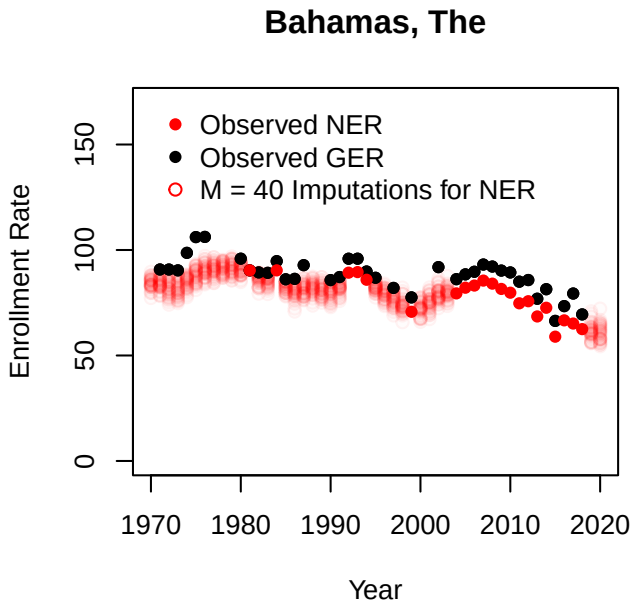

Benin
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

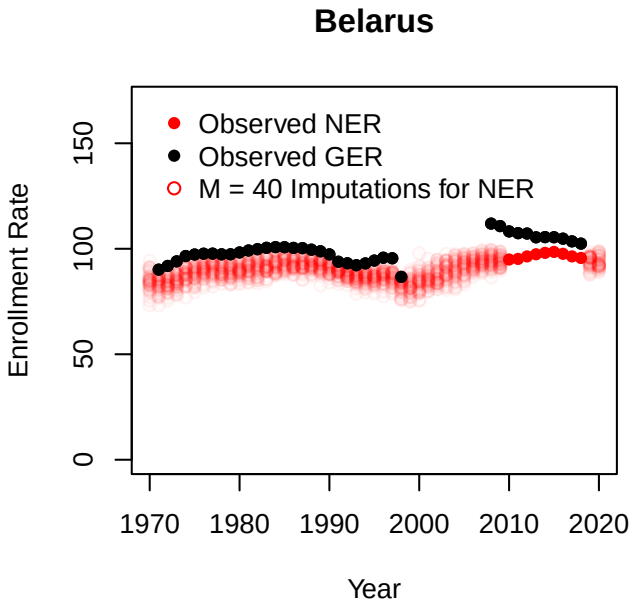

Burkina Faso
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

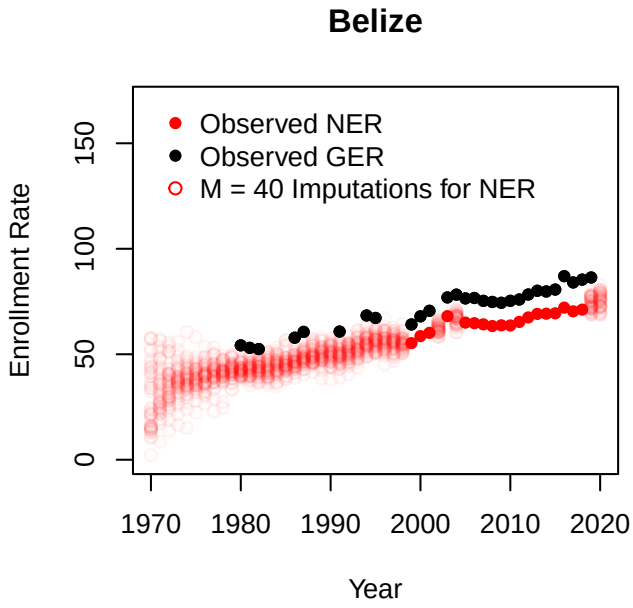

Bangladesh
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

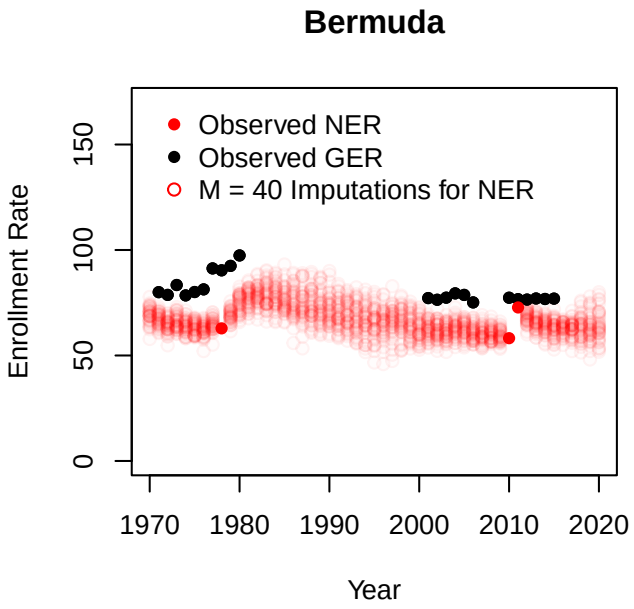

Bulgaria
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

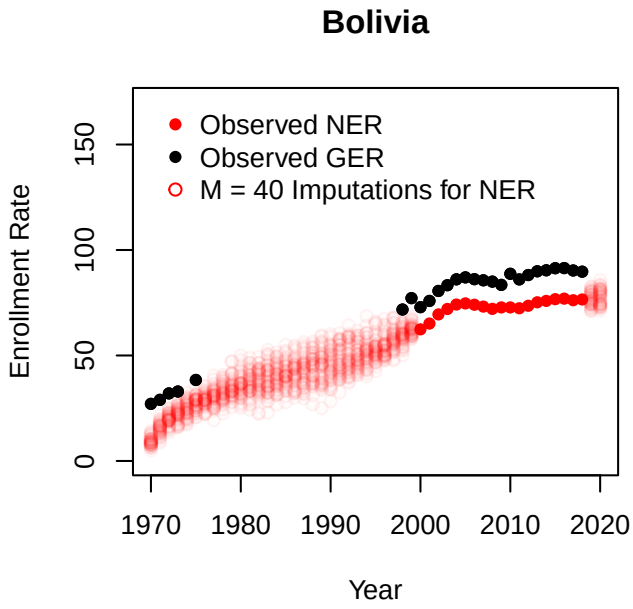

Bahrain
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

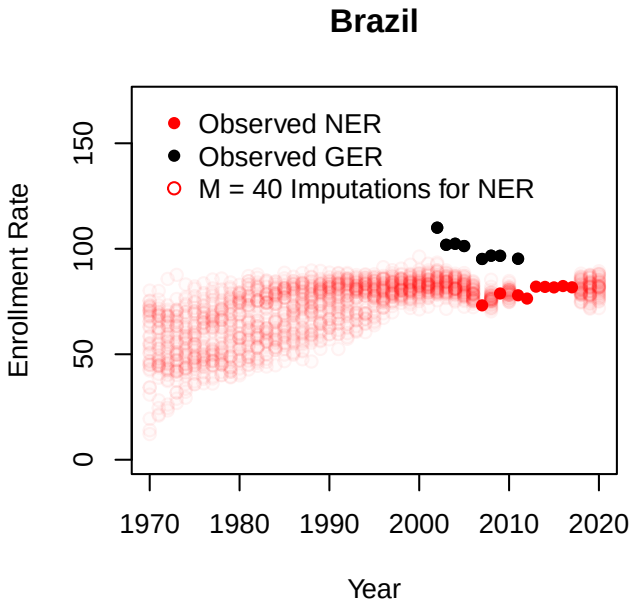

Bahamas, The
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

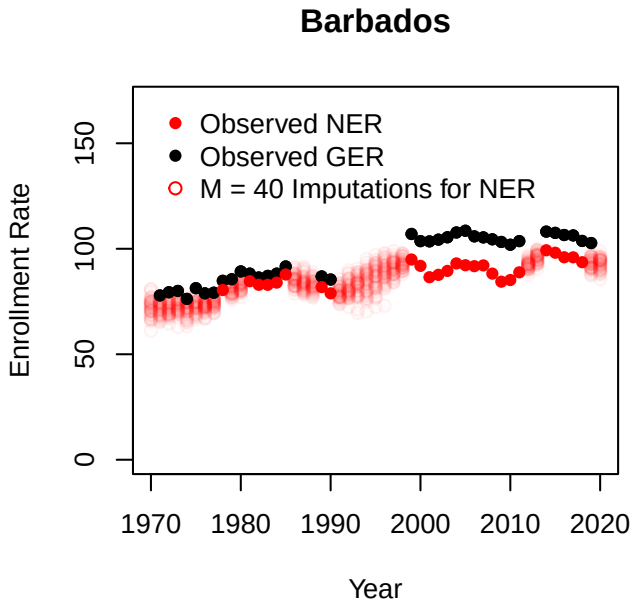

Belarus
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

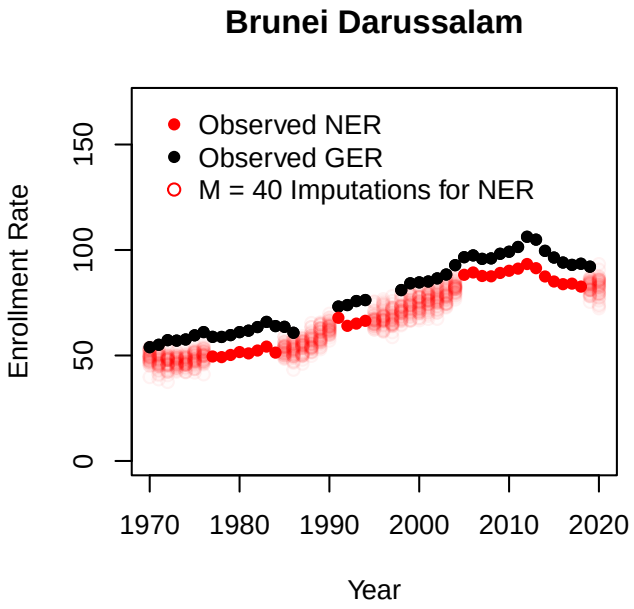

Belize
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

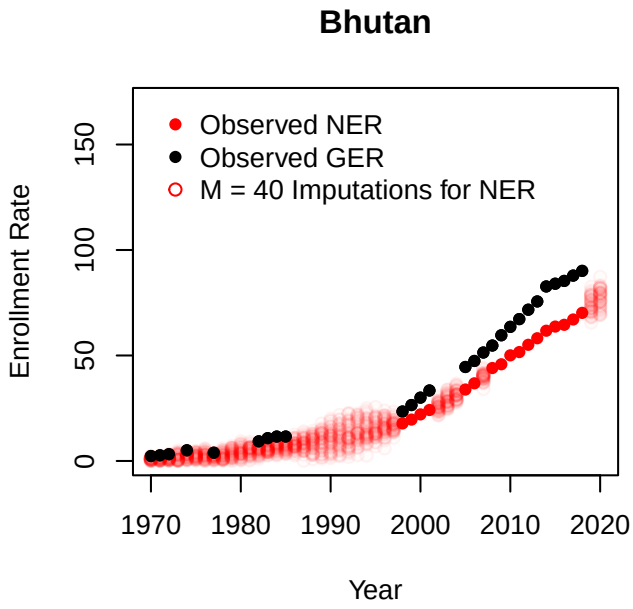

Bermuda
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

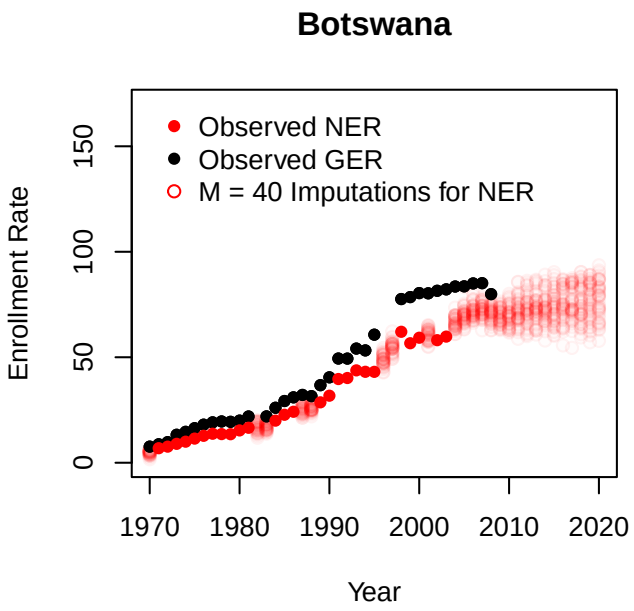

Bolivia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

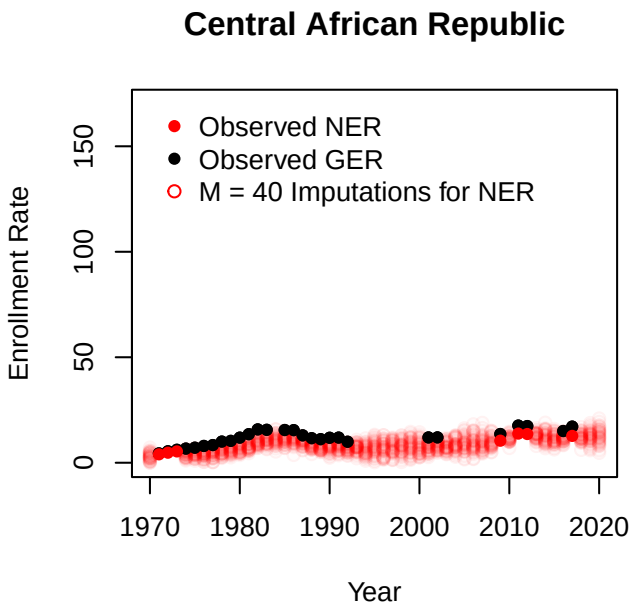

Brazil
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

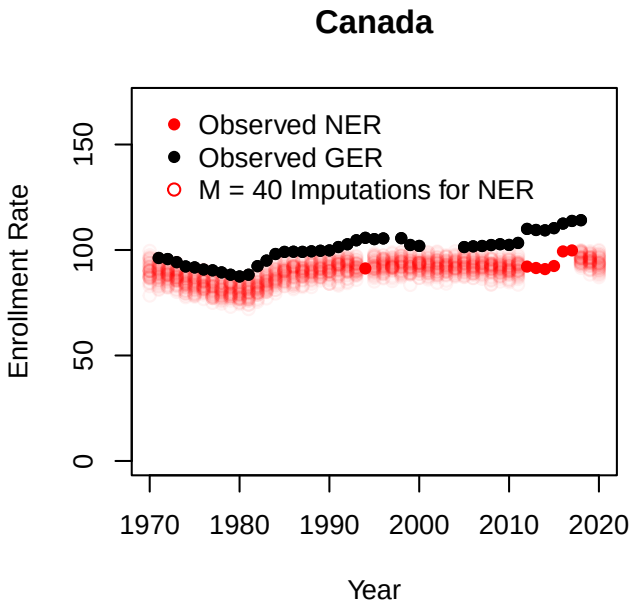

Barbados
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

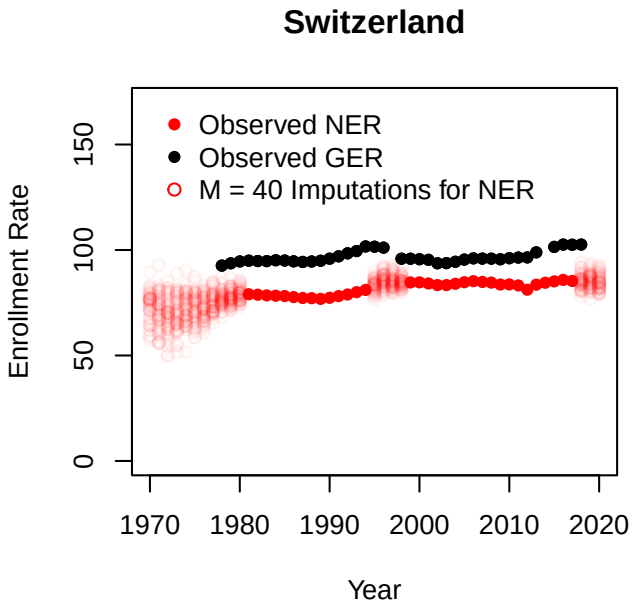

Brunei Darussalam
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

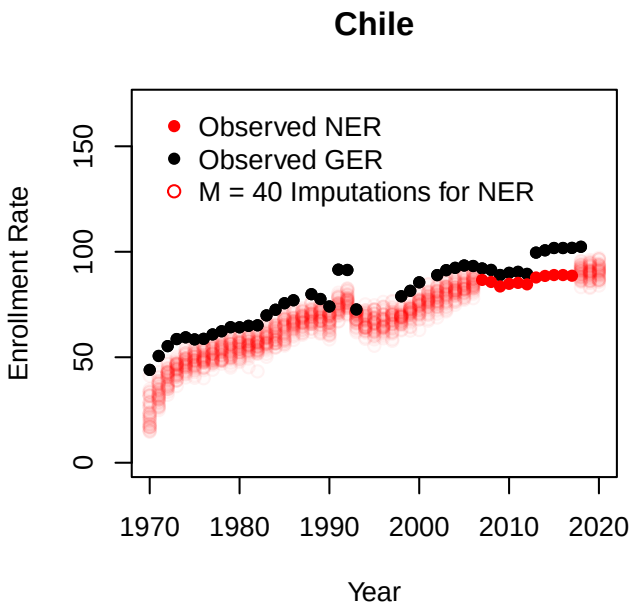

Bhutan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

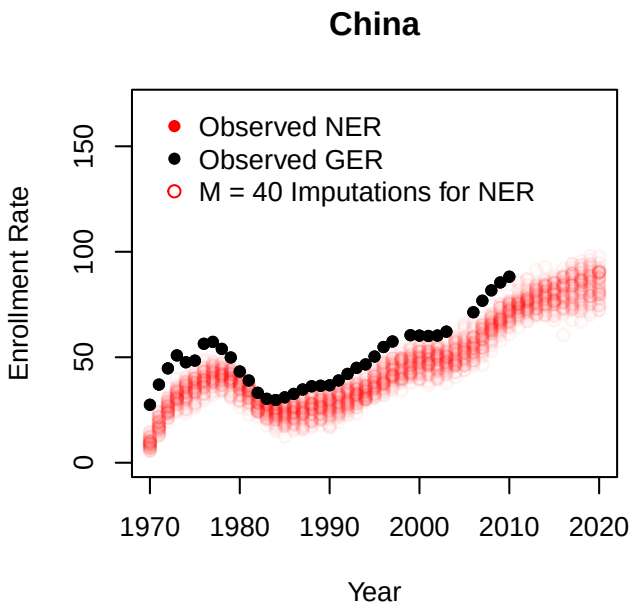

Botswana
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

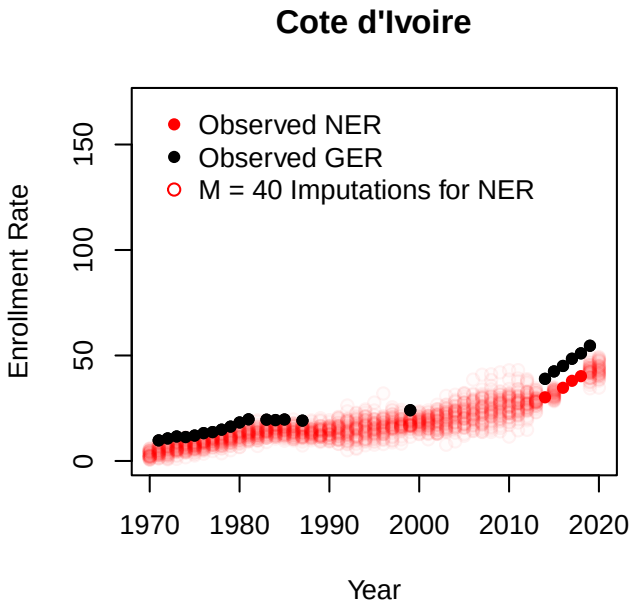

Central African Republic
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

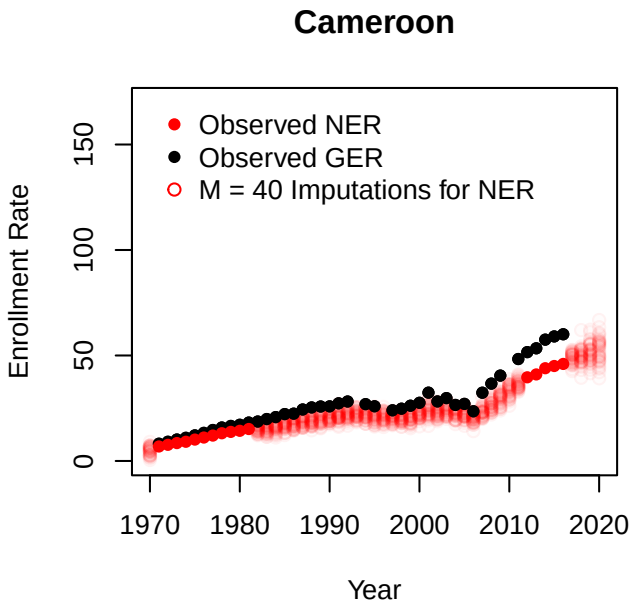

Canada
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

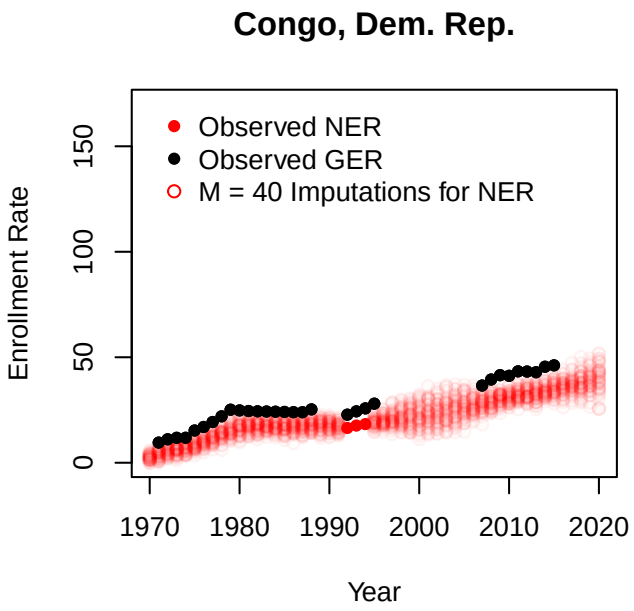

Switzerland
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

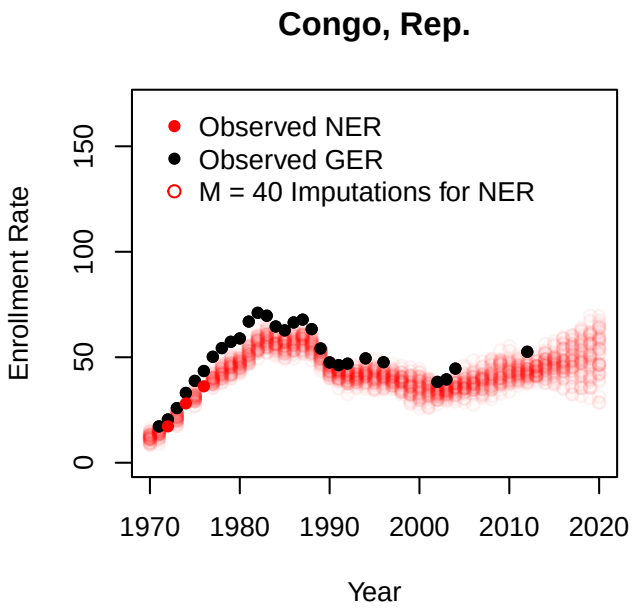

Chile
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

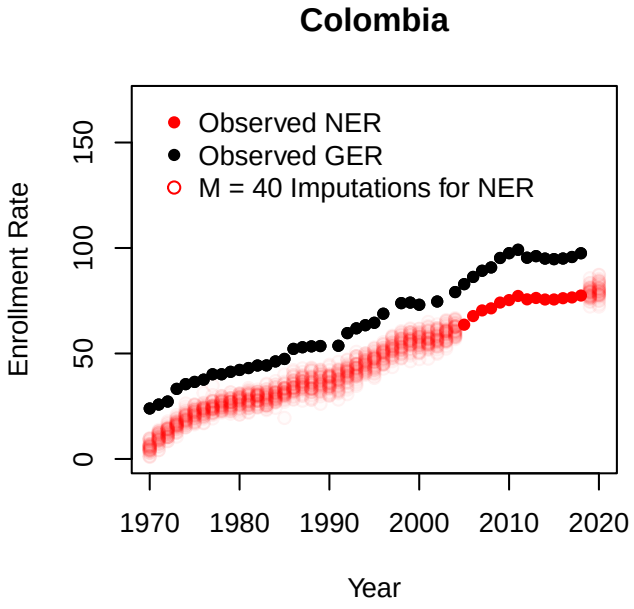

China
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

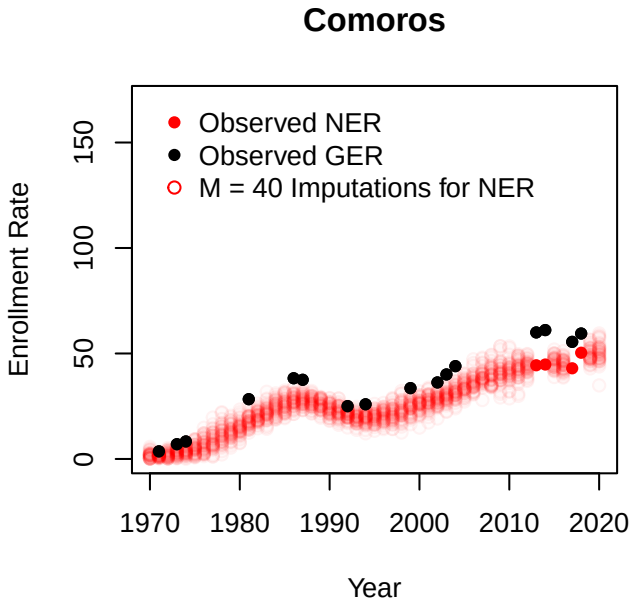

Cote d'Ivoire
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

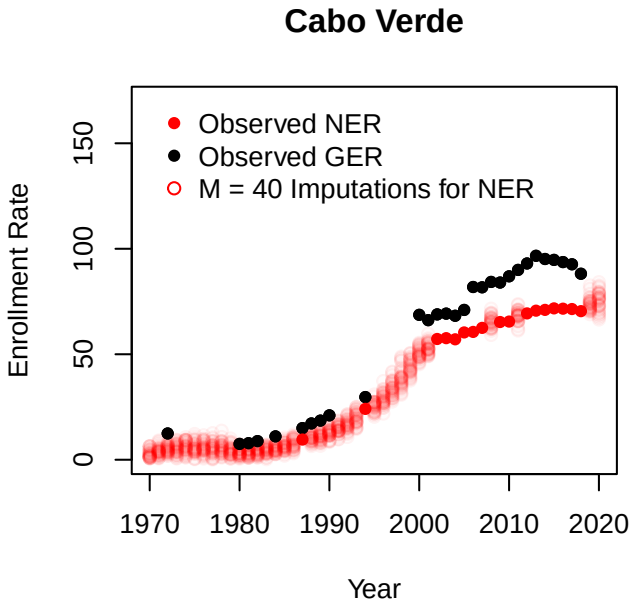

Cameroon
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

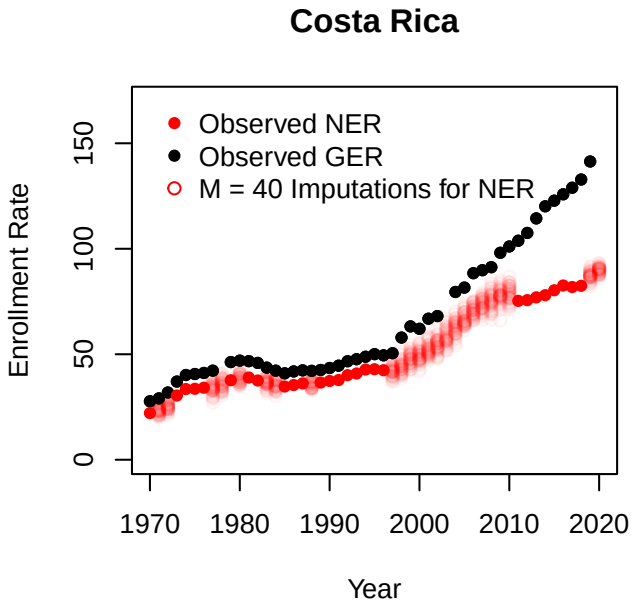

Congo, Dem. Rep.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

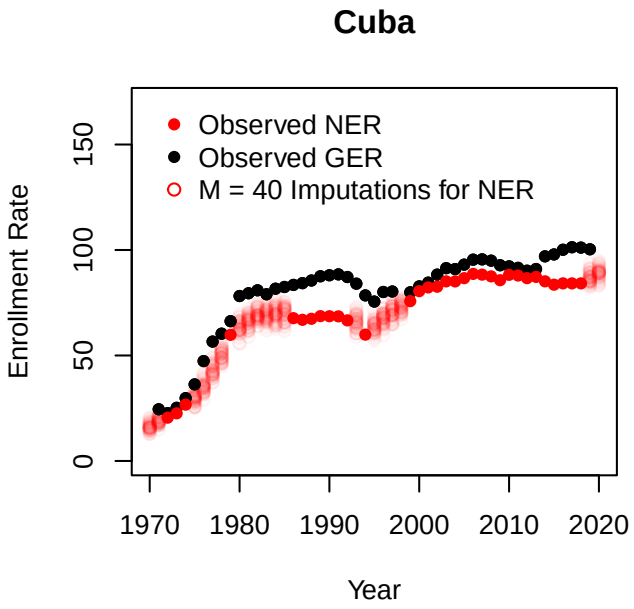

Congo, Rep.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

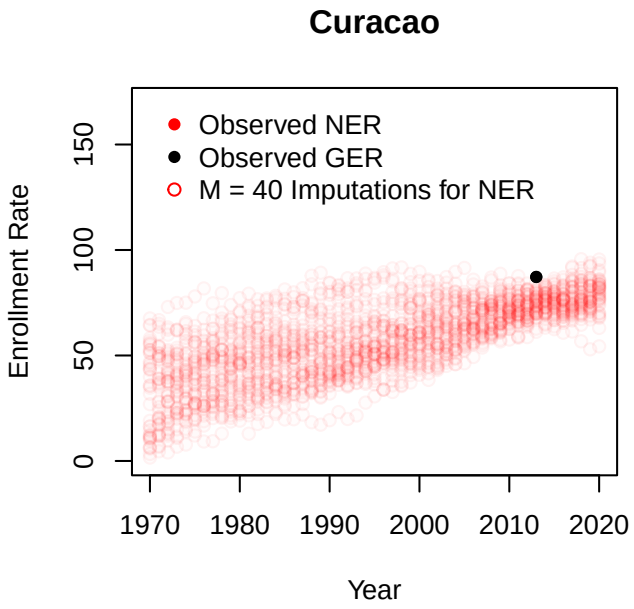

Colombia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

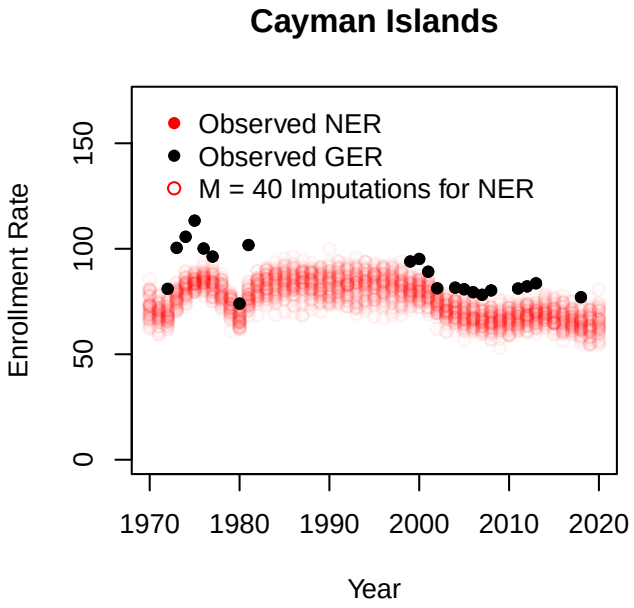

Comoros
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

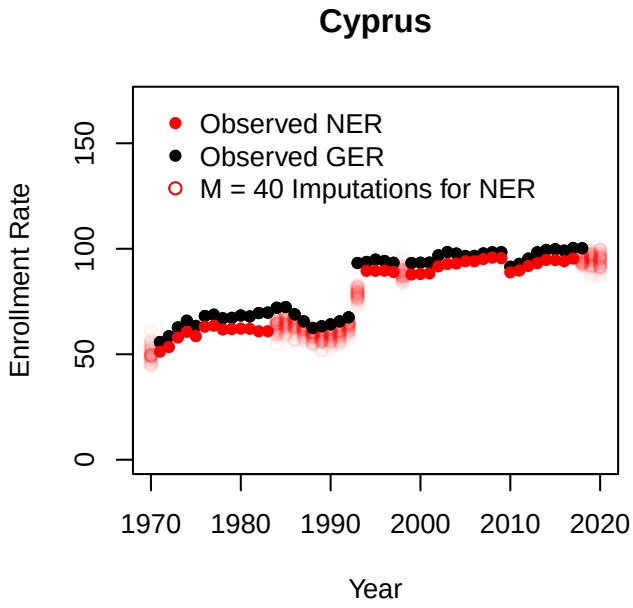

Cabo Verde
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

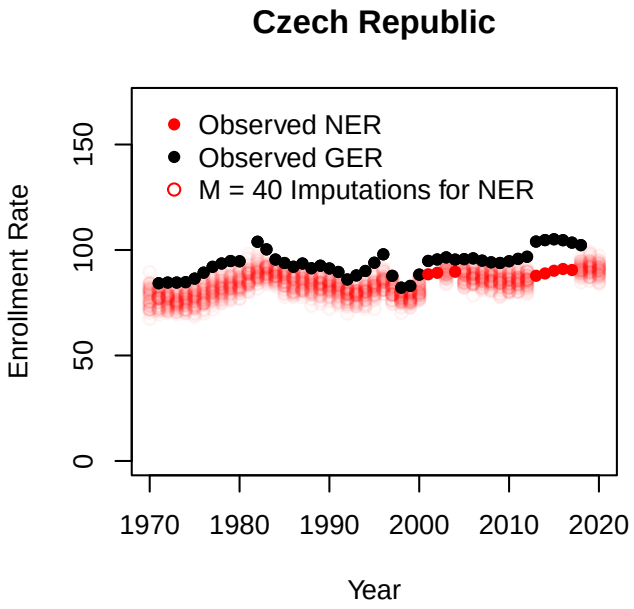

Costa Rica
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

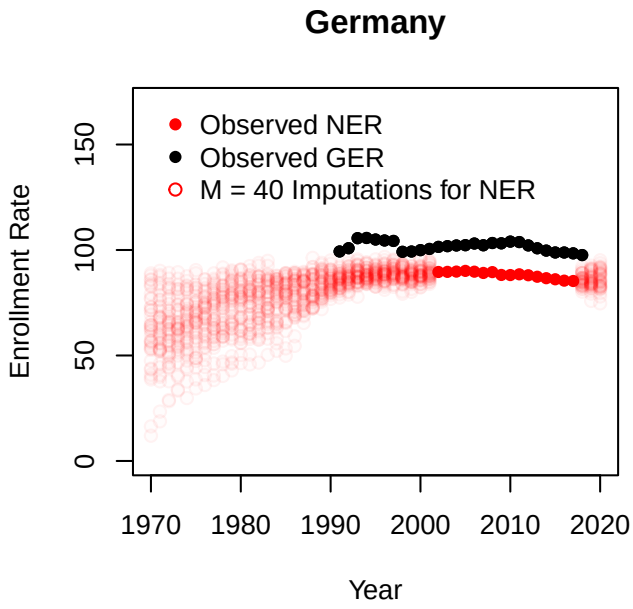

Cuba
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

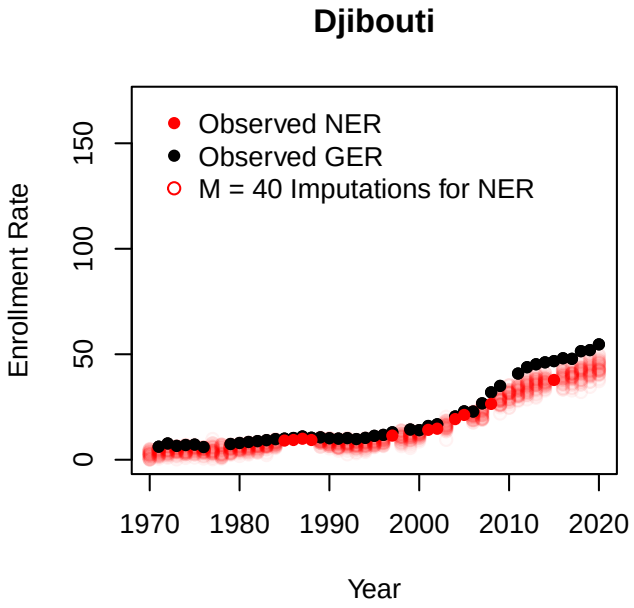

Curacao
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

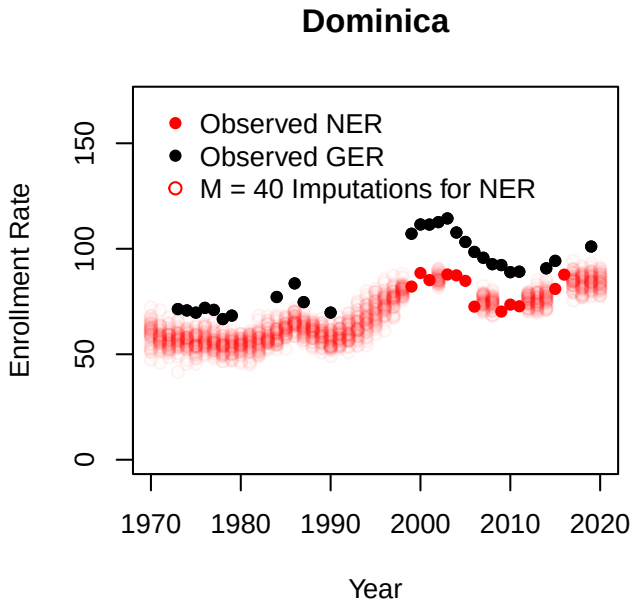

Cayman Islands
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

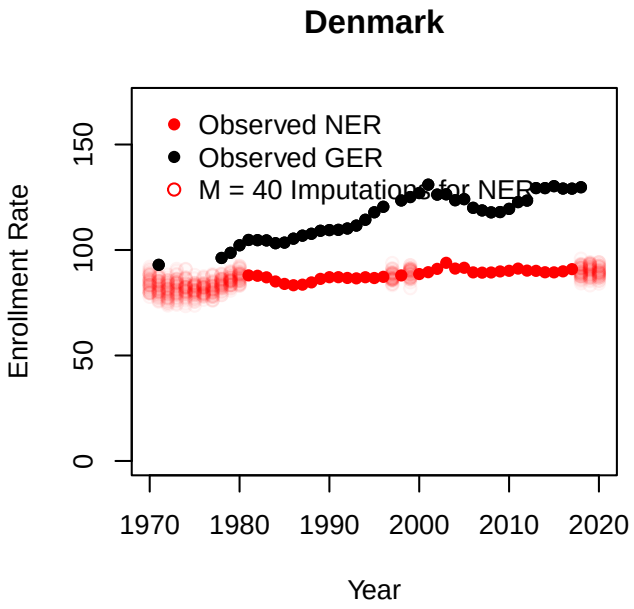

Cyprus
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

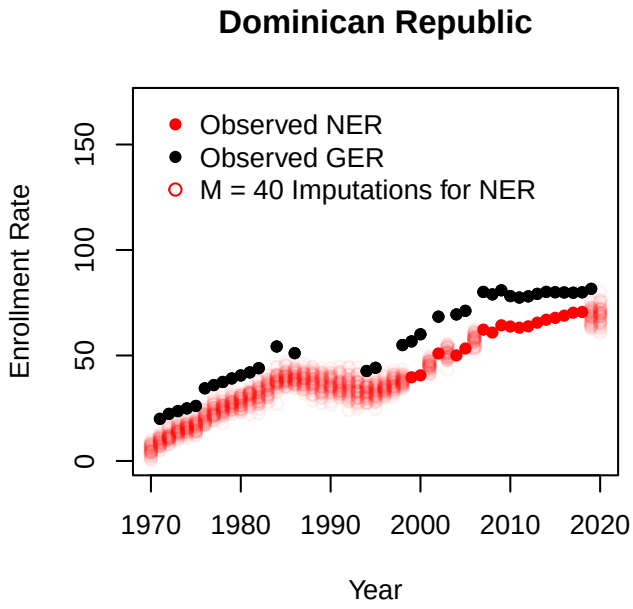

Czech Republic
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

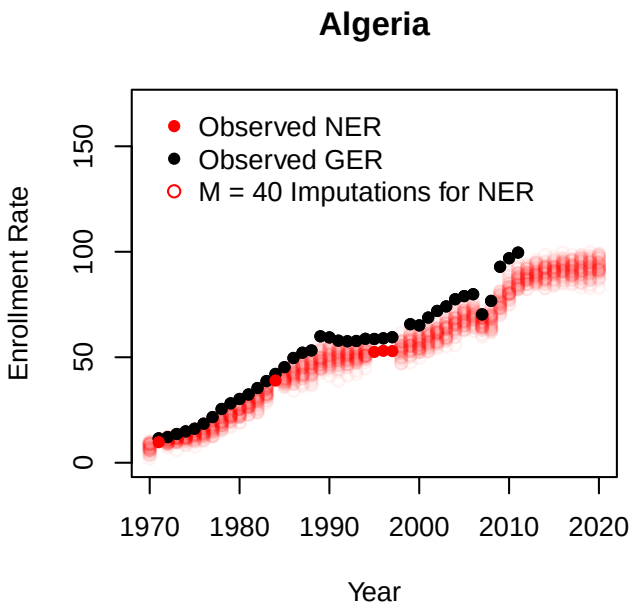

Germany
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

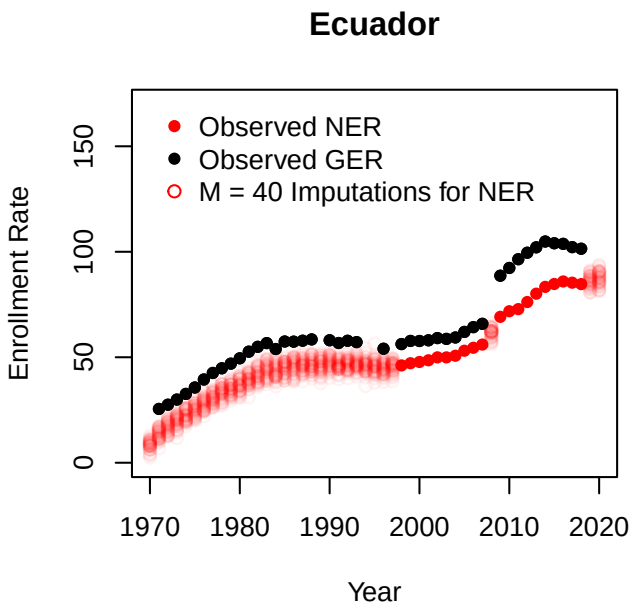

Djibouti
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

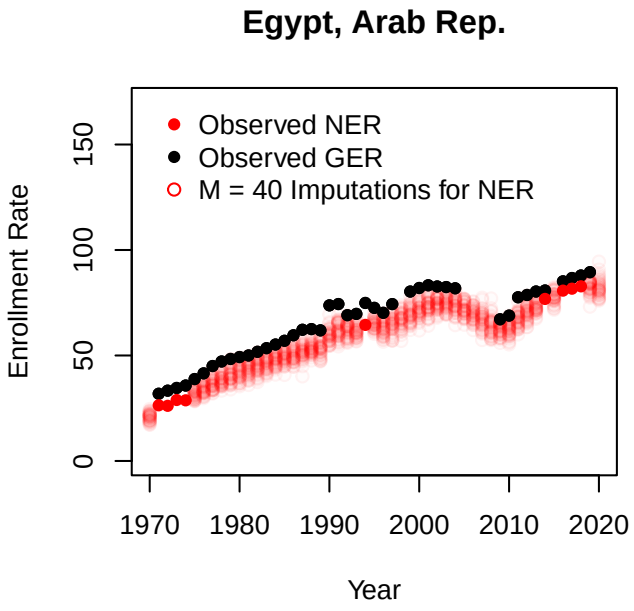

Dominica
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

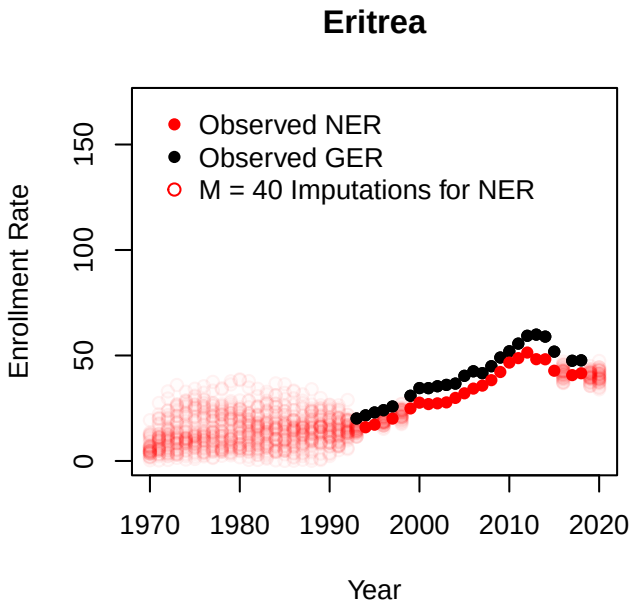

Denmark
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

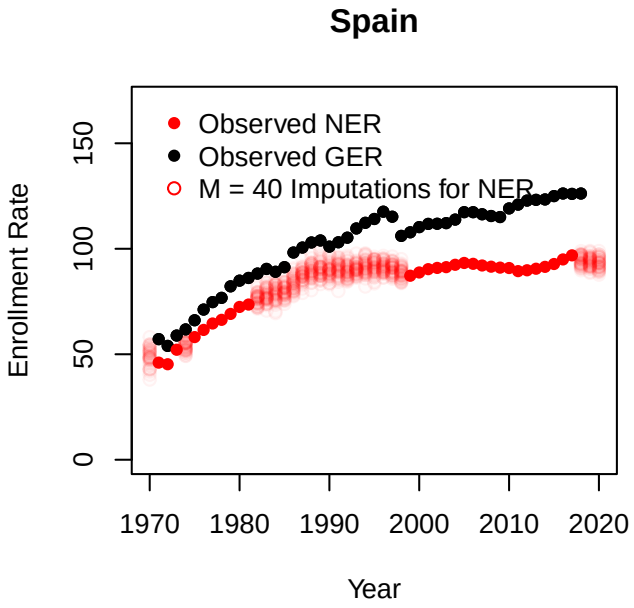

Dominican Republic
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

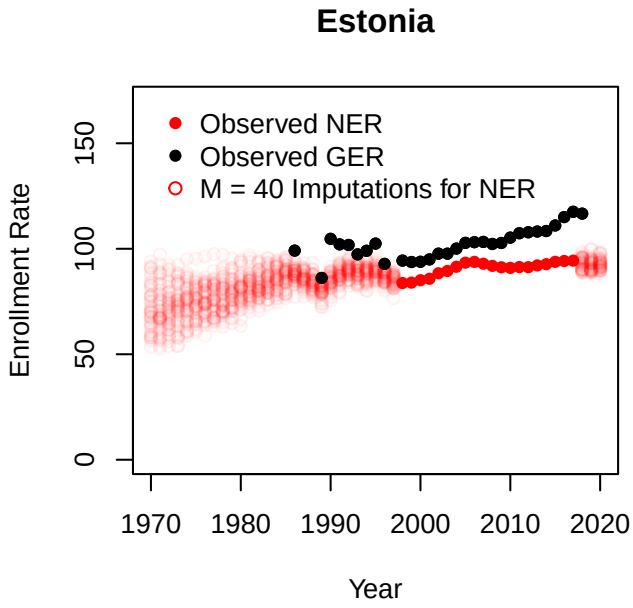

Algeria
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

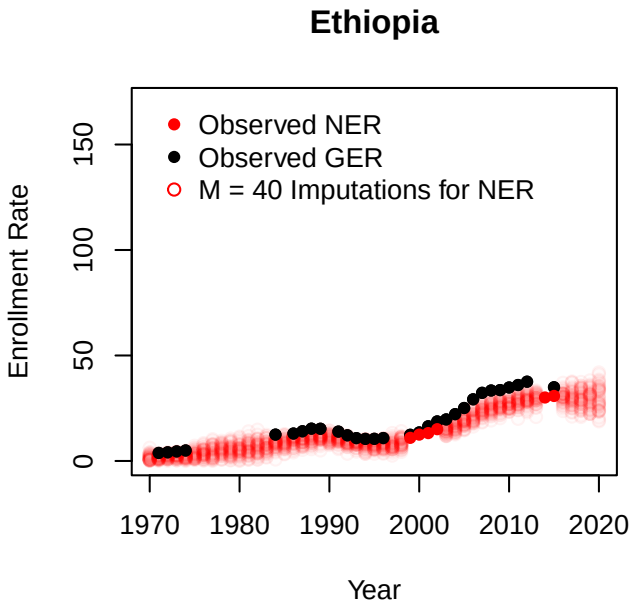

Ecuador
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

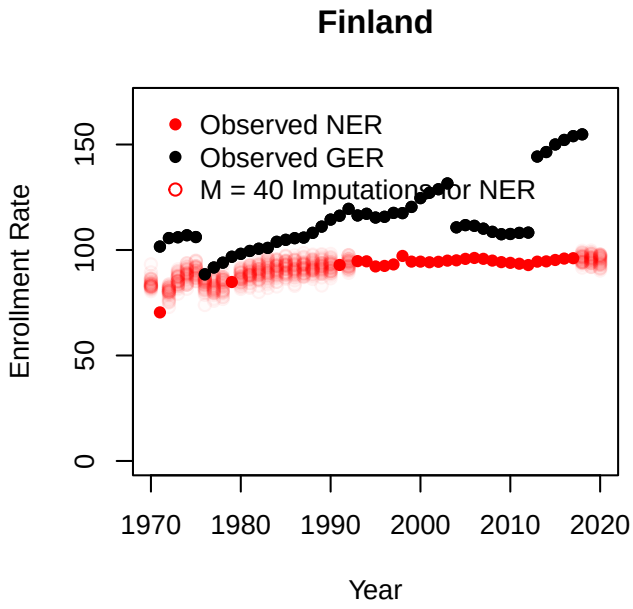

Egypt, Arab Rep.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

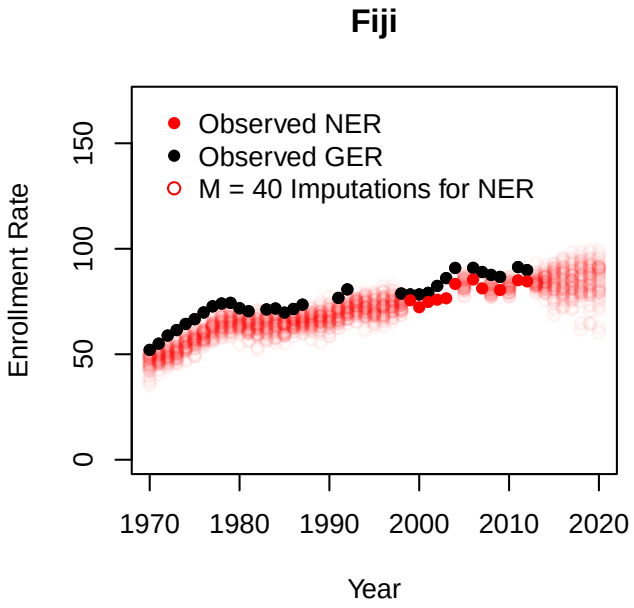

Eritrea
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

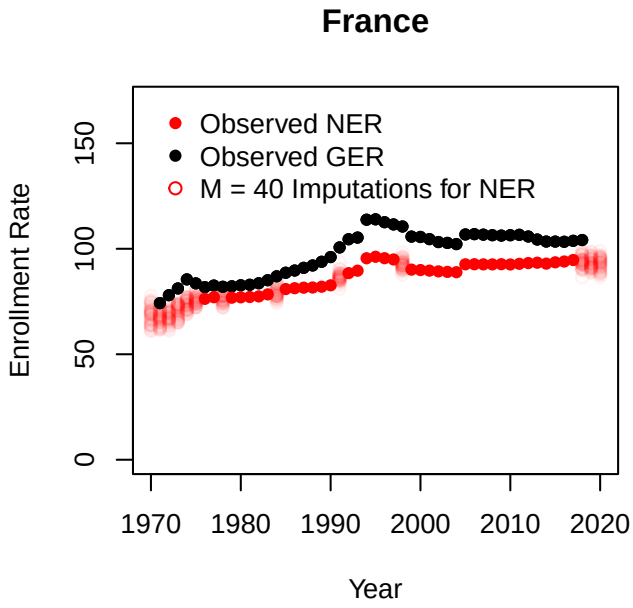

Spain
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

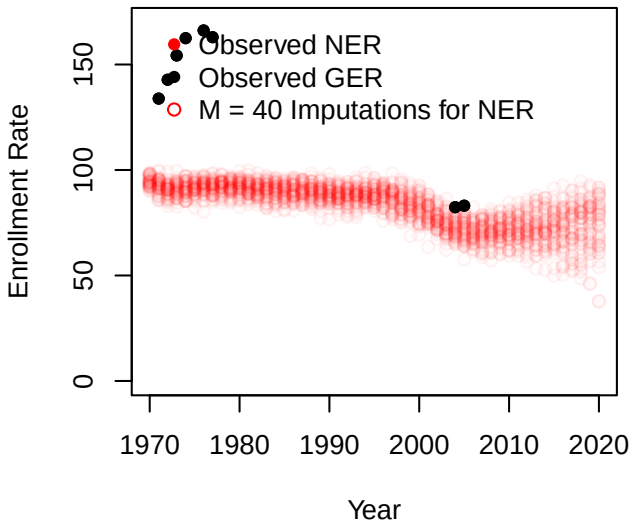

Estonia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

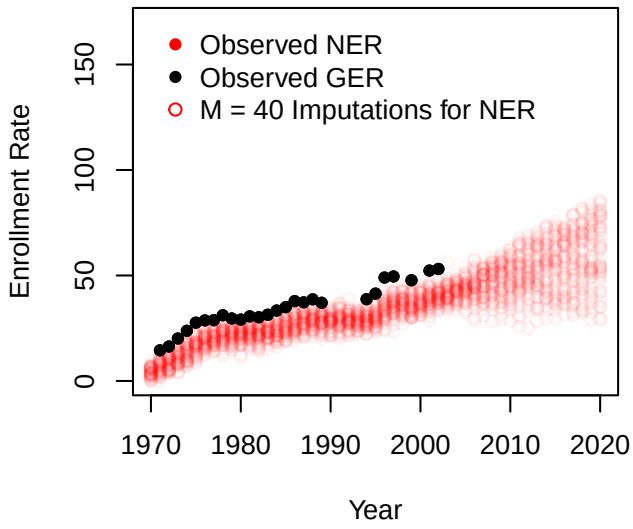

Ethiopia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

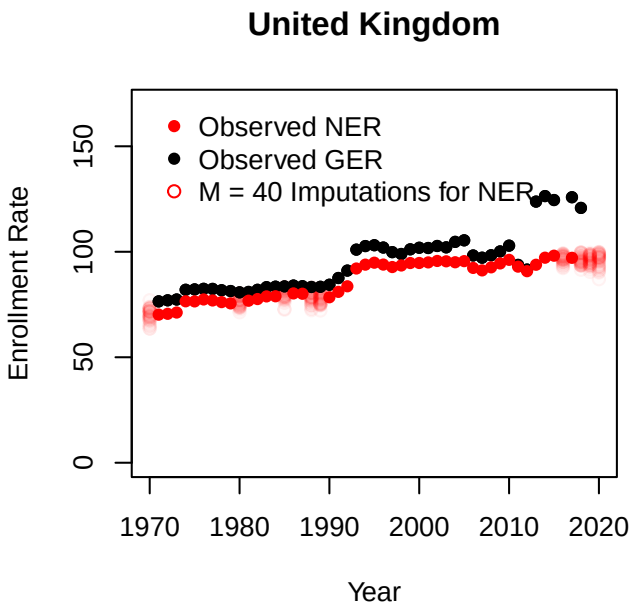

Finland
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

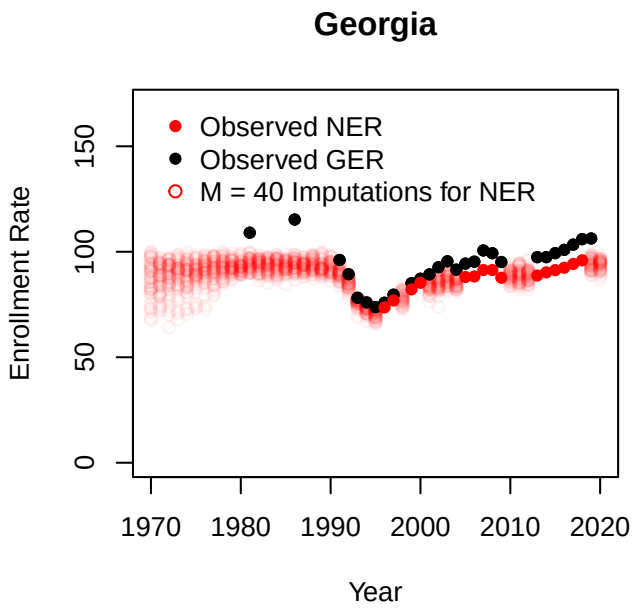

Fiji
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

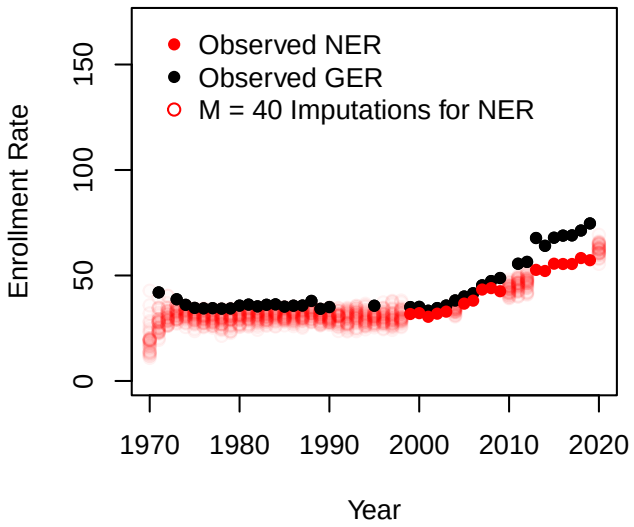

France
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

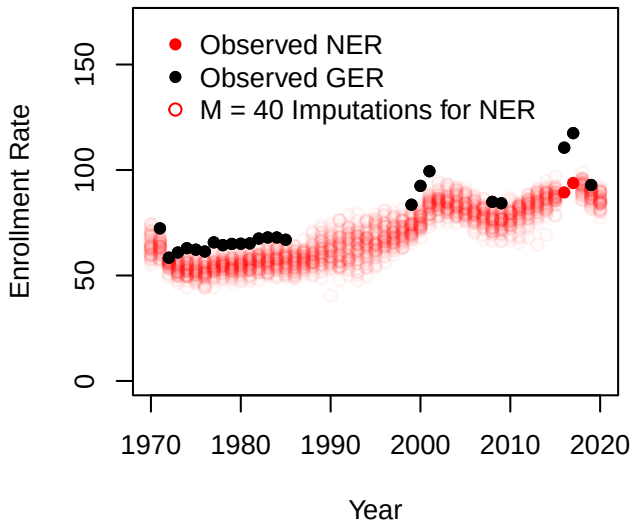

Micronesia, Fed. Sts.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

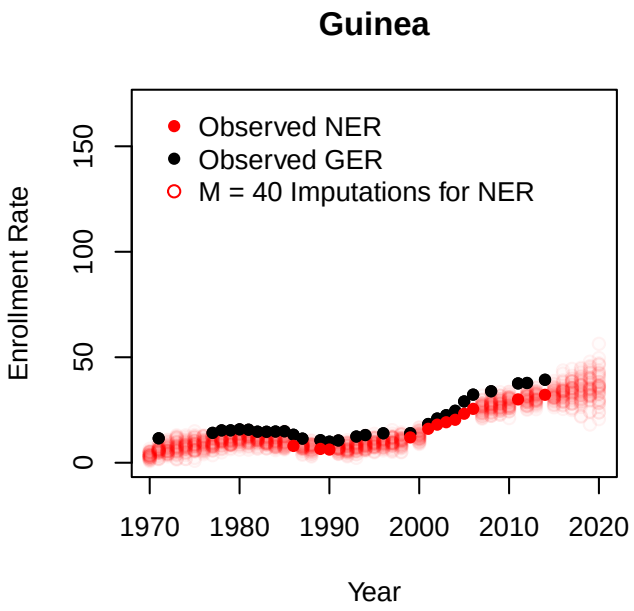

Gabon
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

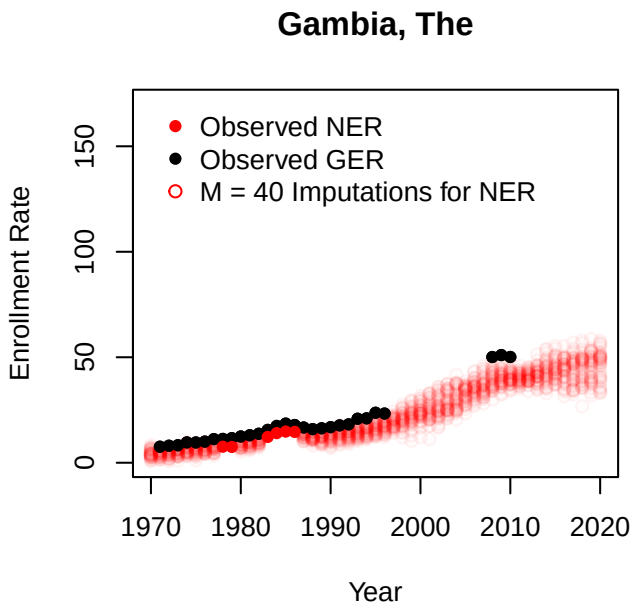

United Kingdom
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

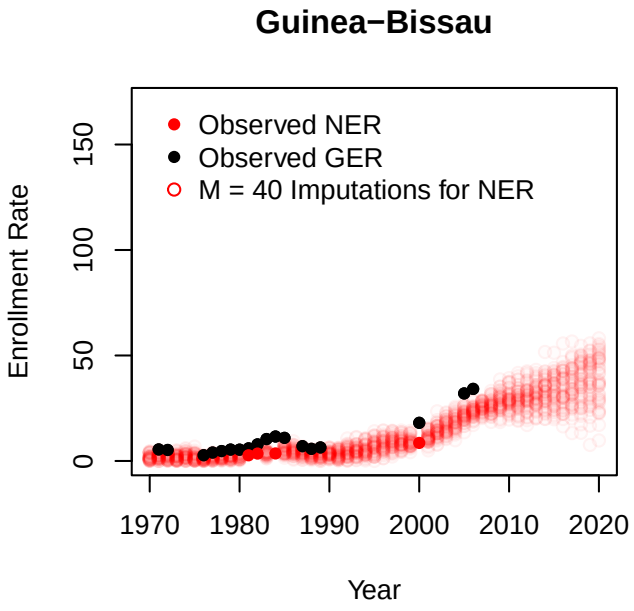

Georgia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

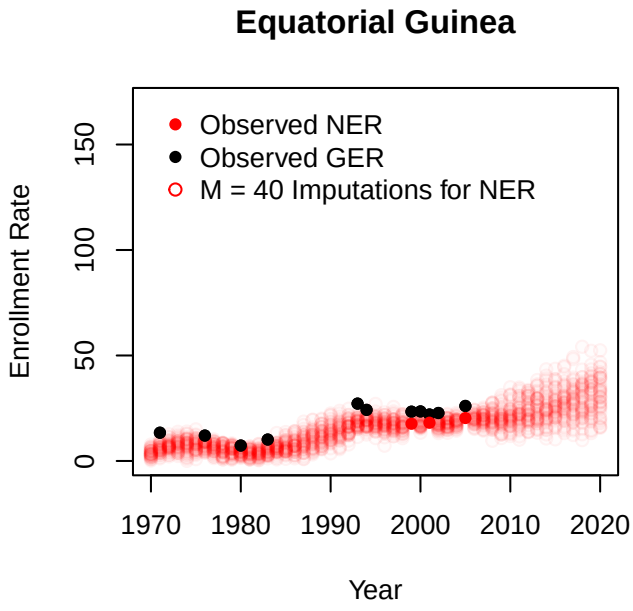

Ghana
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

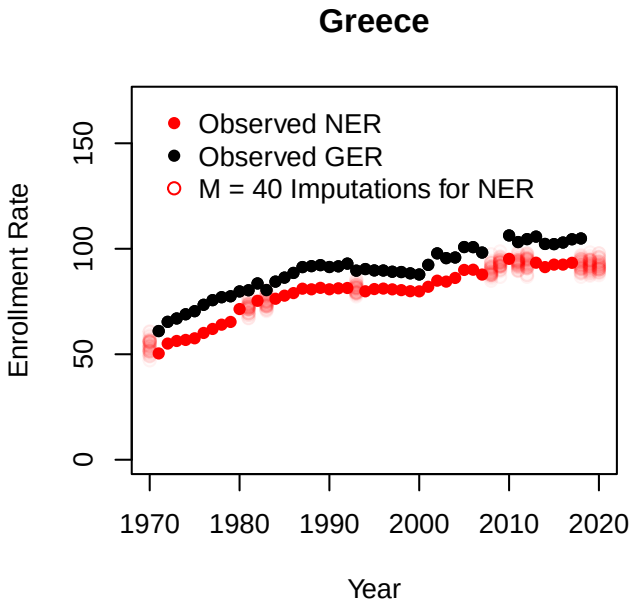

Gibraltar
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

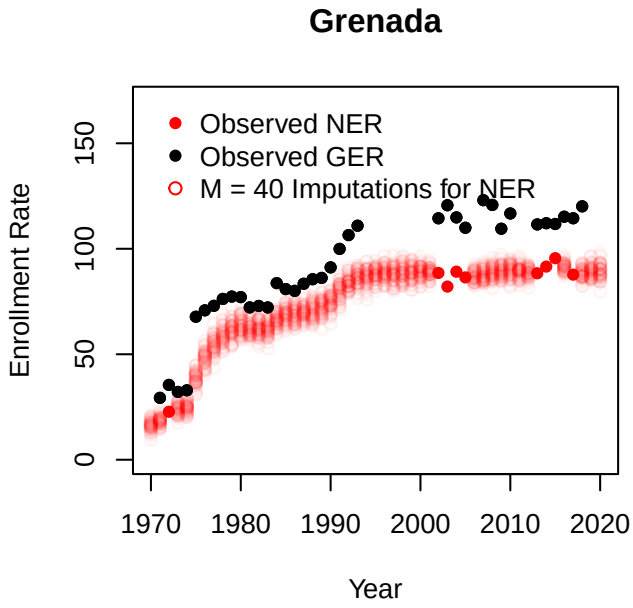

Guinea
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

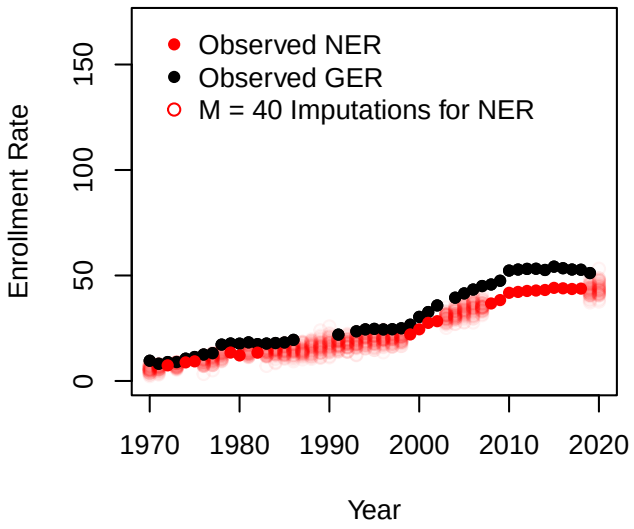

Gambia, The
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

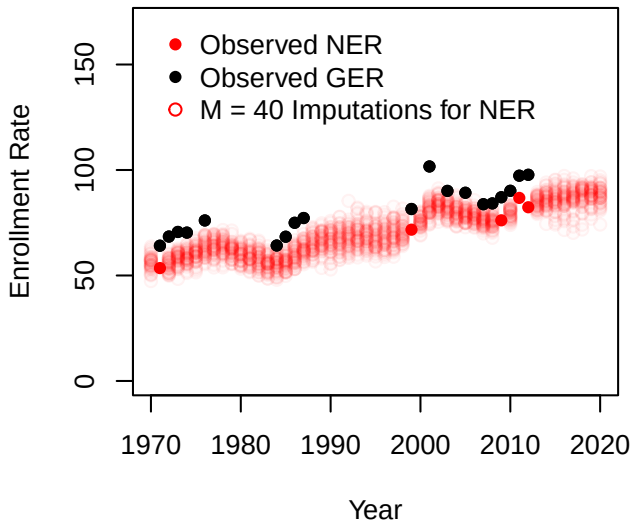

Guinea−Bissau
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

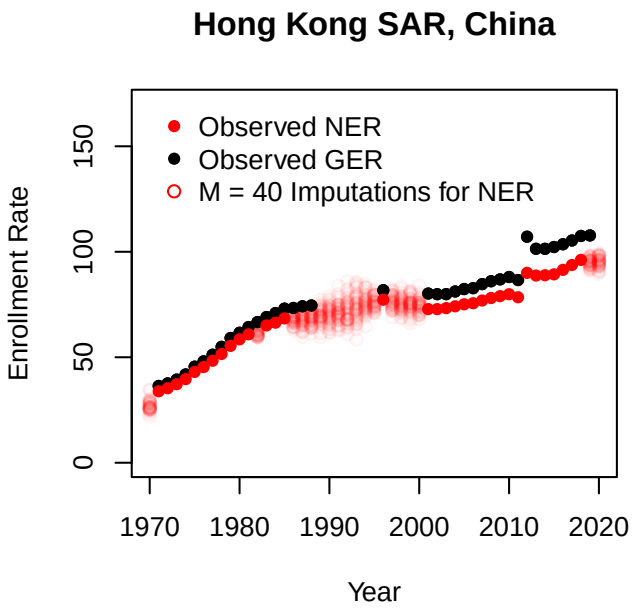

Equatorial Guinea
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

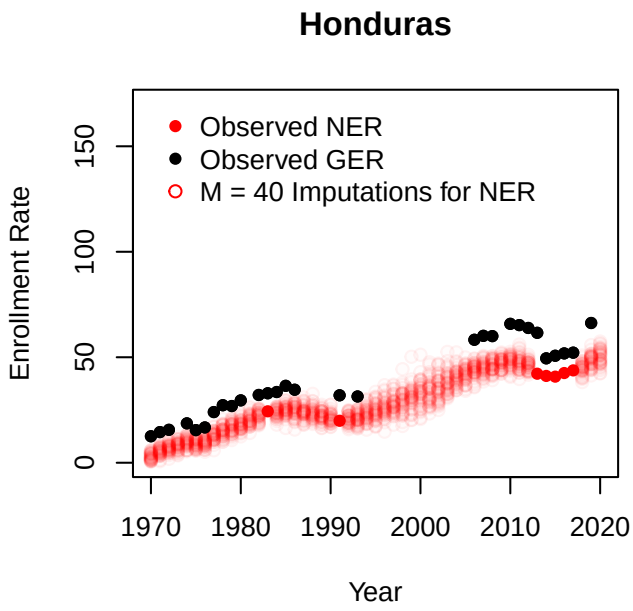

Greece
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

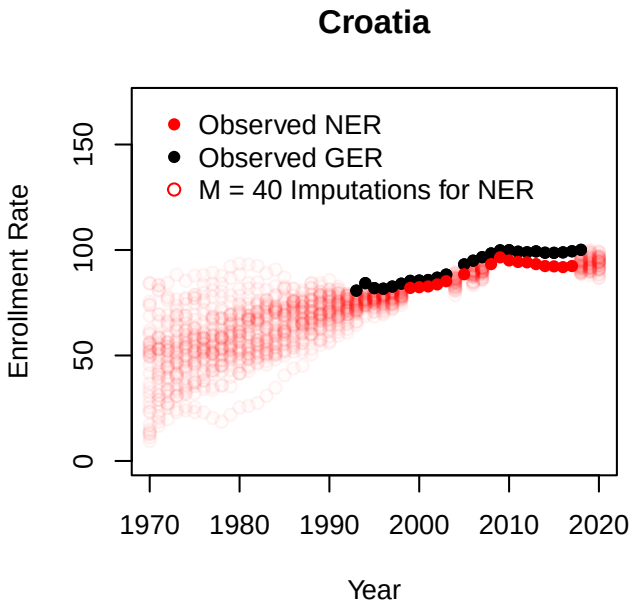

Grenada
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

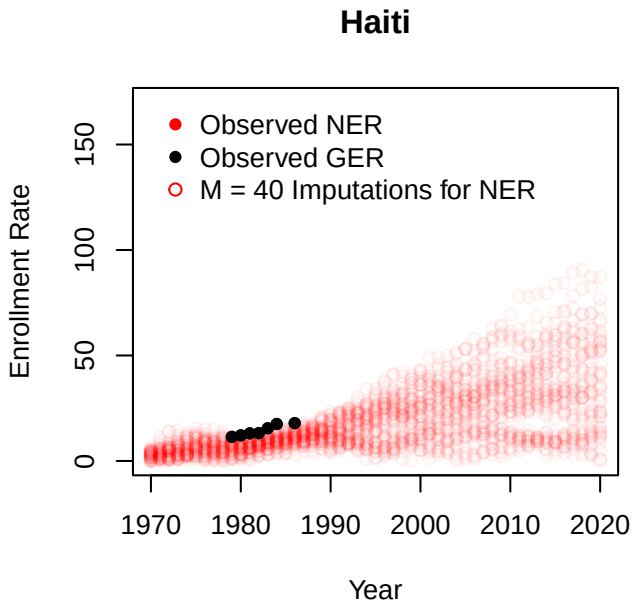

Guatemala
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

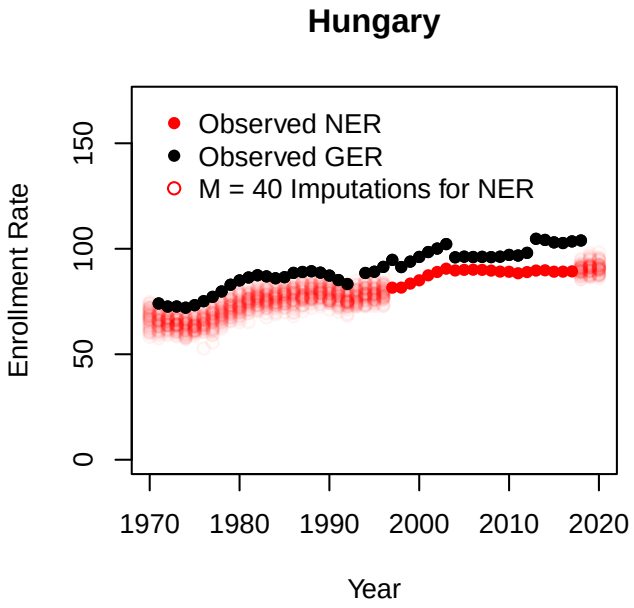

Guyana
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

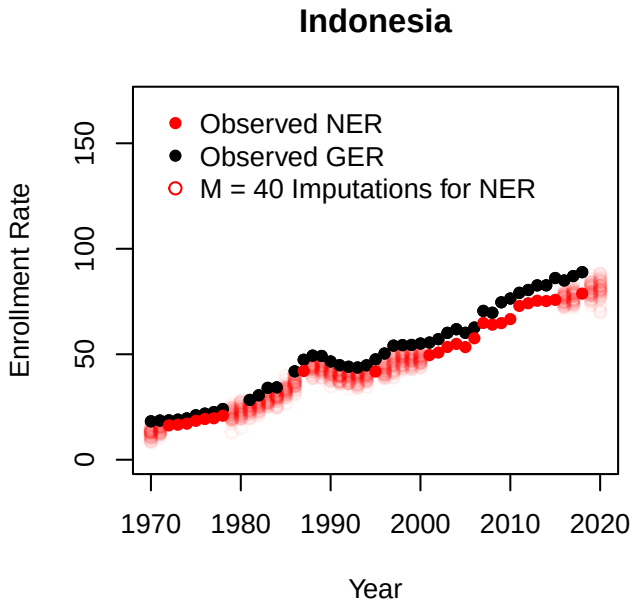

Hong Kong SAR, China
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

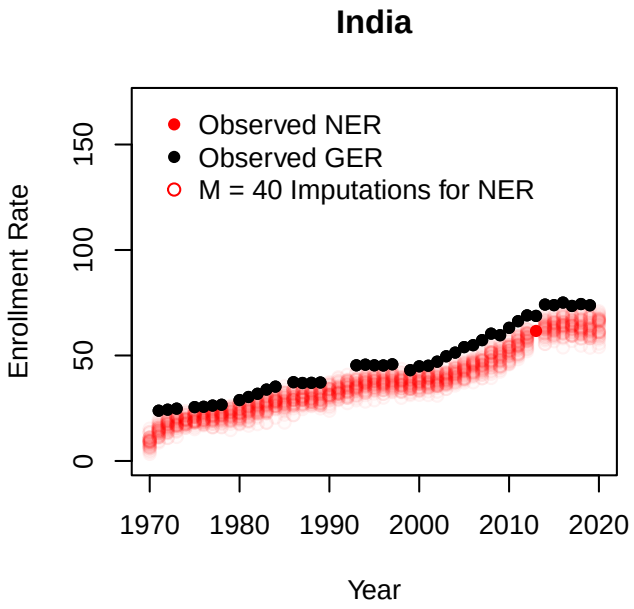

Honduras
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

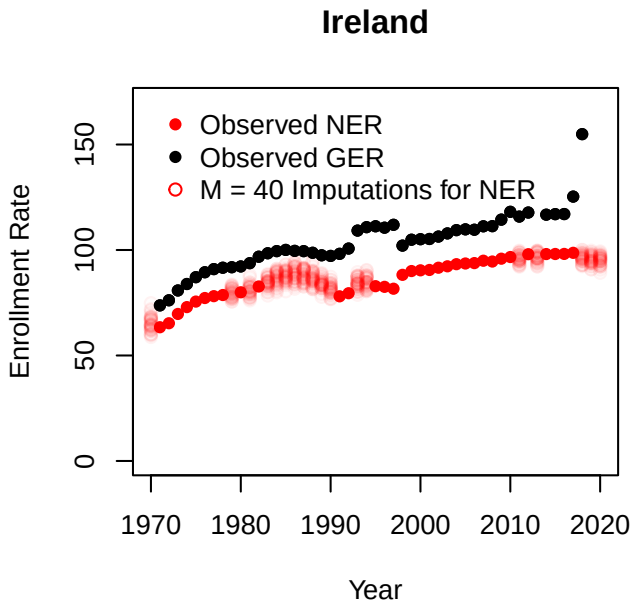

Croatia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

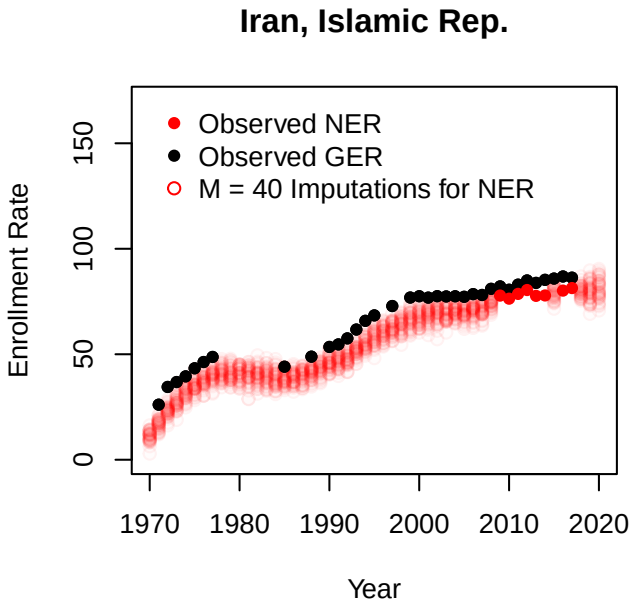

Haiti
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

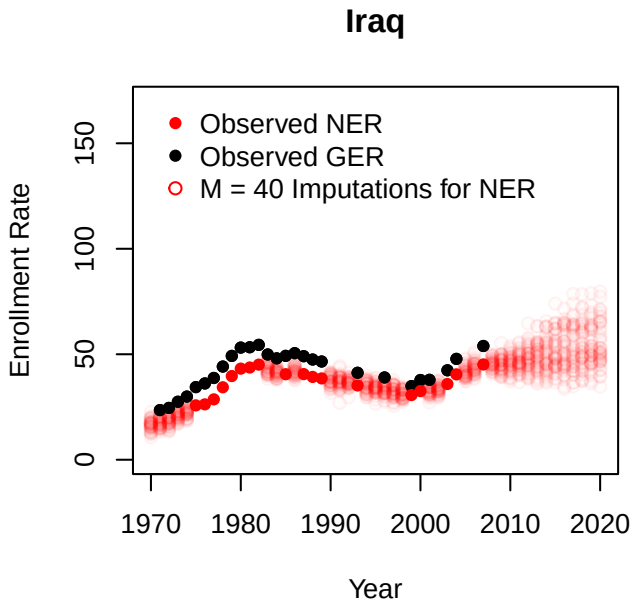

Hungary
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

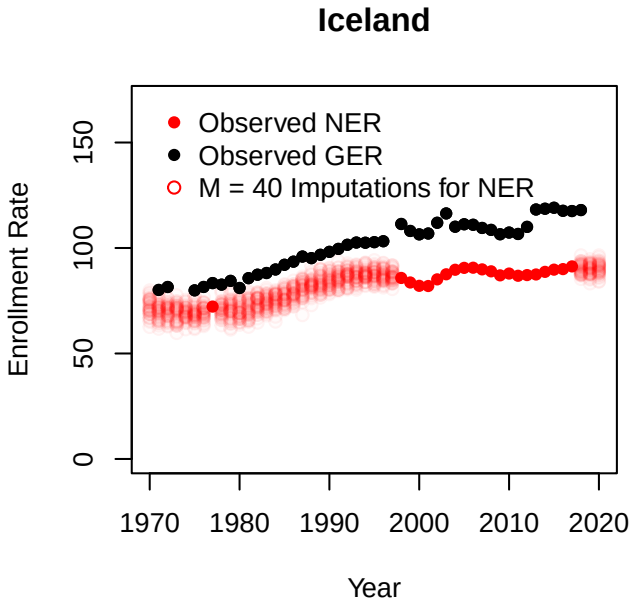

Indonesia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

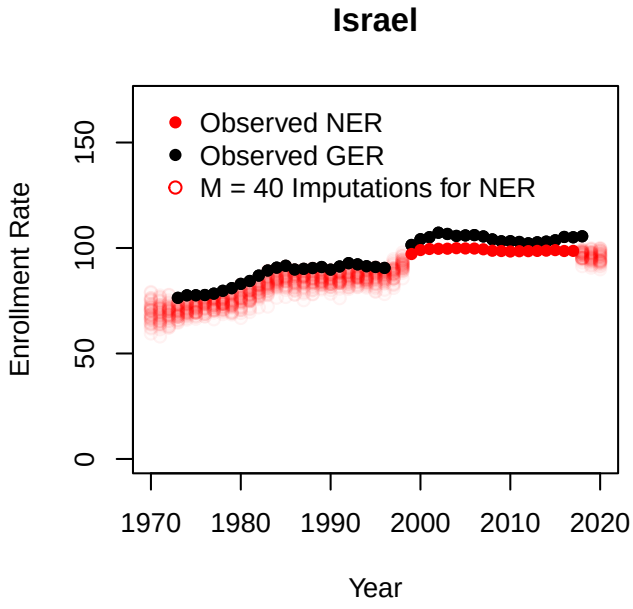

India
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

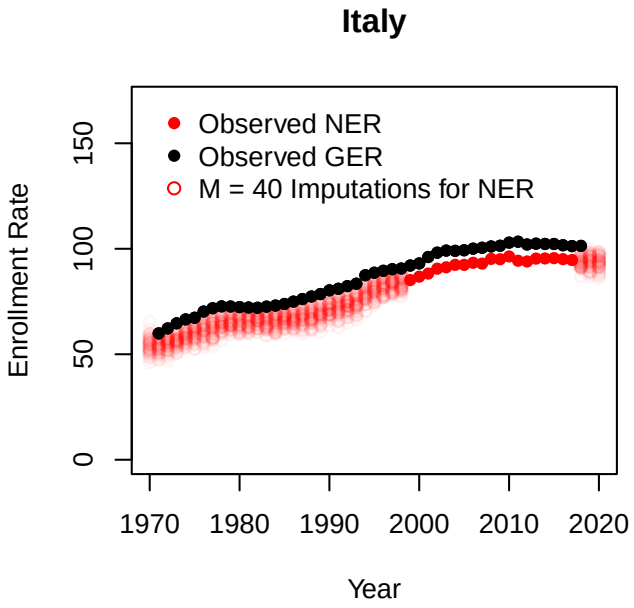

Ireland
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

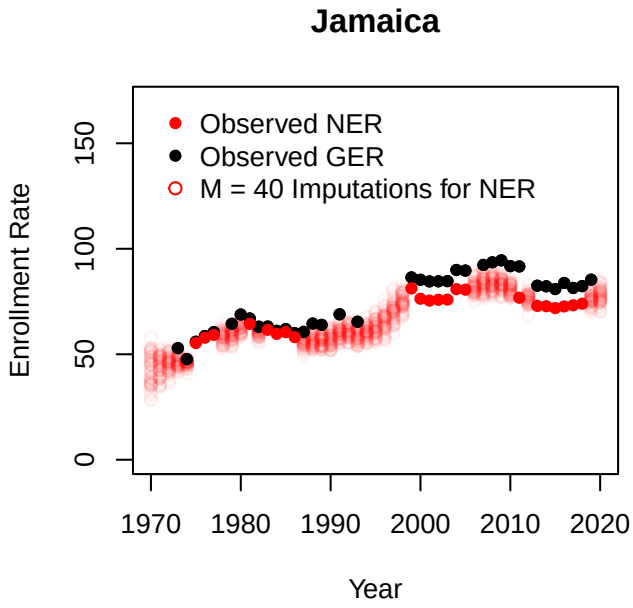

Iran, Islamic Rep.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

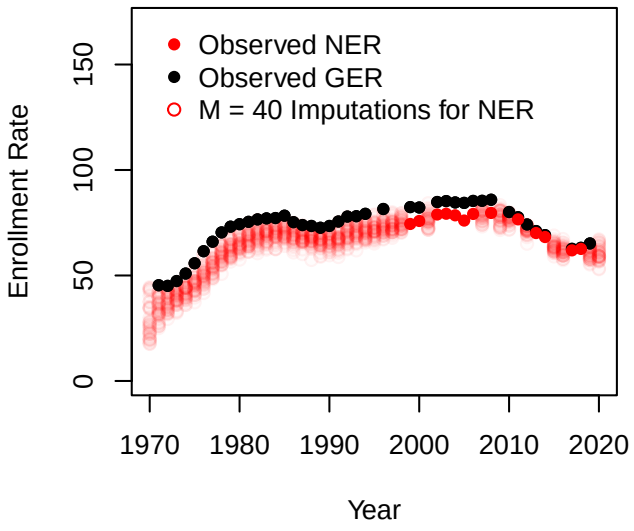

Iraq
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

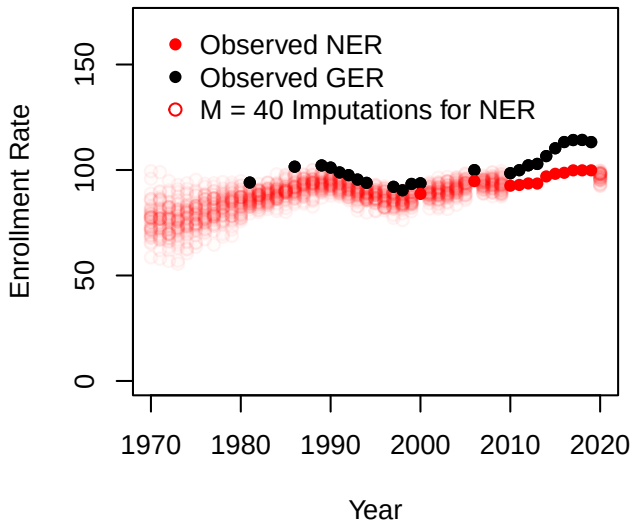

Iceland
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

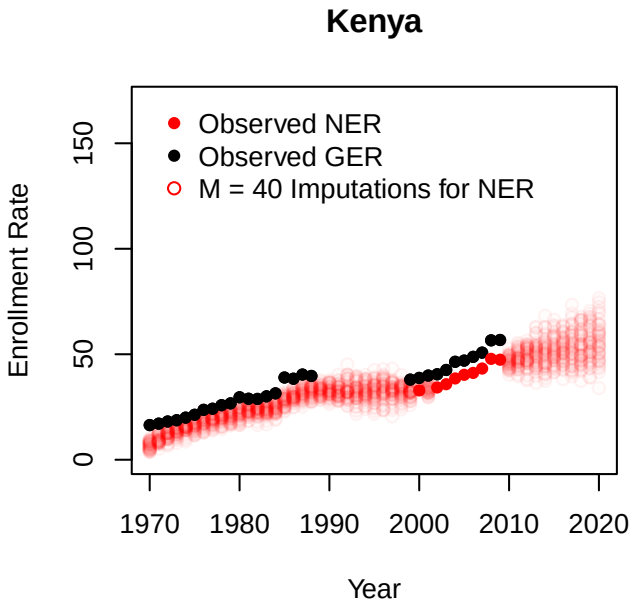

Israel
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

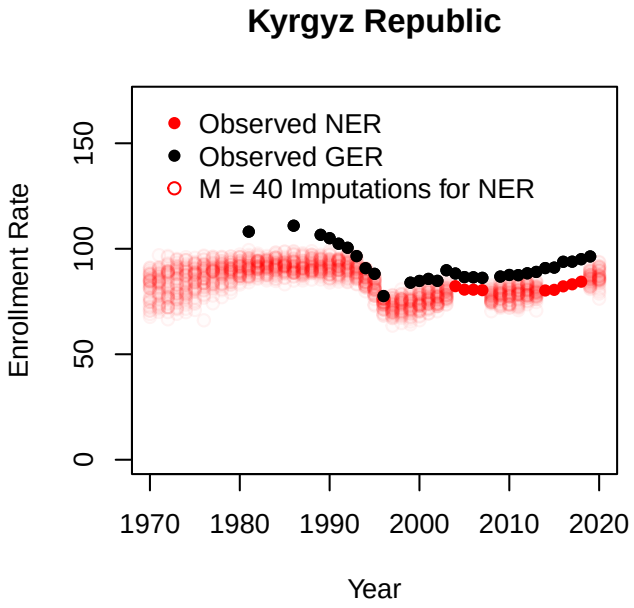

Italy
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

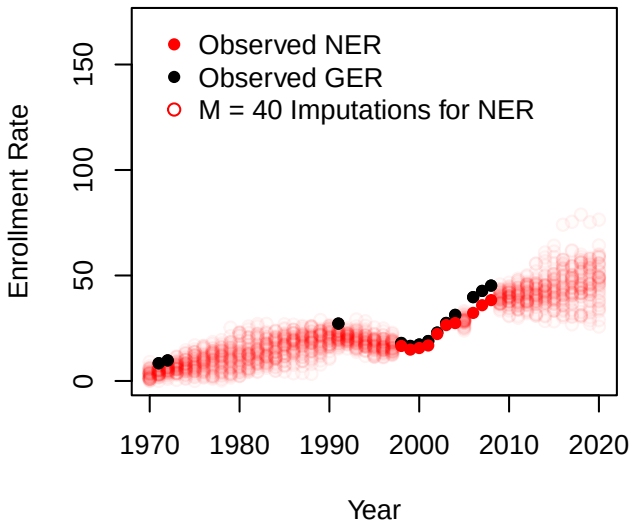

Jamaica
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

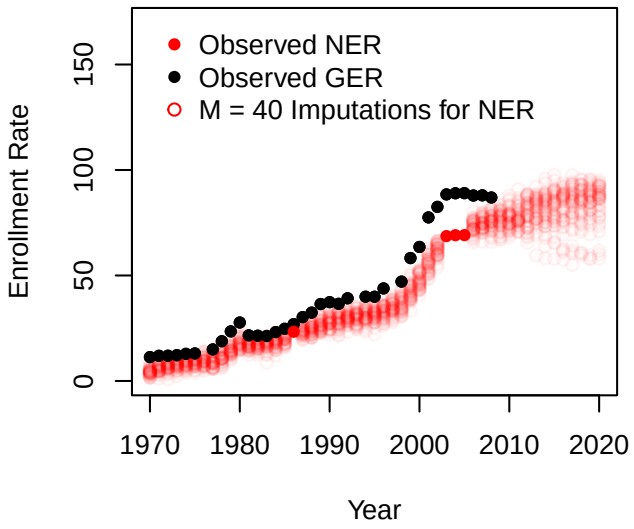

Jordan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

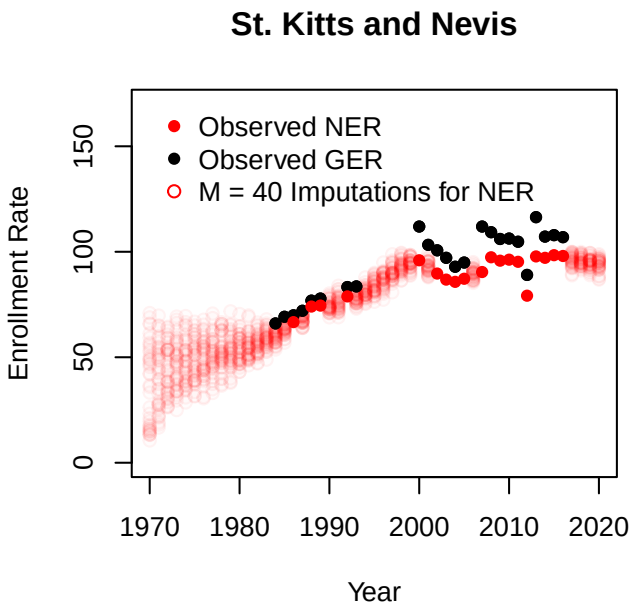

Kazakhstan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

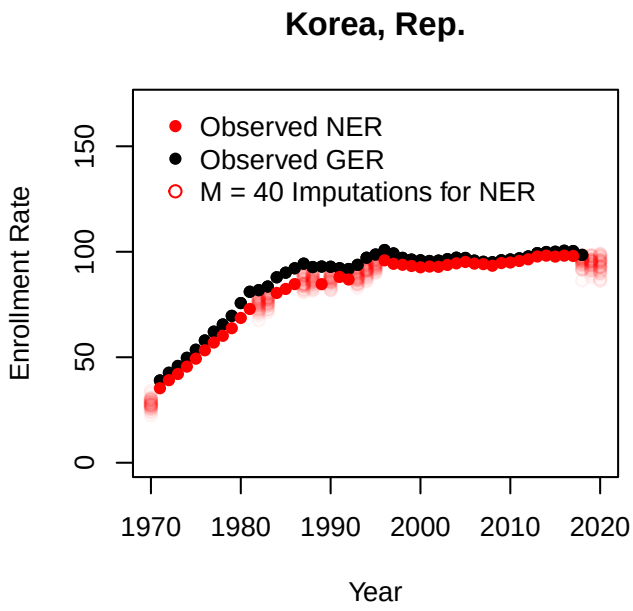

Kenya
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

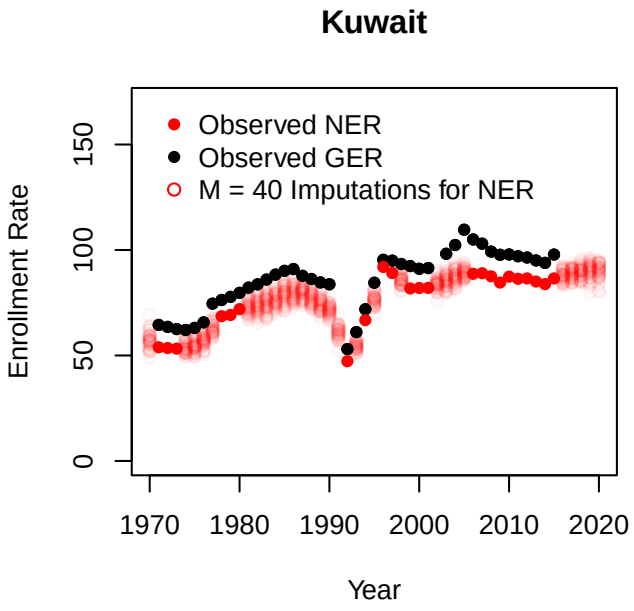

Kyrgyz Republic
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

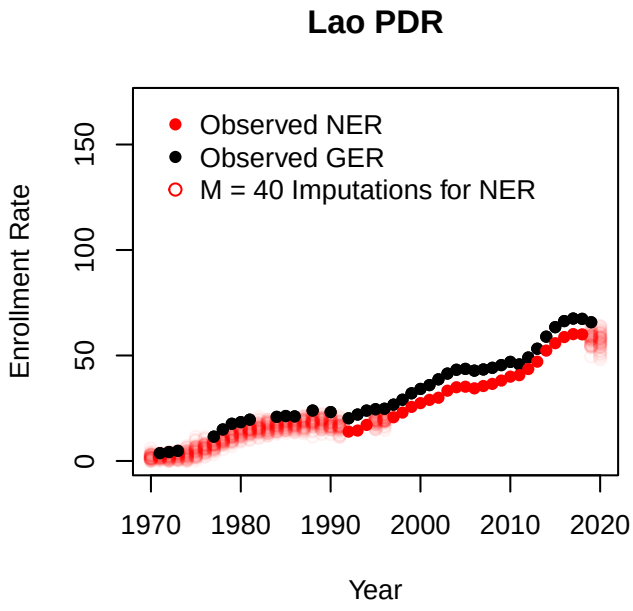

Cambodia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

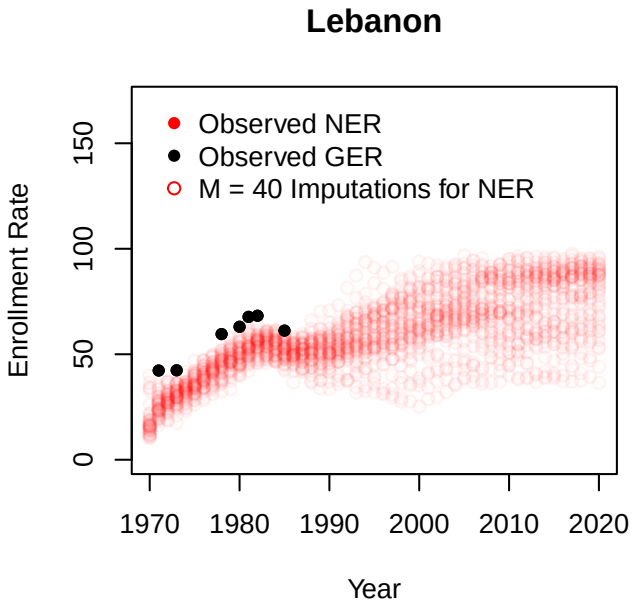

Kiribati
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

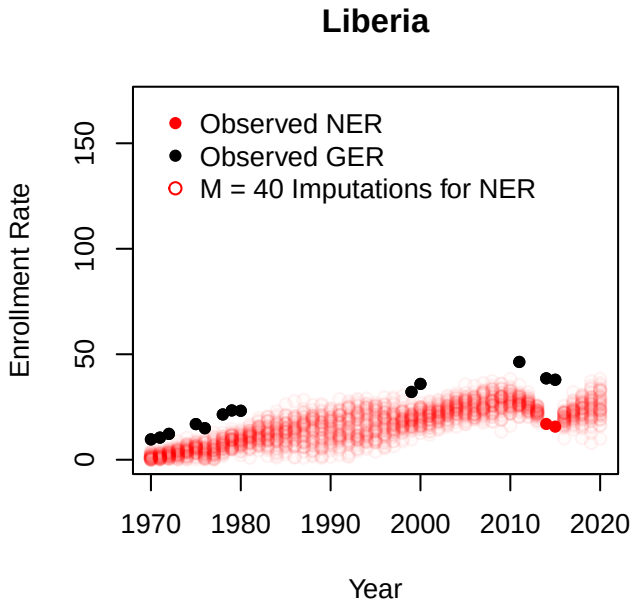

St. Kitts and Nevis
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

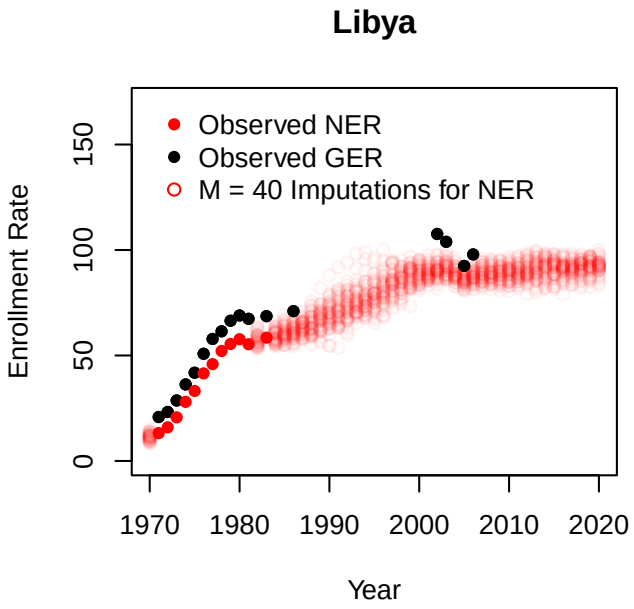

Korea, Rep.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

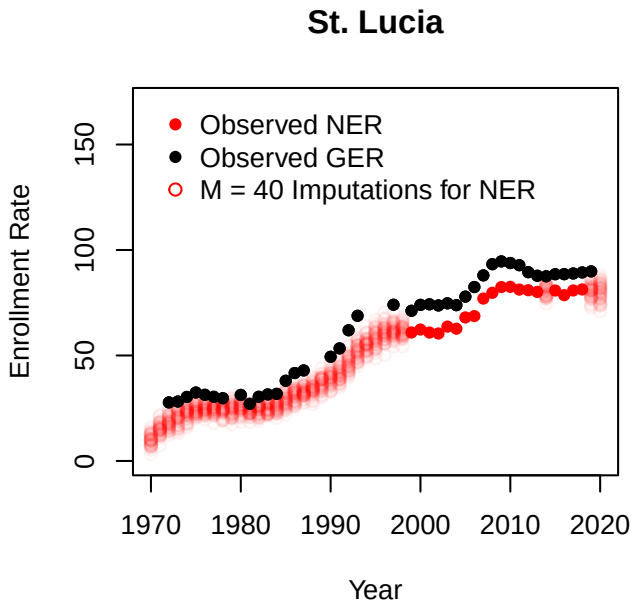

Kuwait
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

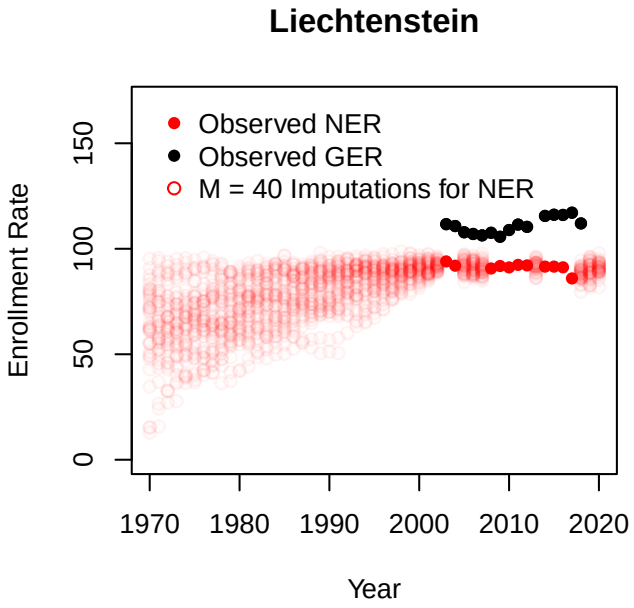

Lao PDR
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

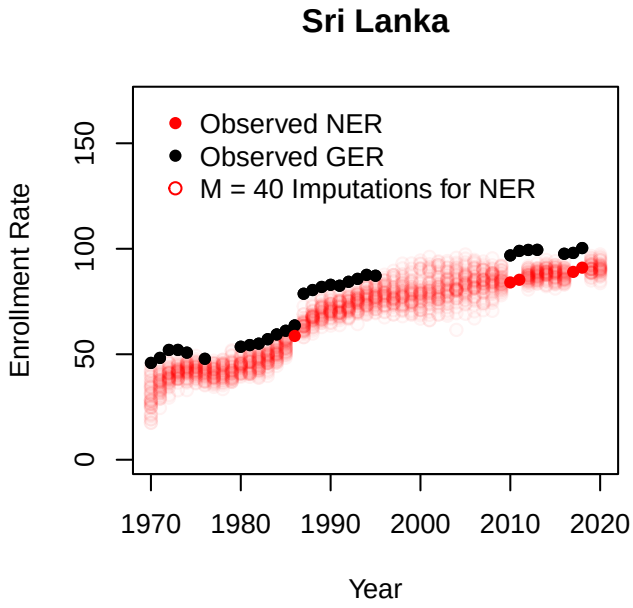

Lebanon
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

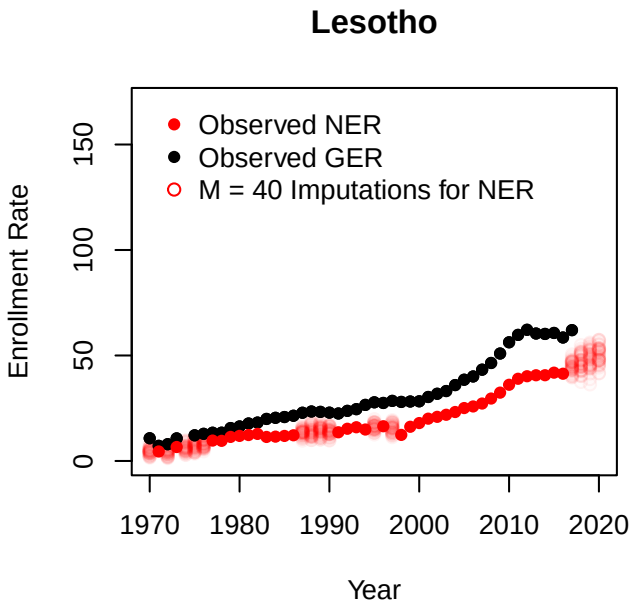

Liberia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

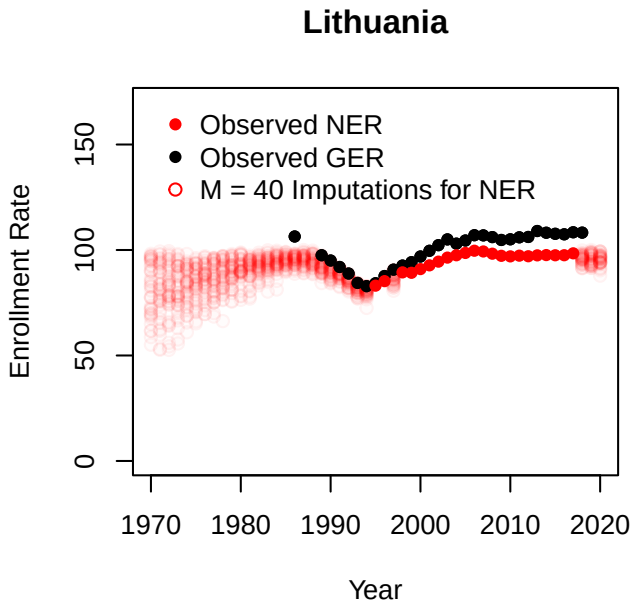

Libya
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

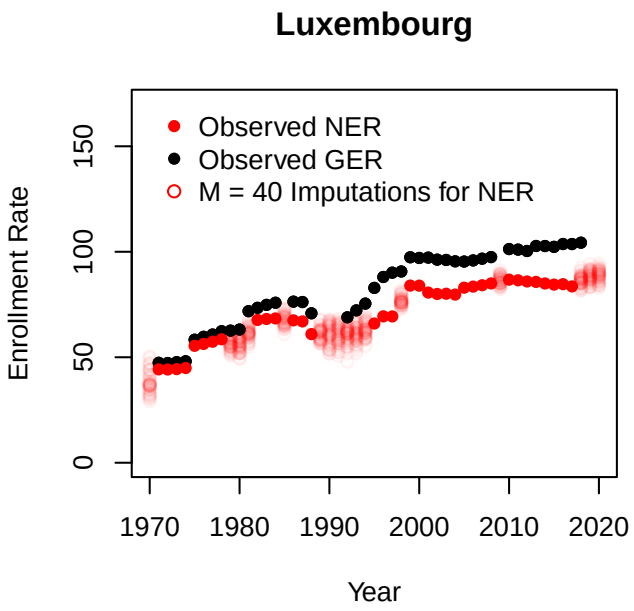

St. Lucia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

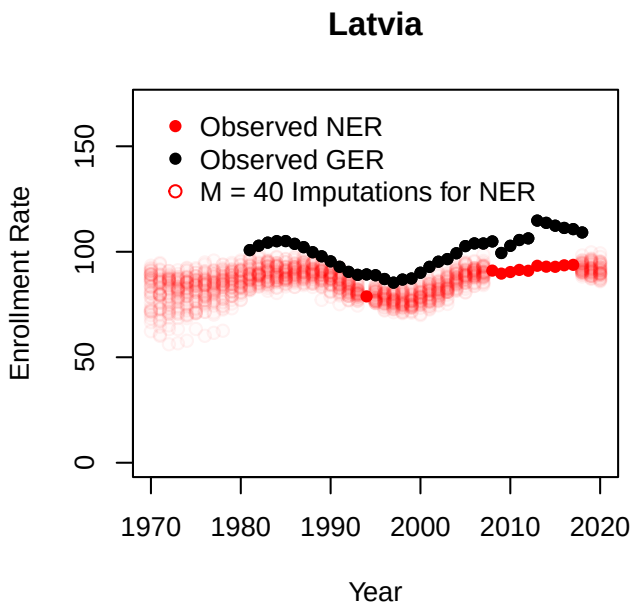

Liechtenstein
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

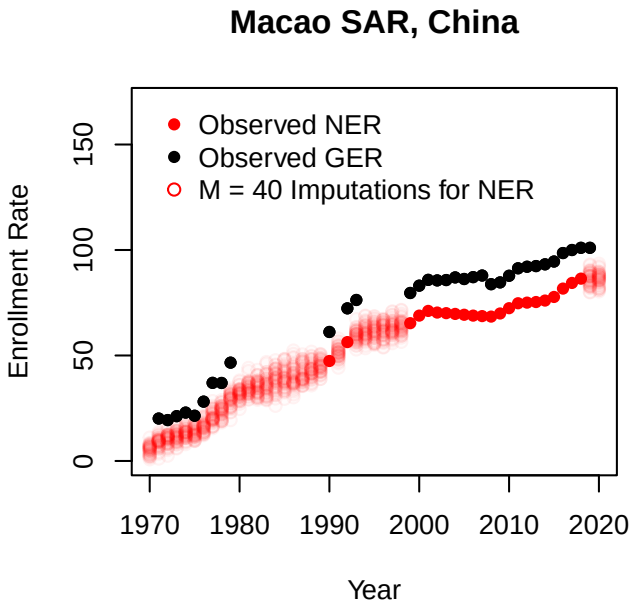

Sri Lanka
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

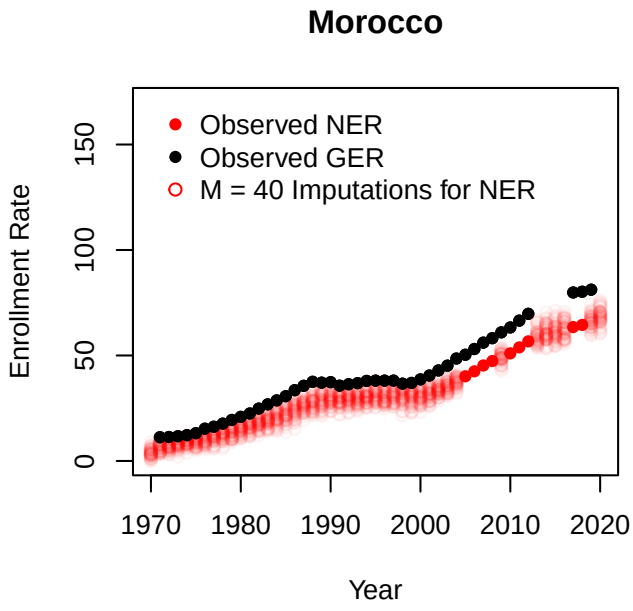

Lesotho
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

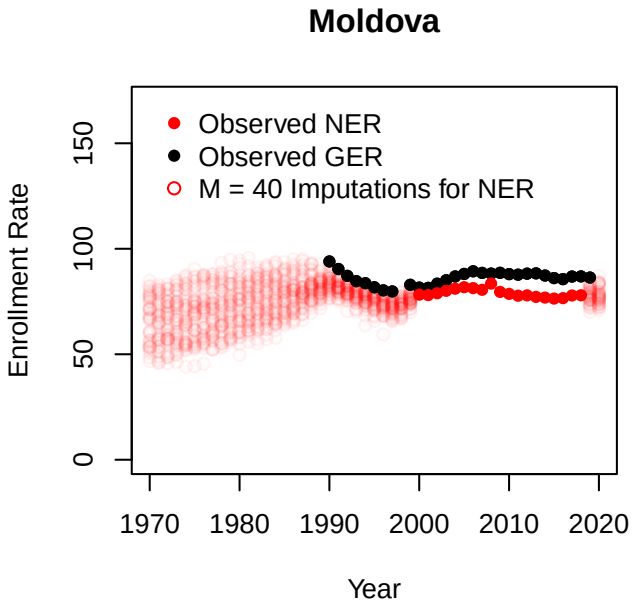

Lithuania
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

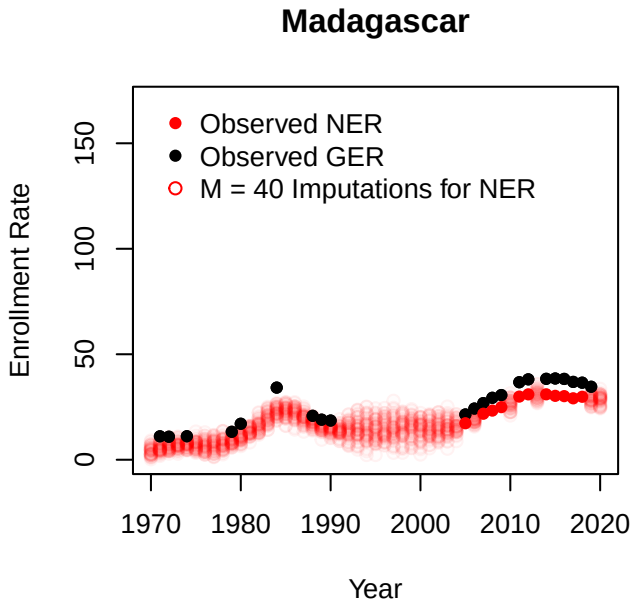

Luxembourg
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

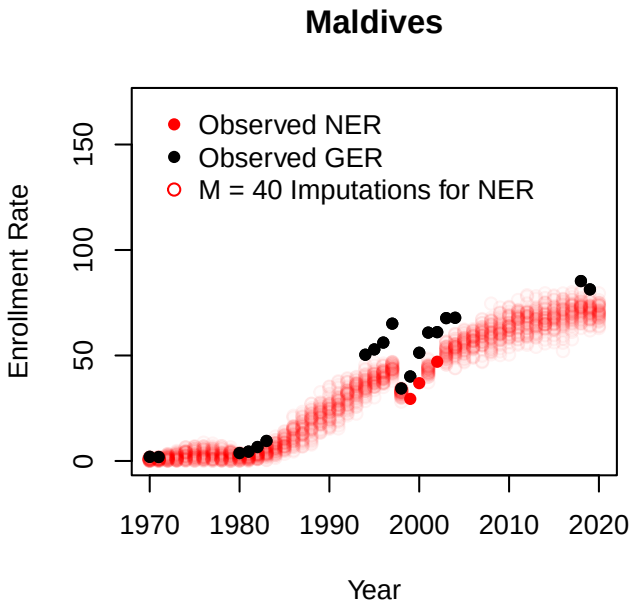

Latvia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

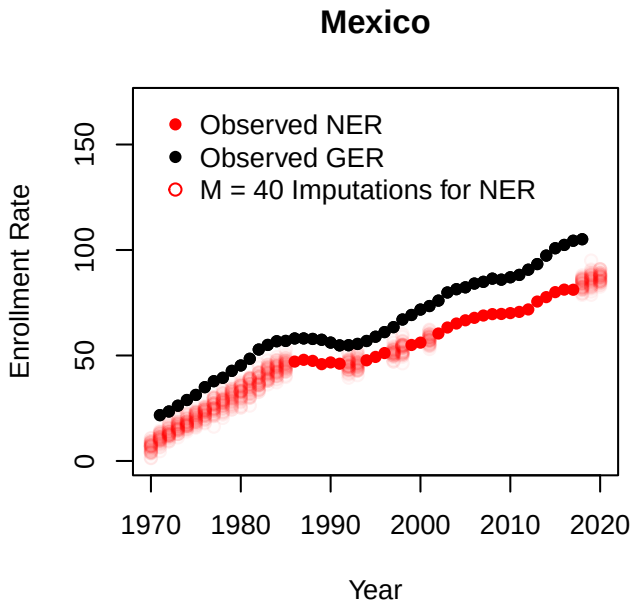

Macao SAR, China
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

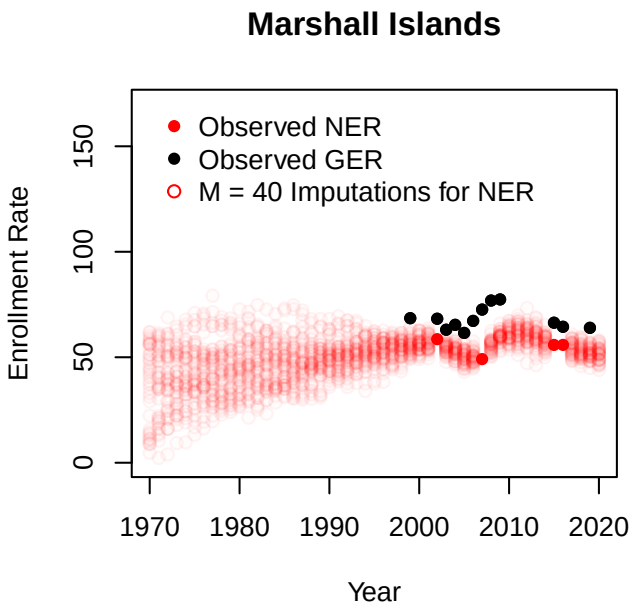

Morocco
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

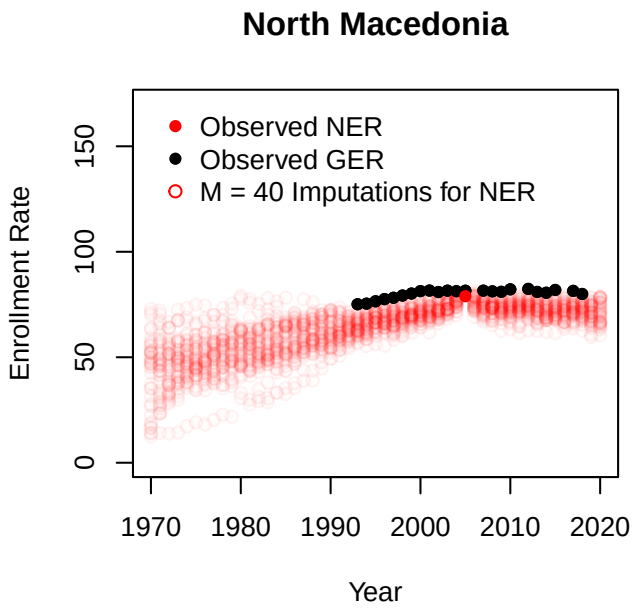

Moldova
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

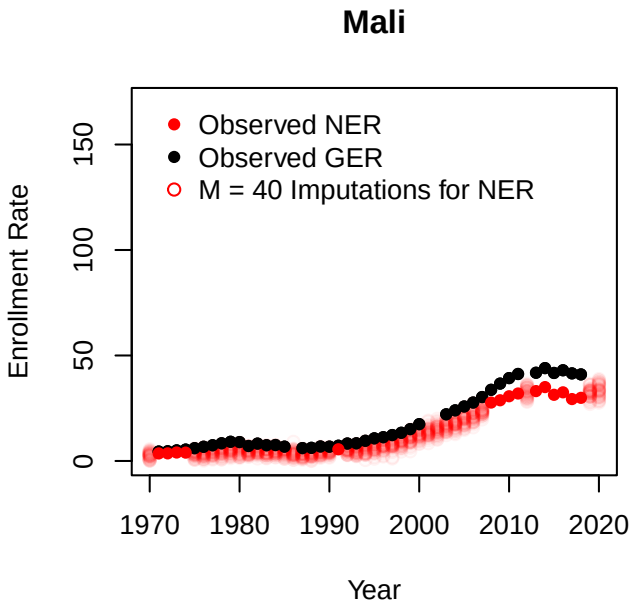

Madagascar
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

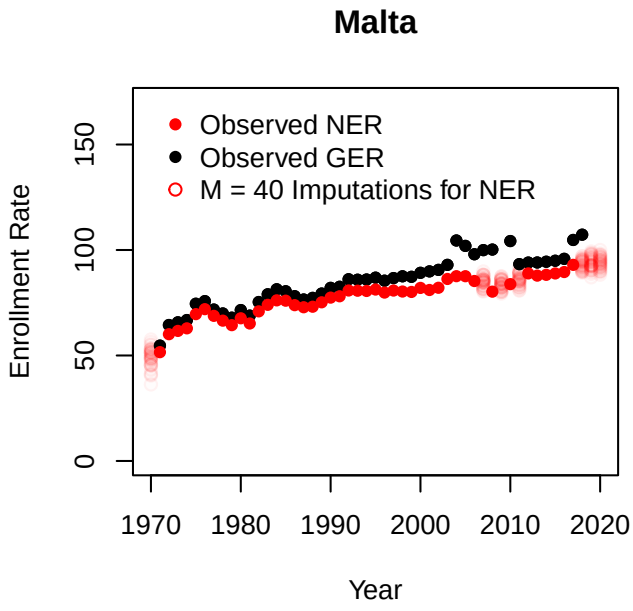

Maldives
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

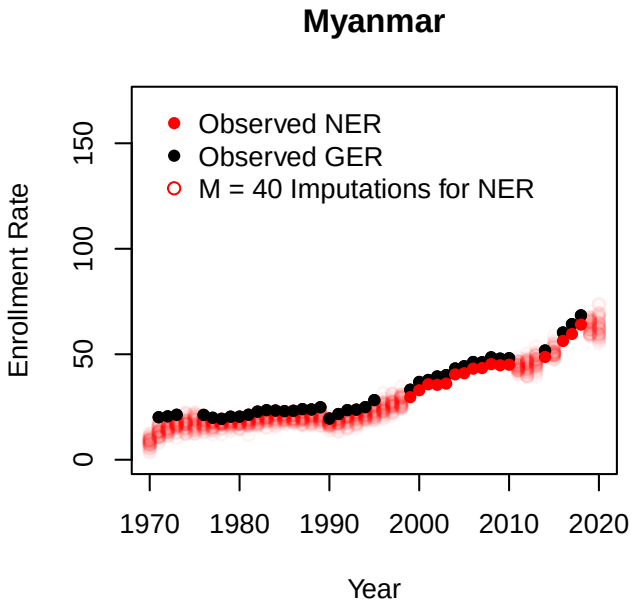

Mexico
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

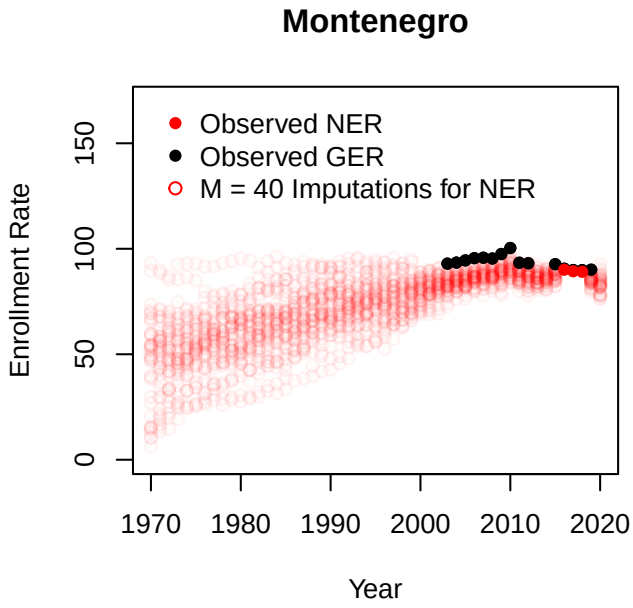

Marshall Islands
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

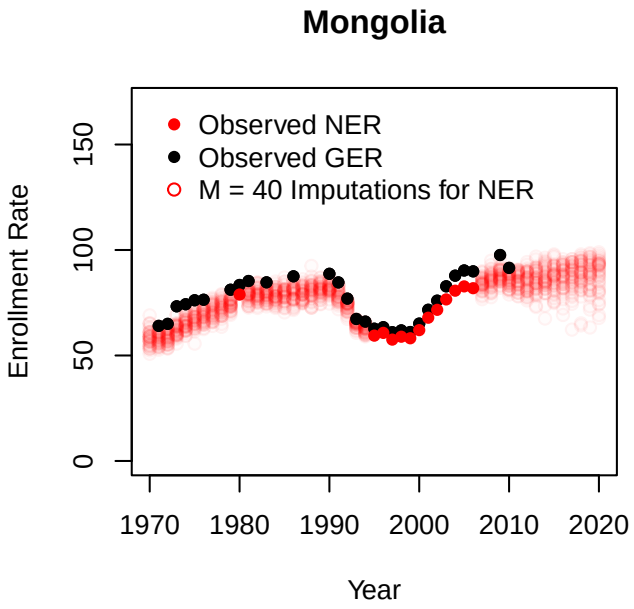

North Macedonia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

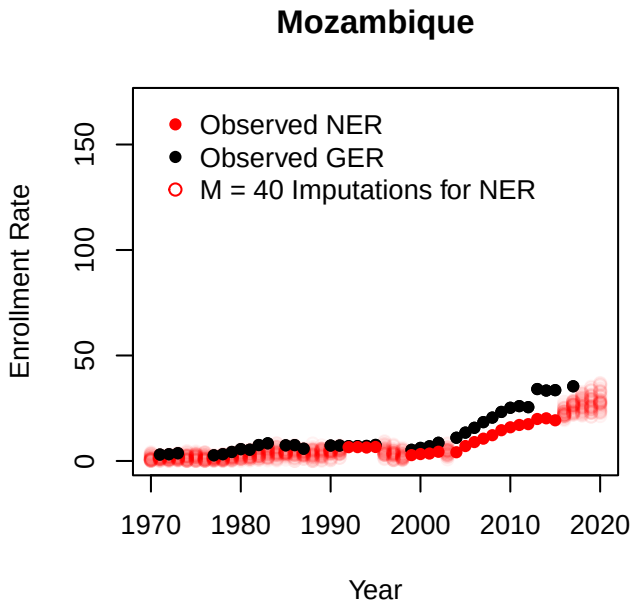

Mali
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

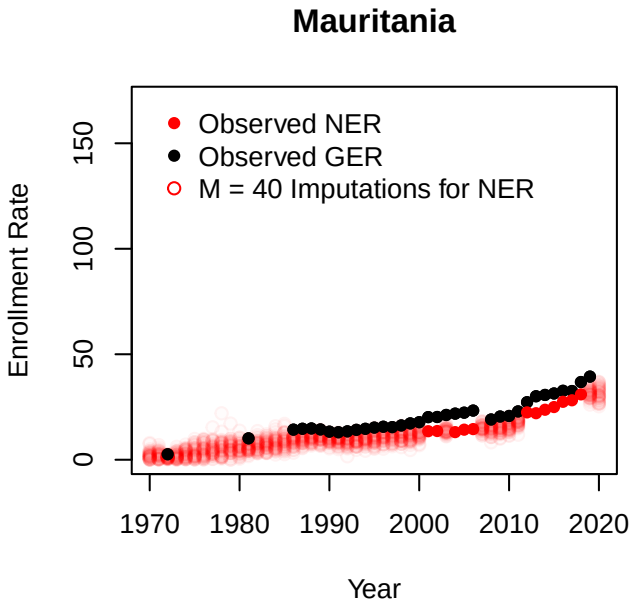

Malta
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

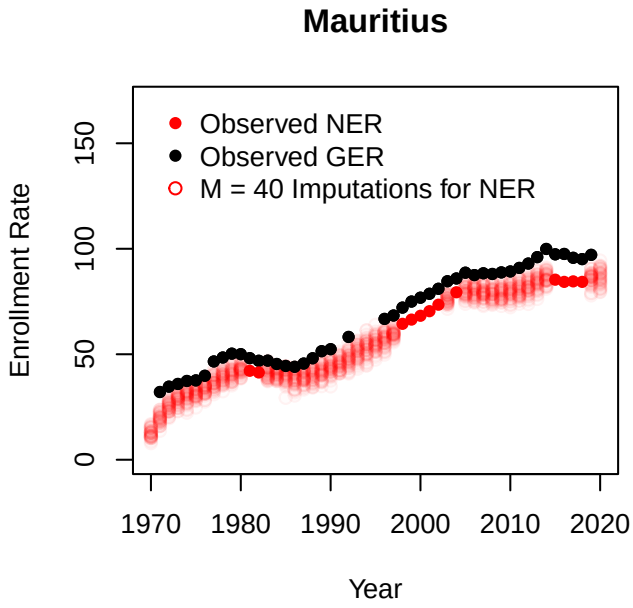

Myanmar
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

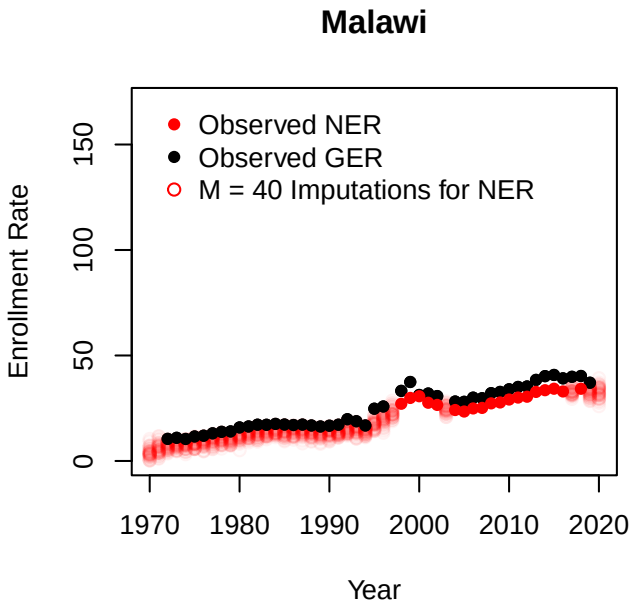

Montenegro
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

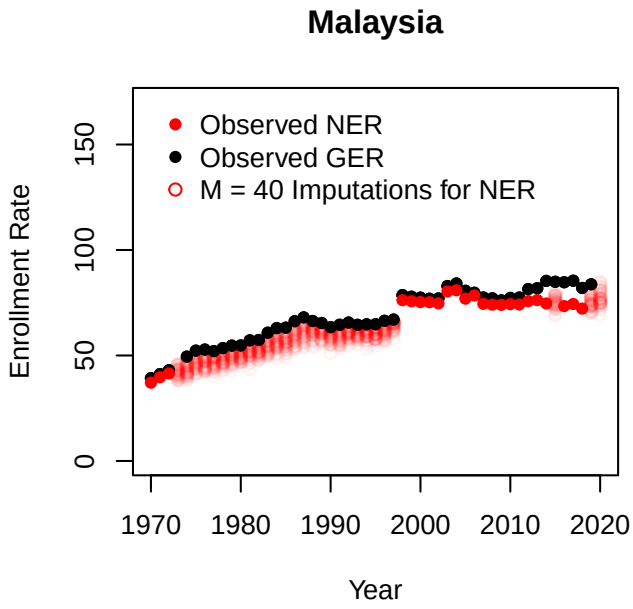

Mongolia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

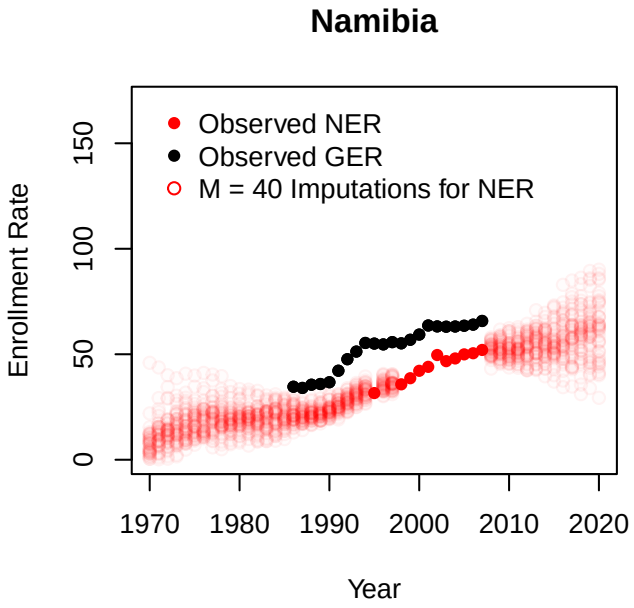

Mozambique
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

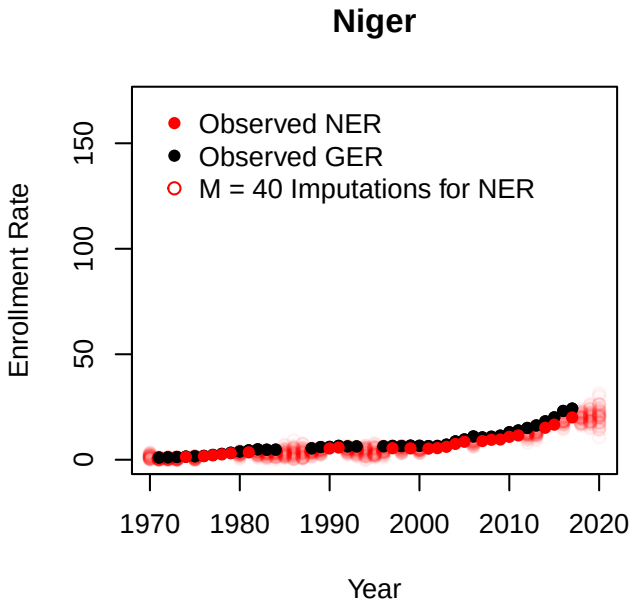

Mauritania
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

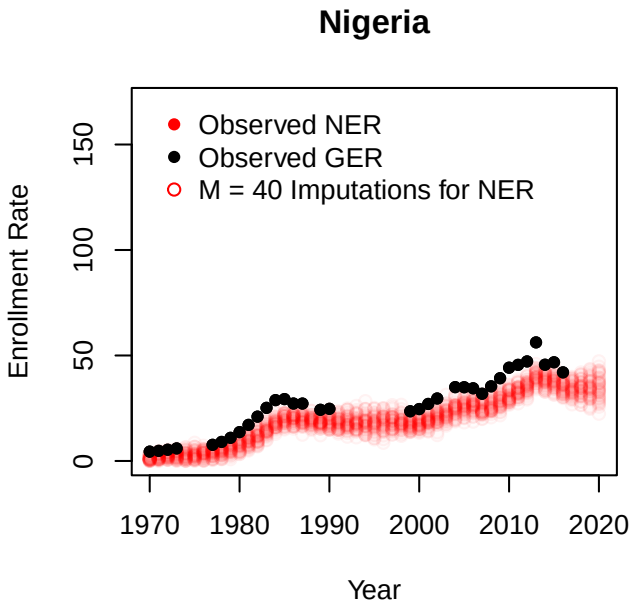

Mauritius
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

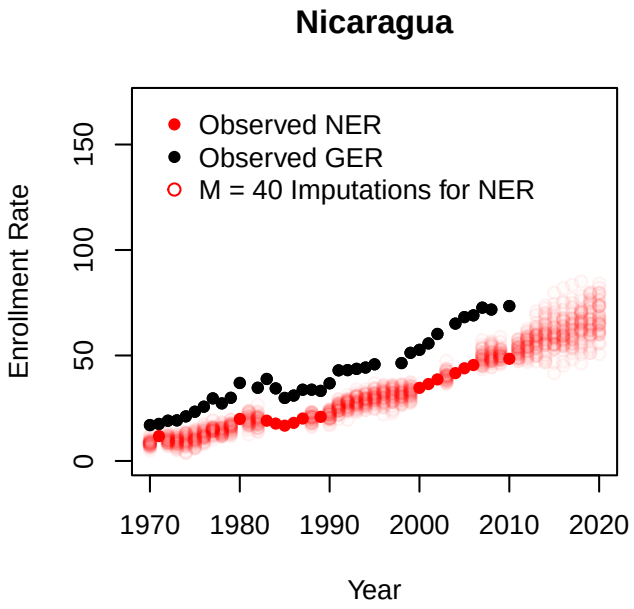

Malawi
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

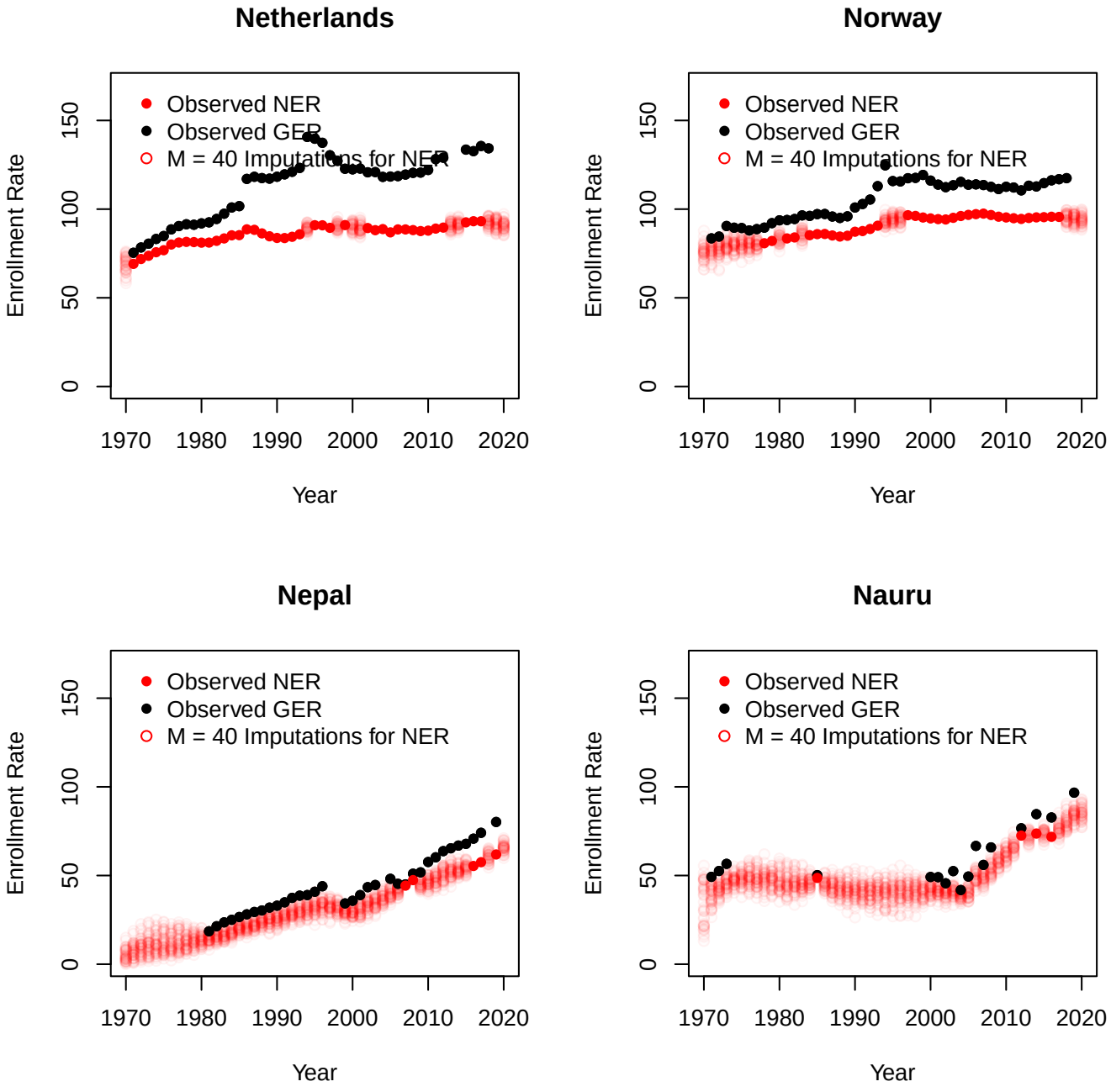

Malaysia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

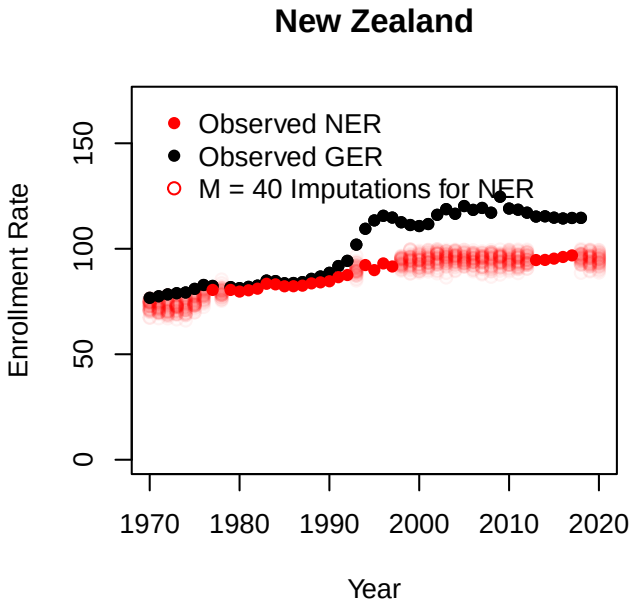

Namibia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

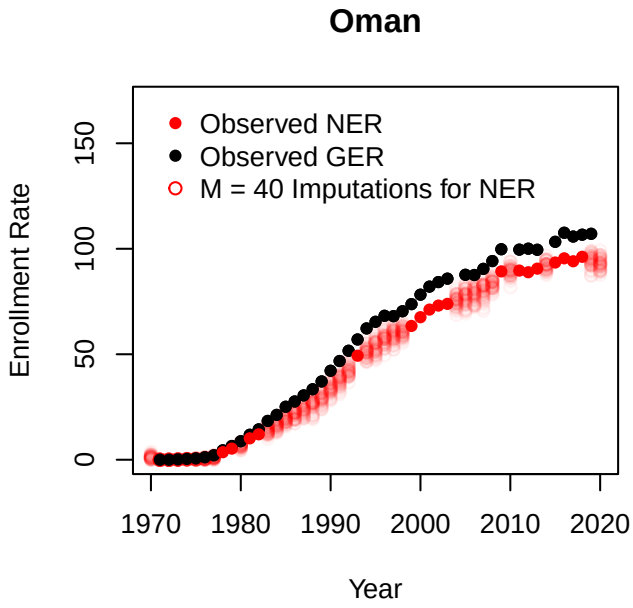

Niger
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

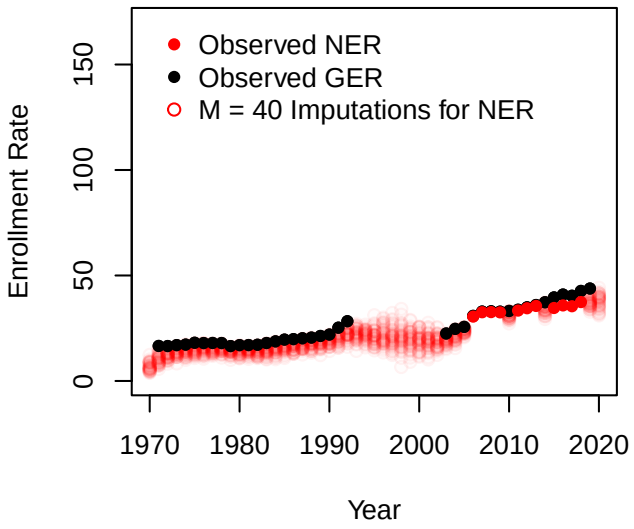

Nigeria
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

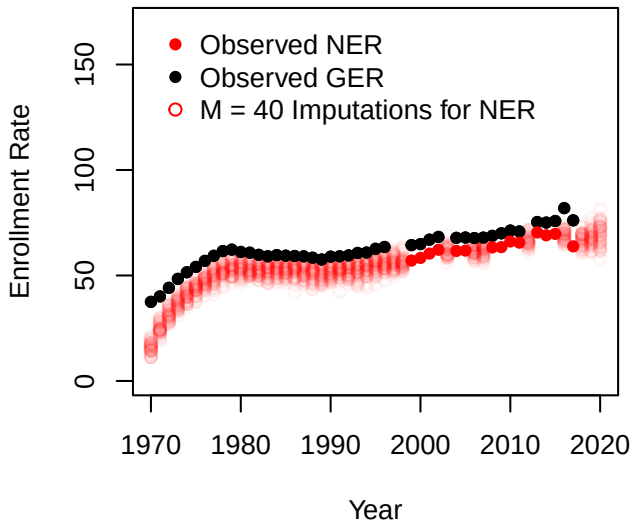

Nicaragua
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

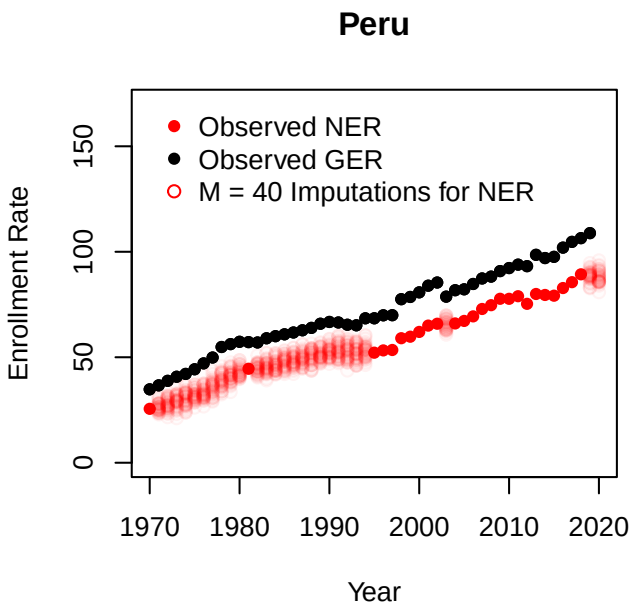

Netherlands
Norway
Nepal
Nauru
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
Year
0
50
100
150
1970
1980
1990
2000
2010
2020

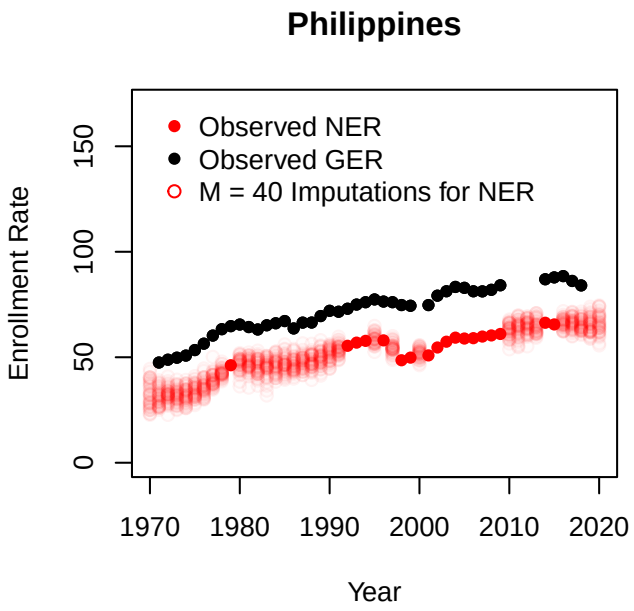

New Zealand
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

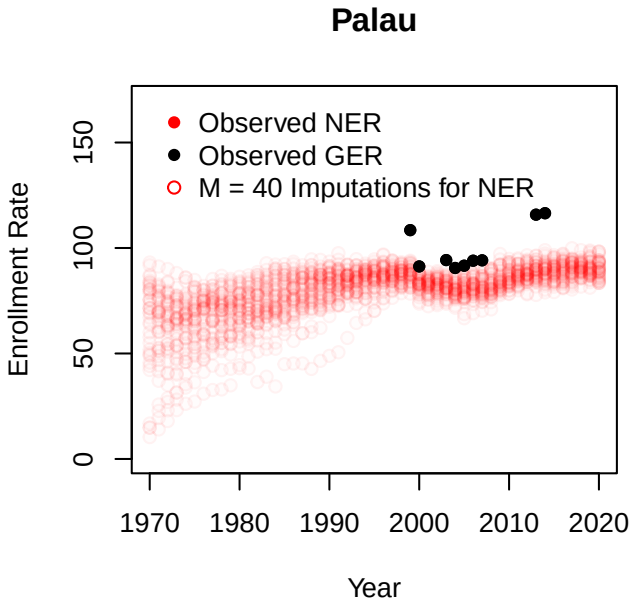

Oman
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

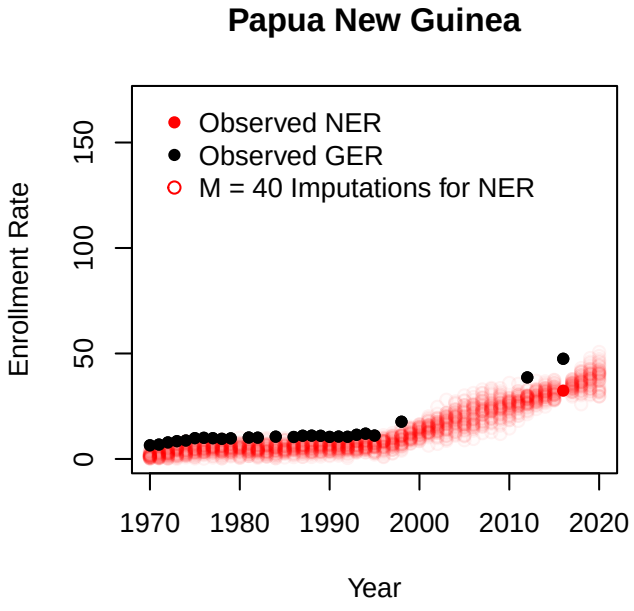

Pakistan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

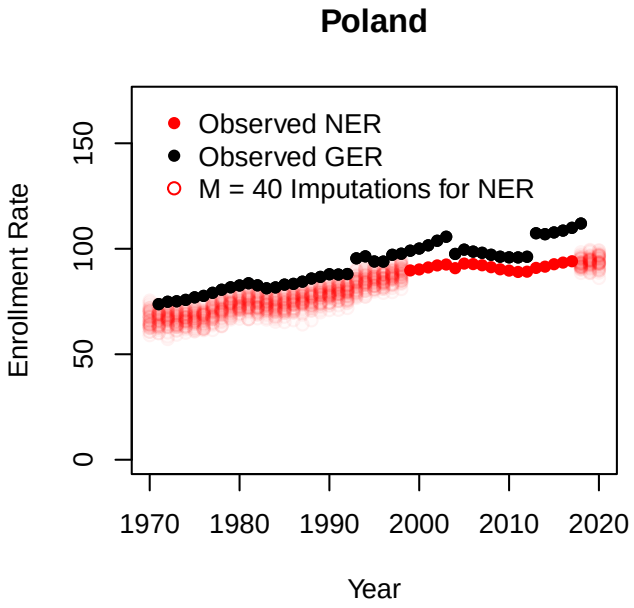

Panama
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

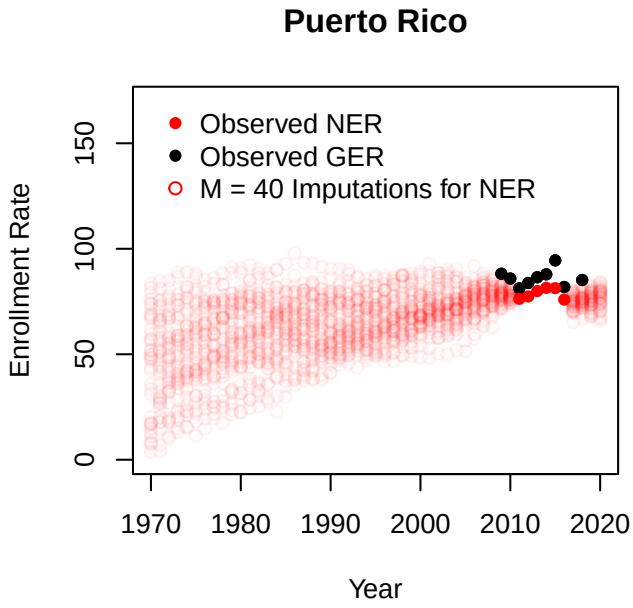

Peru
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

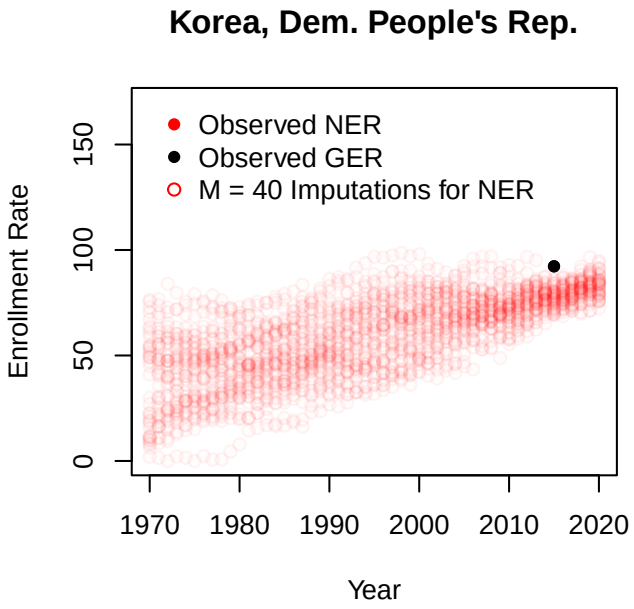

Philippines
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

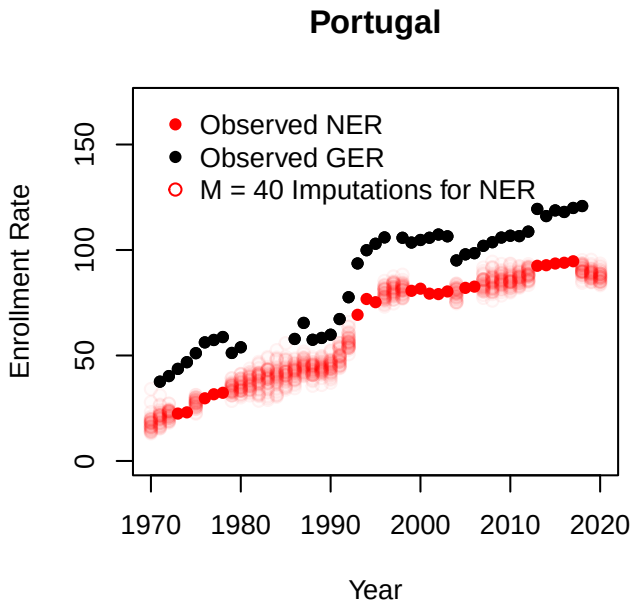

Palau
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

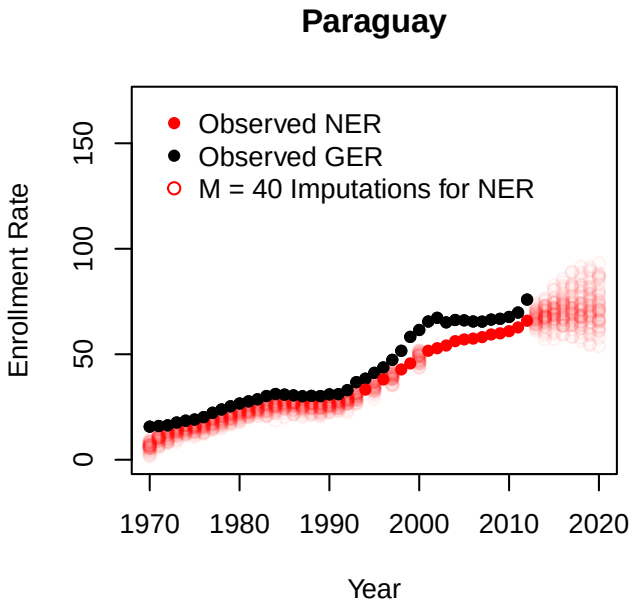

Papua New Guinea
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

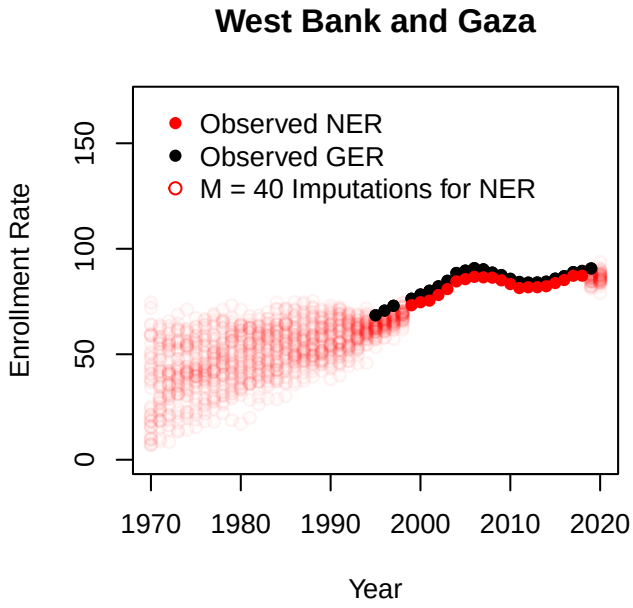

Poland
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

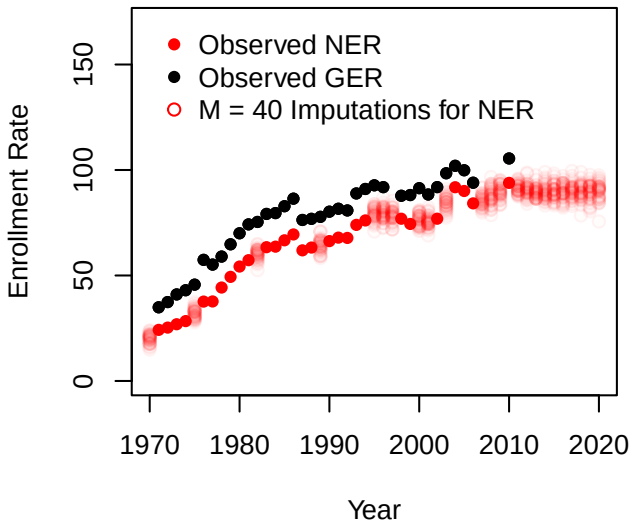

Puerto Rico
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

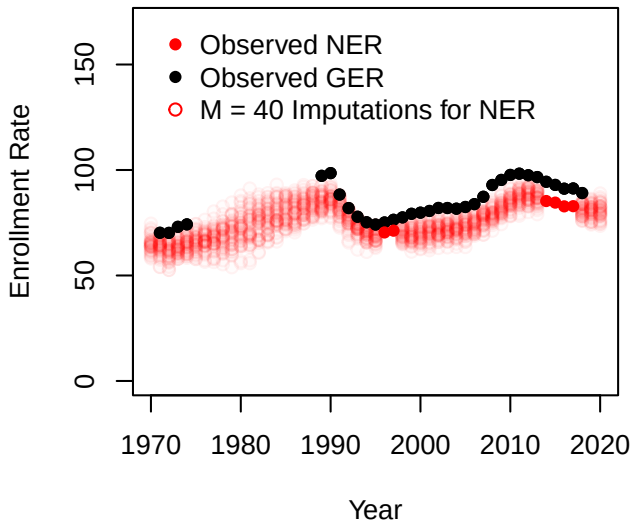

Korea, Dem. People's Rep.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

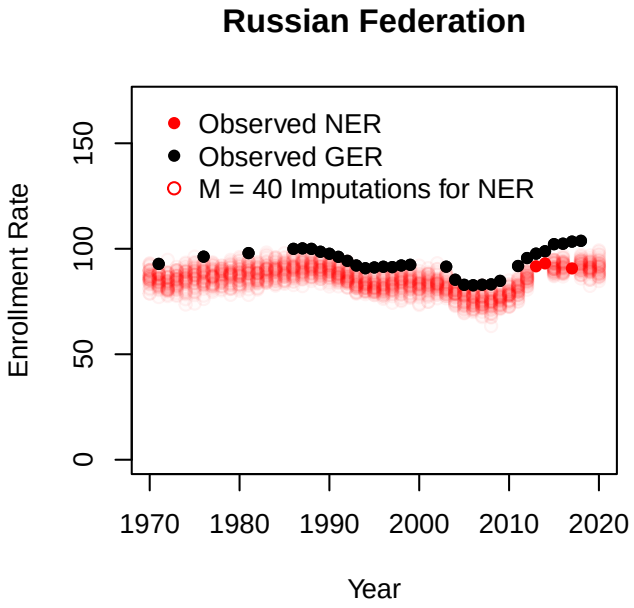

Portugal
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

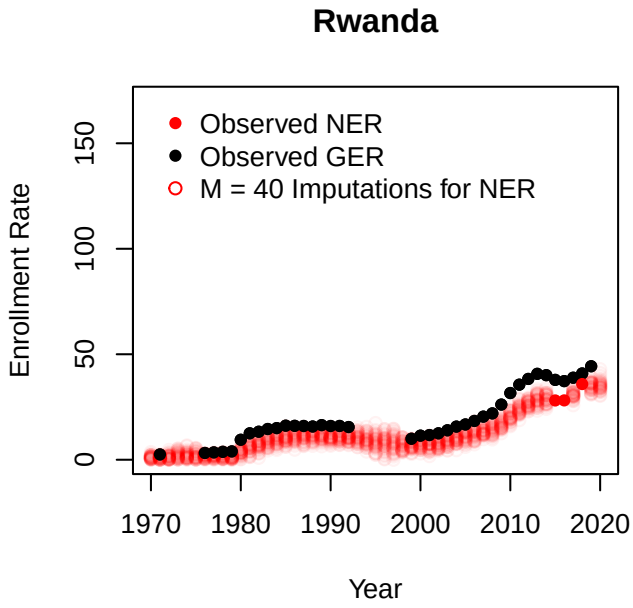

Paraguay
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

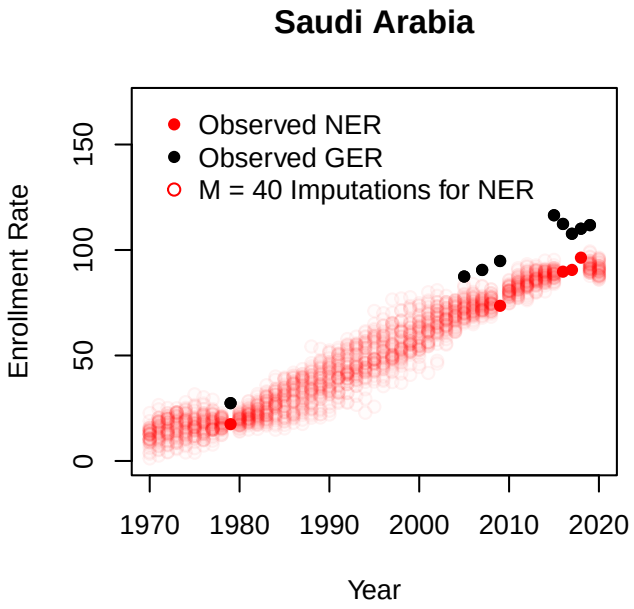

West Bank and Gaza
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

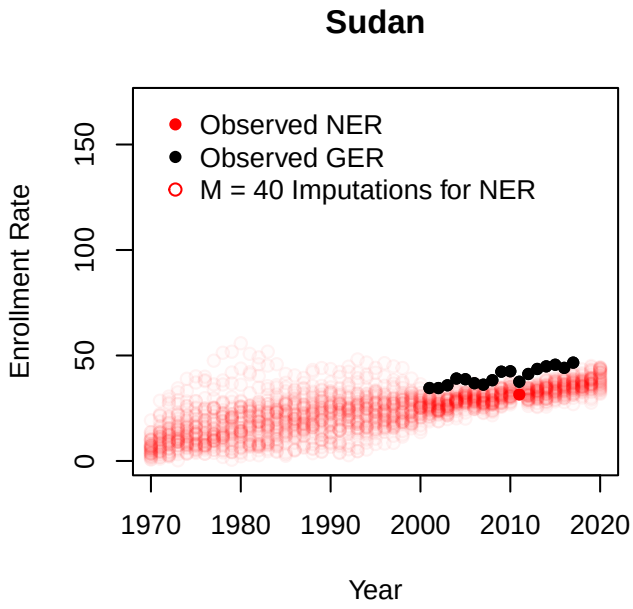

Qatar
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

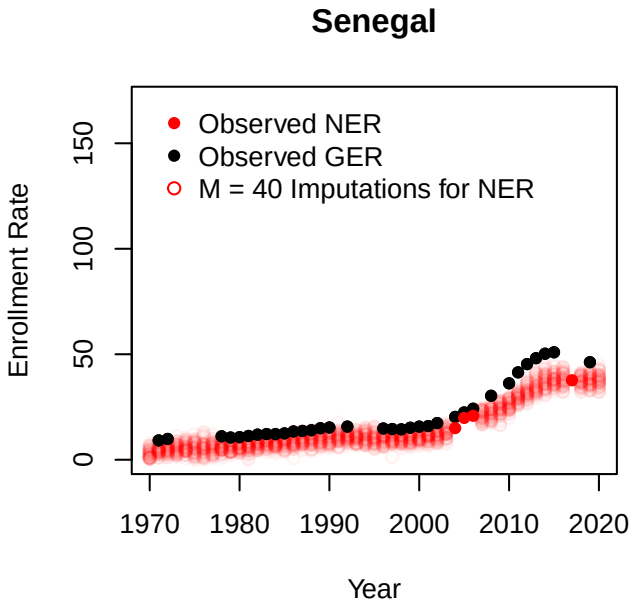

Romania
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

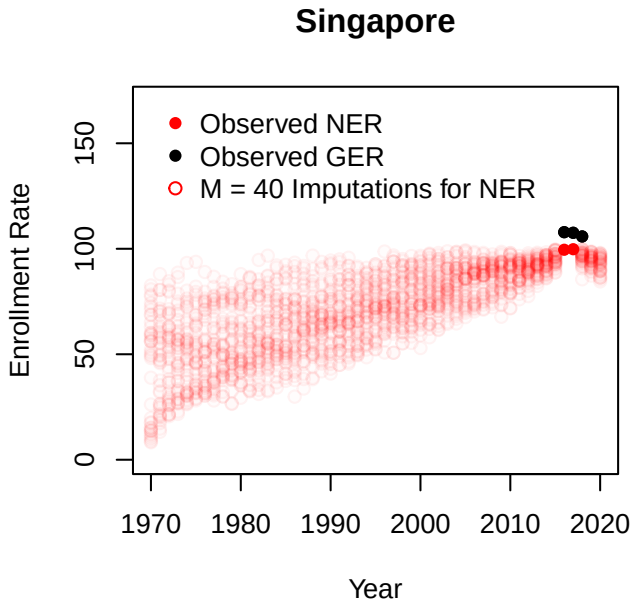

Russian Federation
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

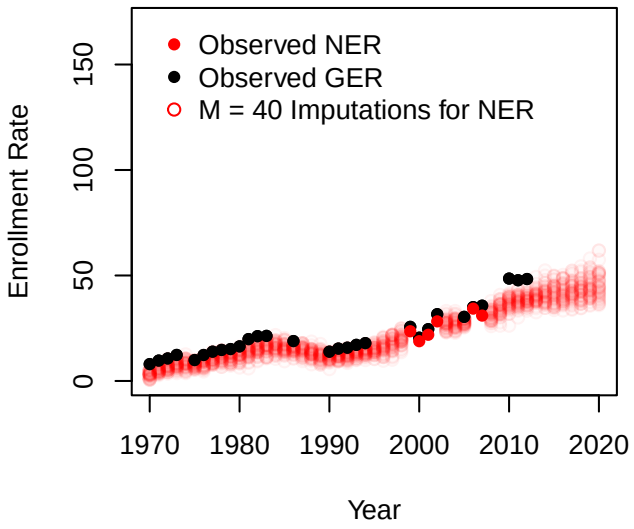

Rwanda
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

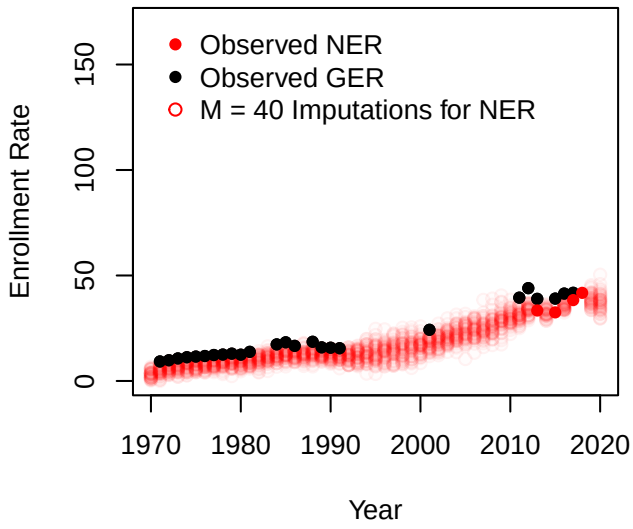

Saudi Arabia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

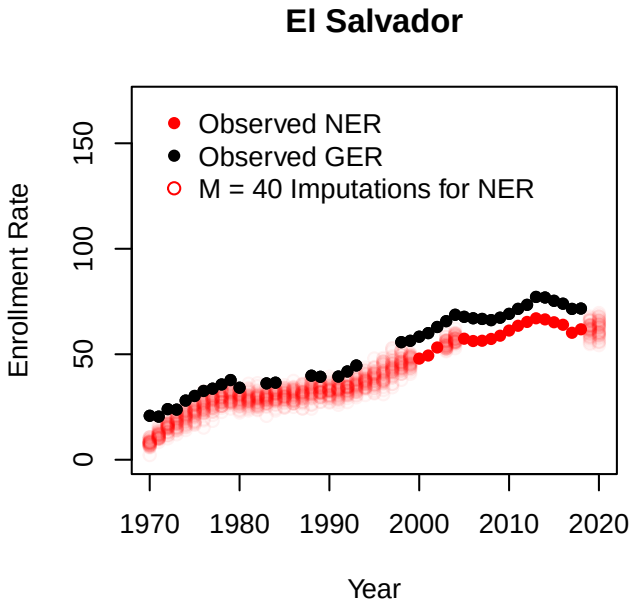

Sudan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

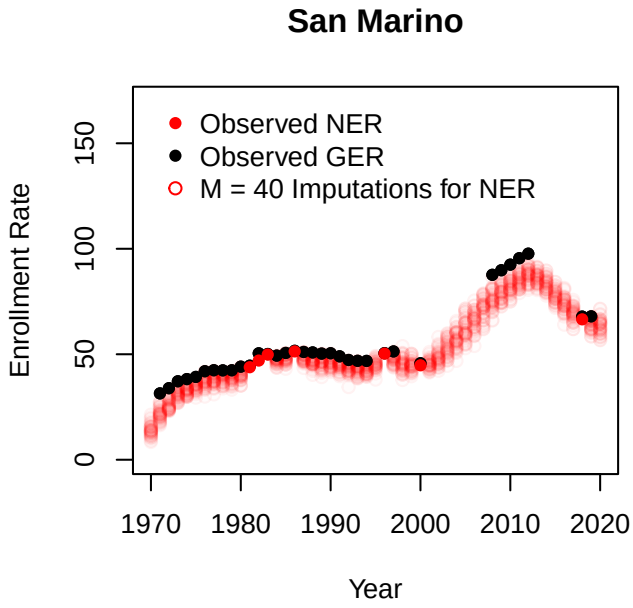

Senegal
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

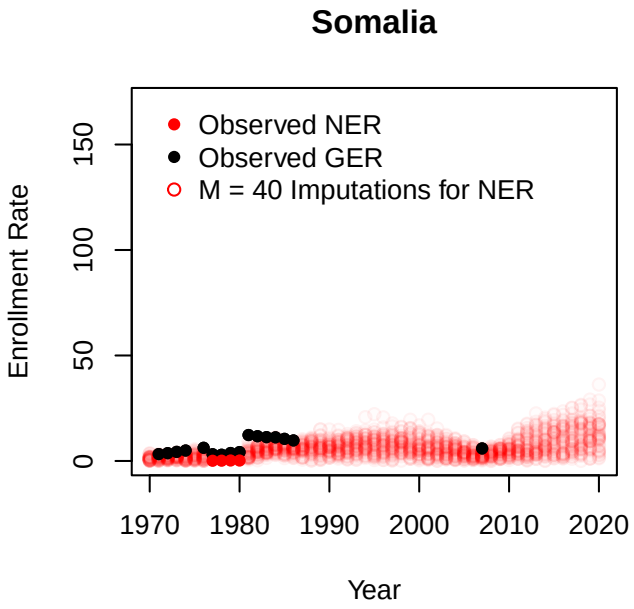

Singapore
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

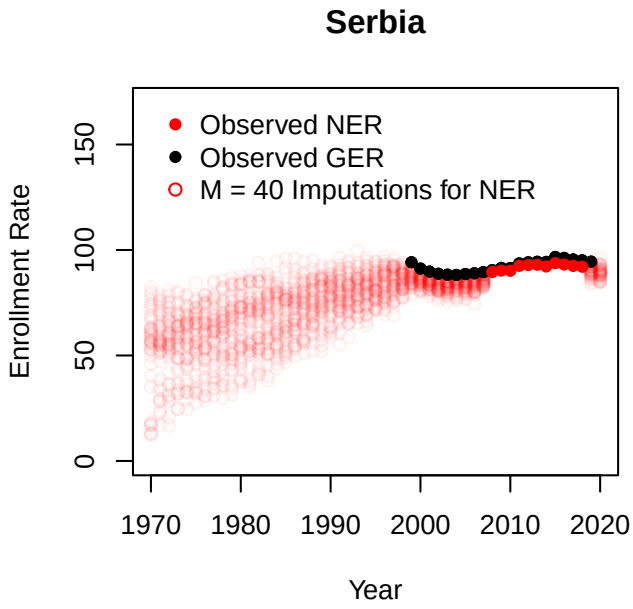

Solomon Islands
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

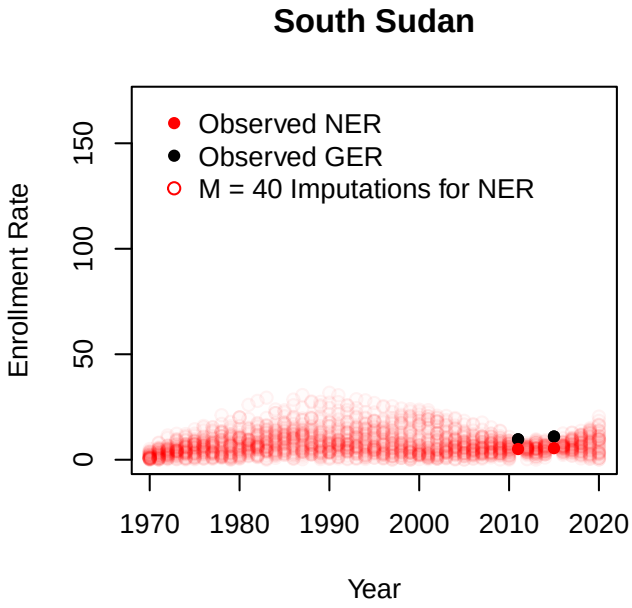

Sierra Leone
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

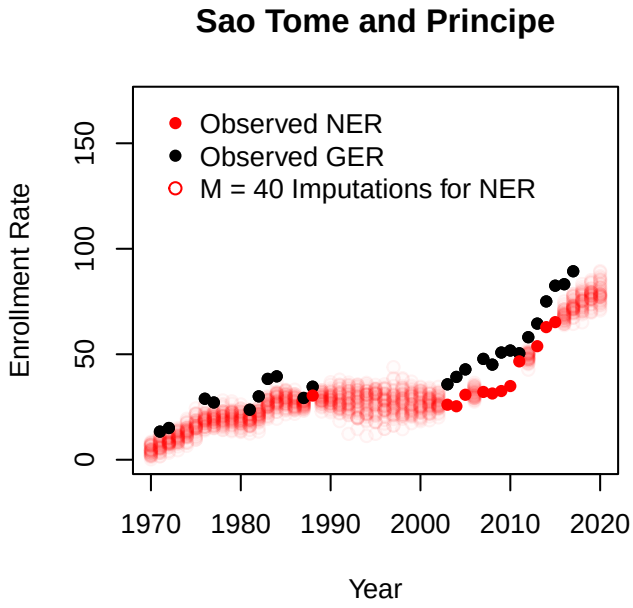

El Salvador
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

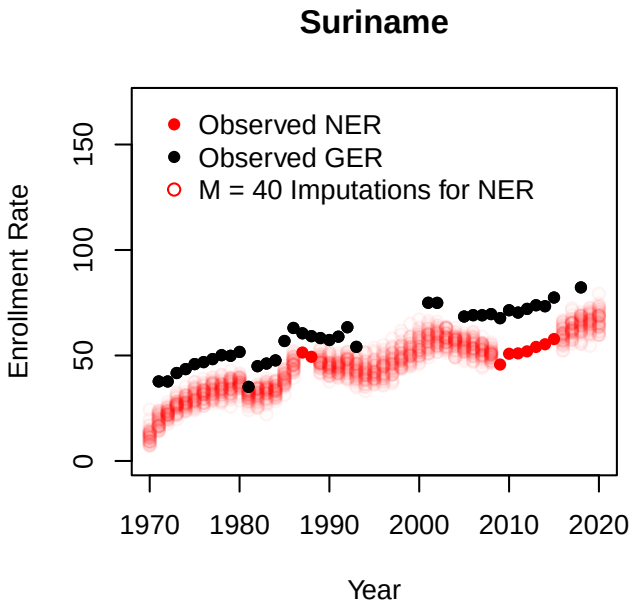

San Marino
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

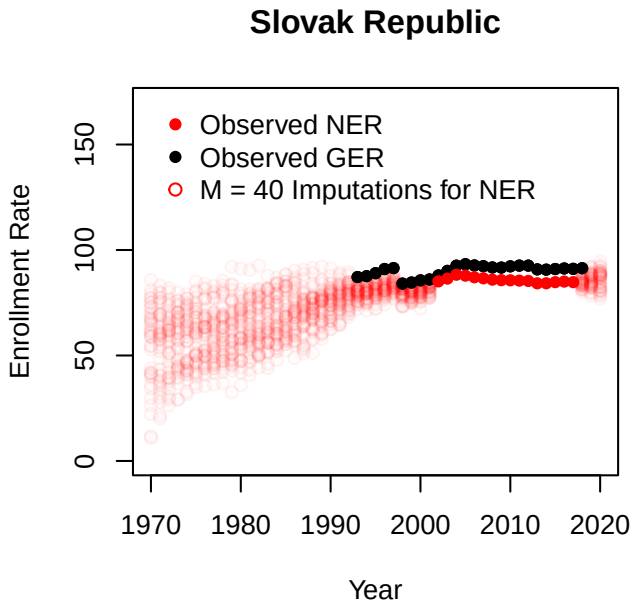

Somalia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

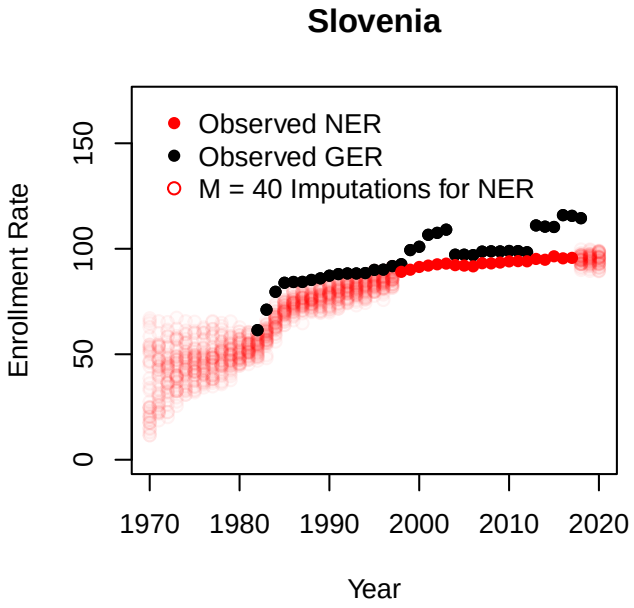

Serbia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

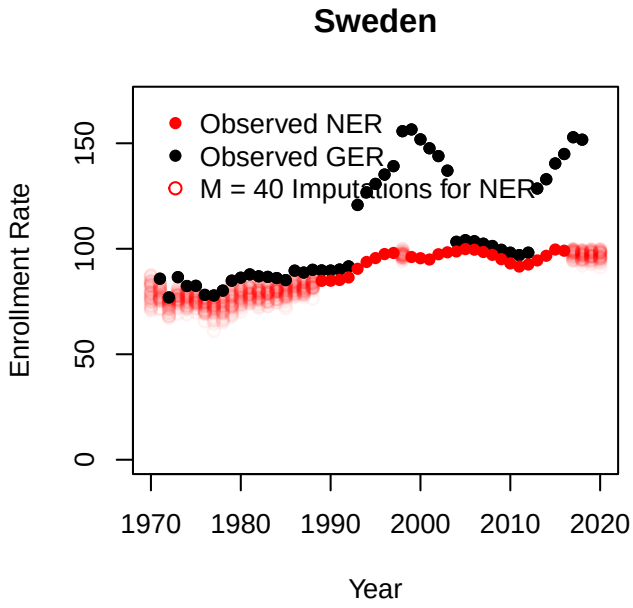

South Sudan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

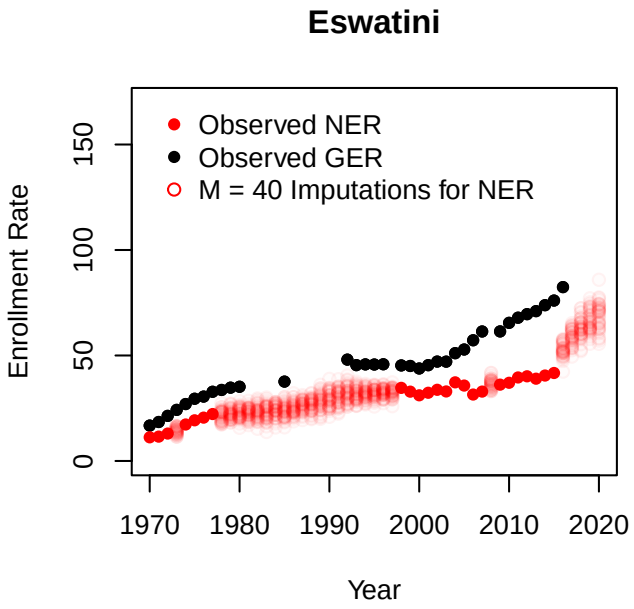

Sao Tome and Principe
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

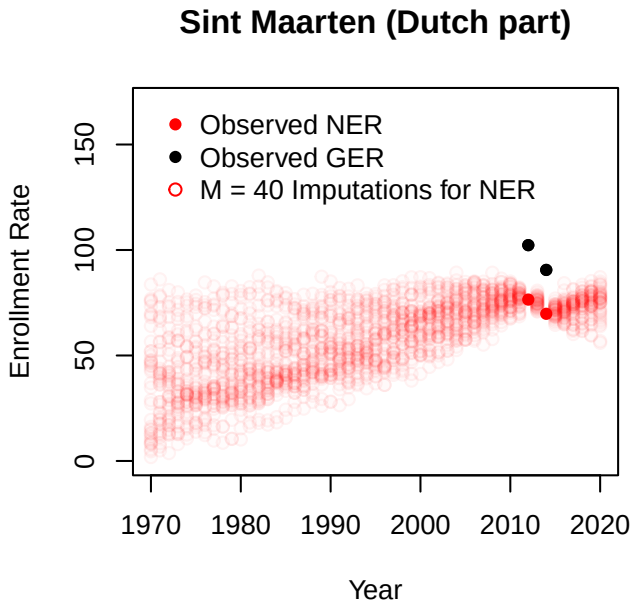

Suriname
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

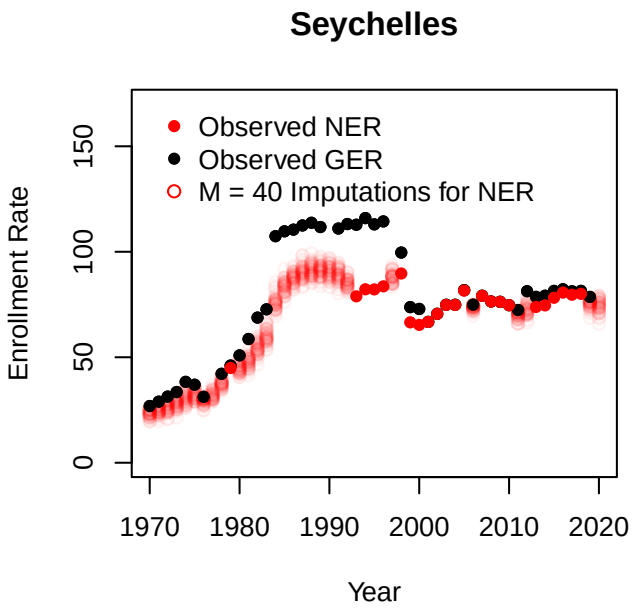

Slovak Republic
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

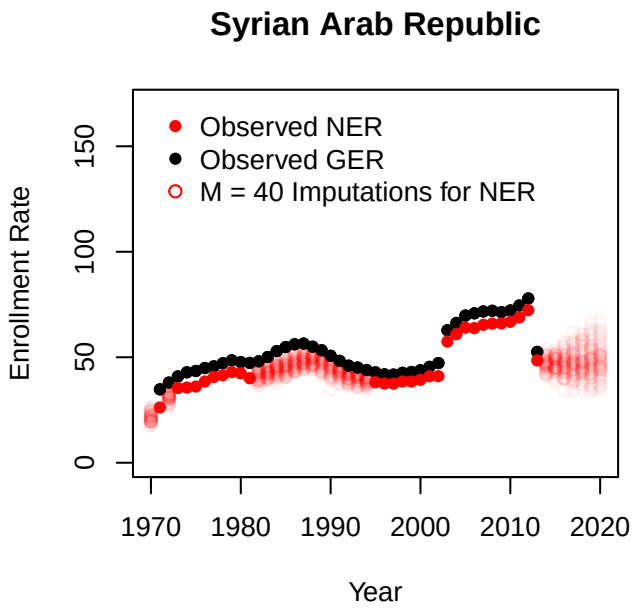

Slovenia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

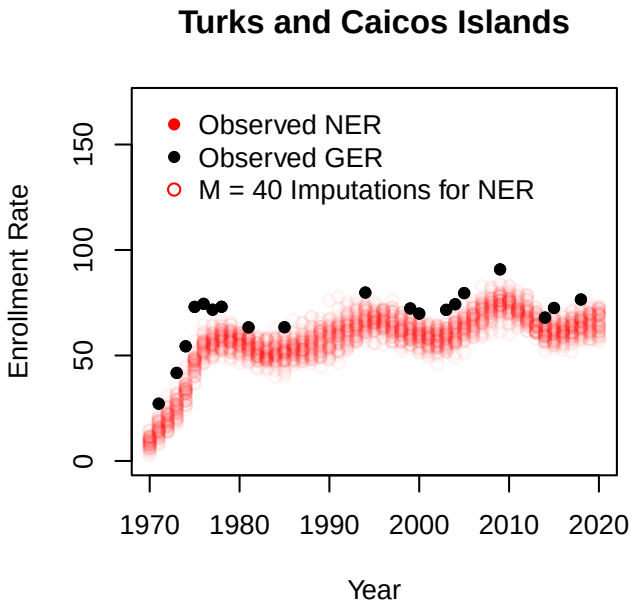

Sweden
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

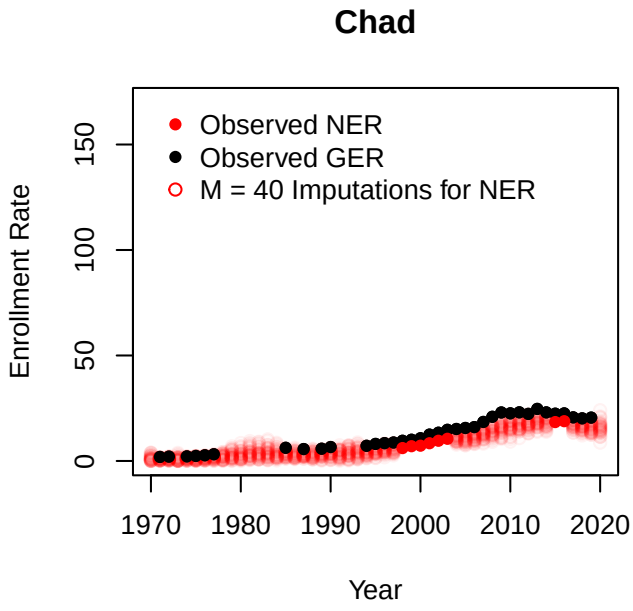

Eswatini
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

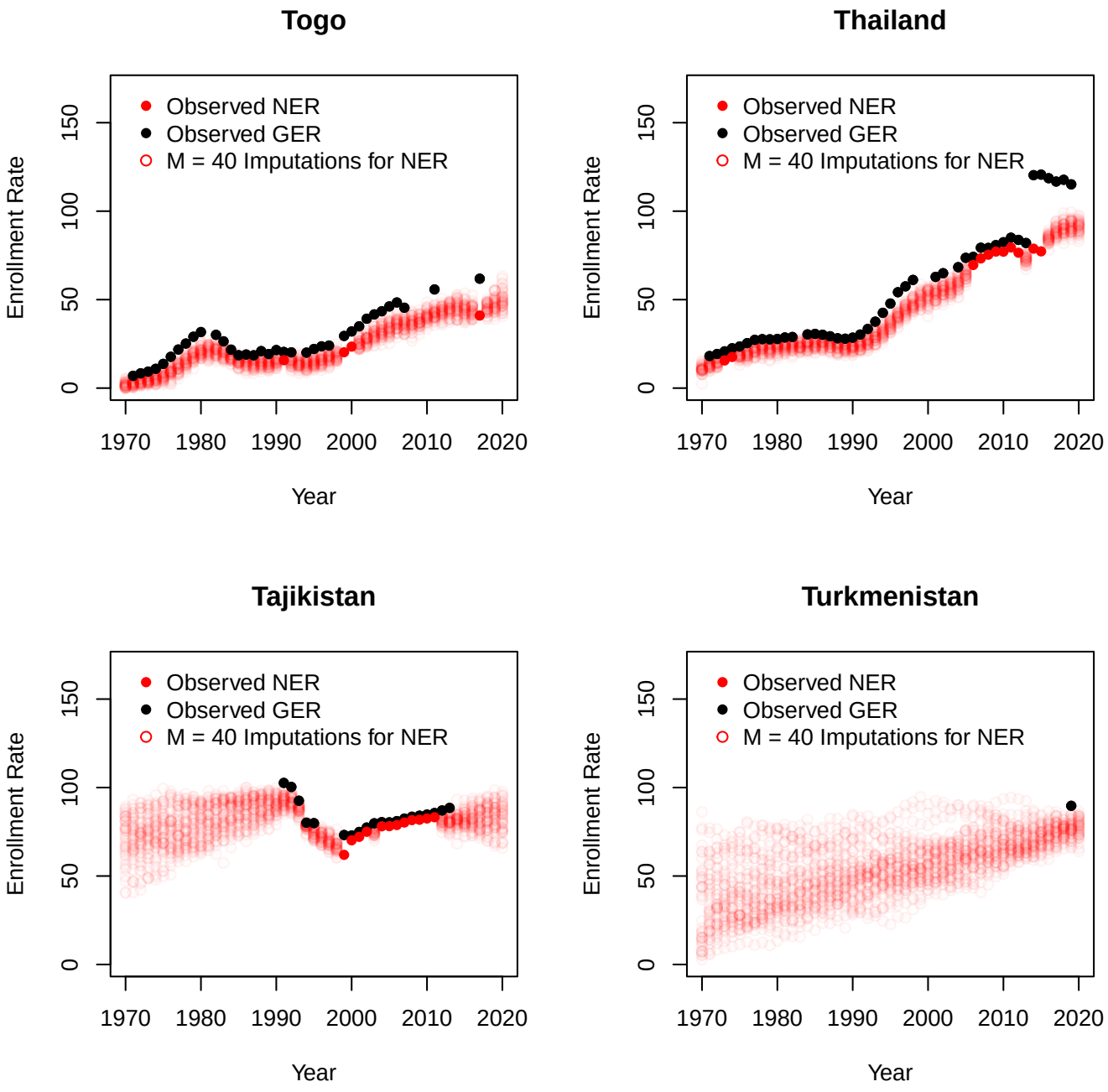

Sint Maarten (Dutch part)
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

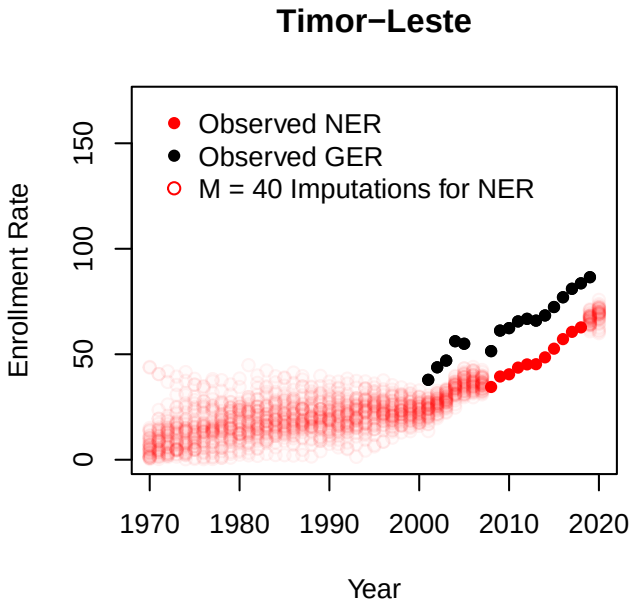

Seychelles
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

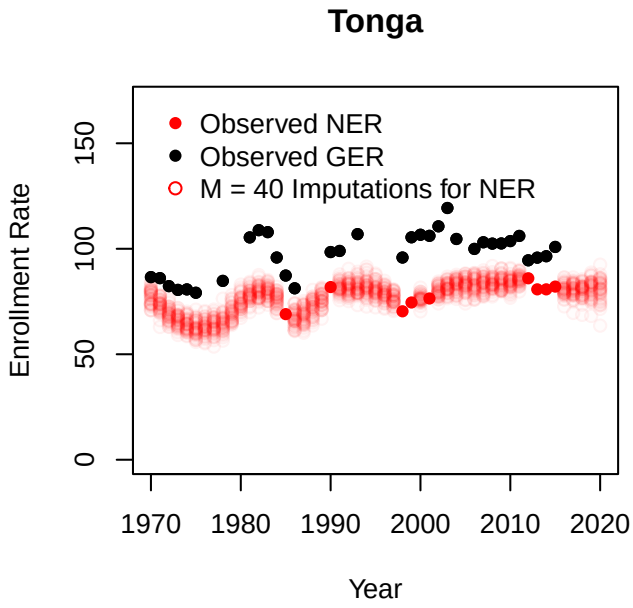

Syrian Arab Republic
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

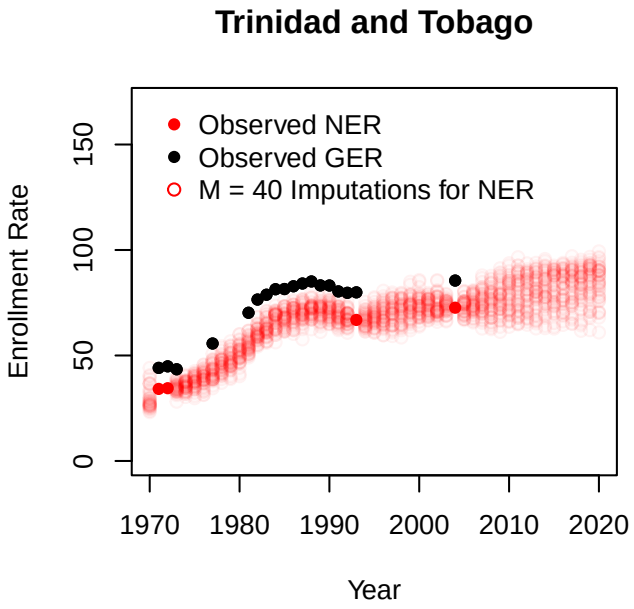

Turks and Caicos Islands
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

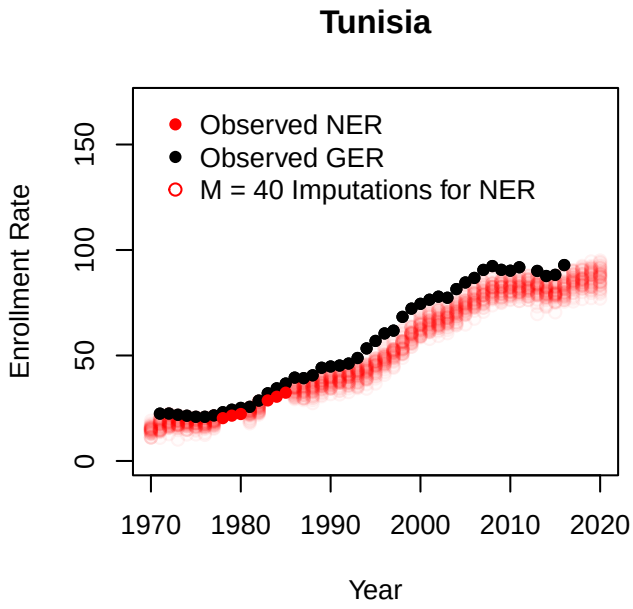

Chad
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

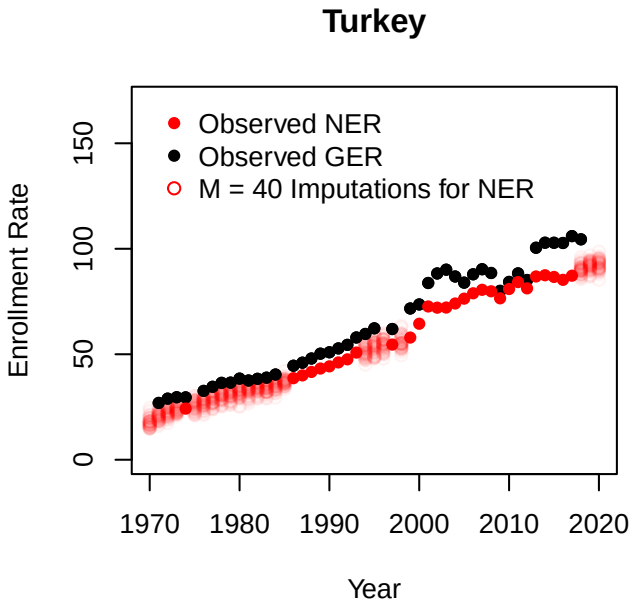

Togo
Thailand
Tajikistan
Turkmenistan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
Year
0
50
100
150
1970
1980
1990
2000
2010
2020

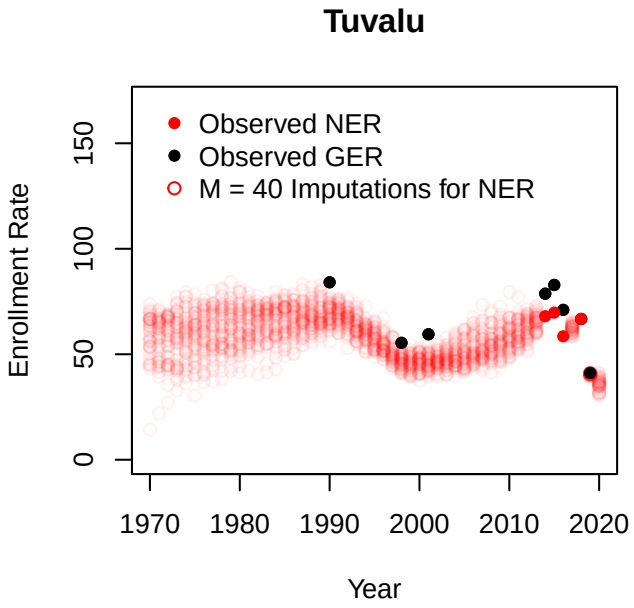

Timor−Leste
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

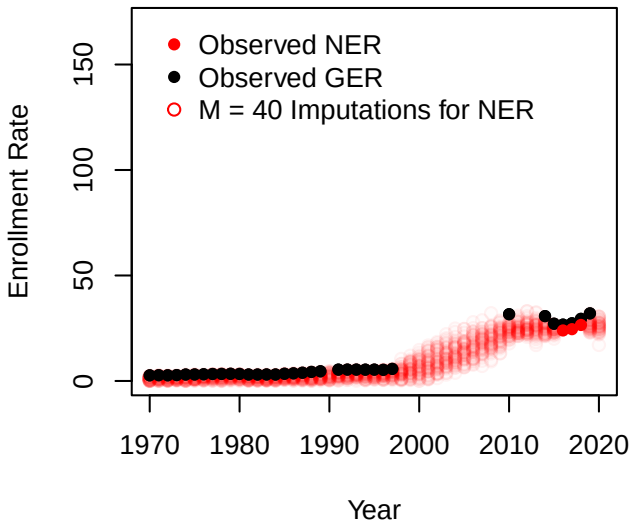

Tonga
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

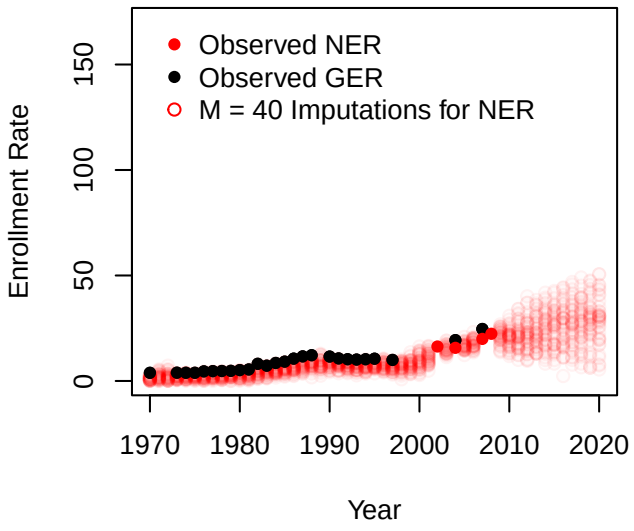

Trinidad and Tobago
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

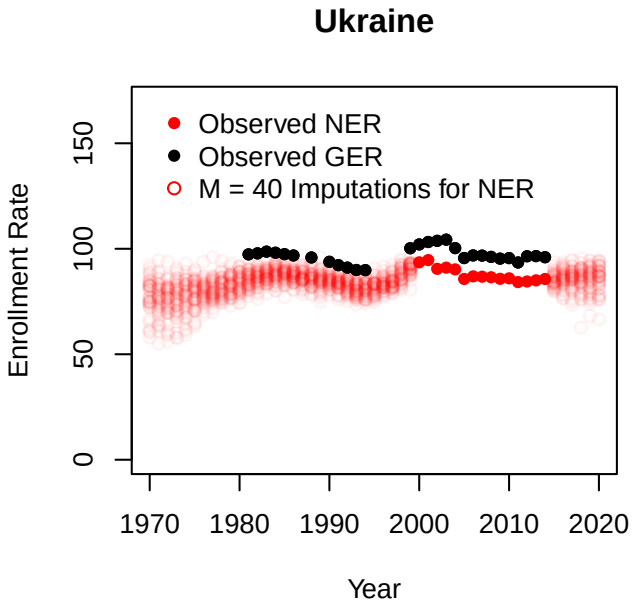

Tunisia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

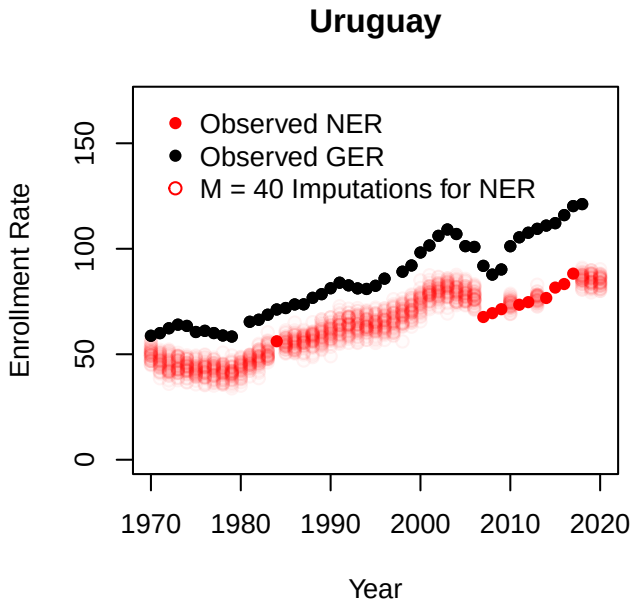

Turkey
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

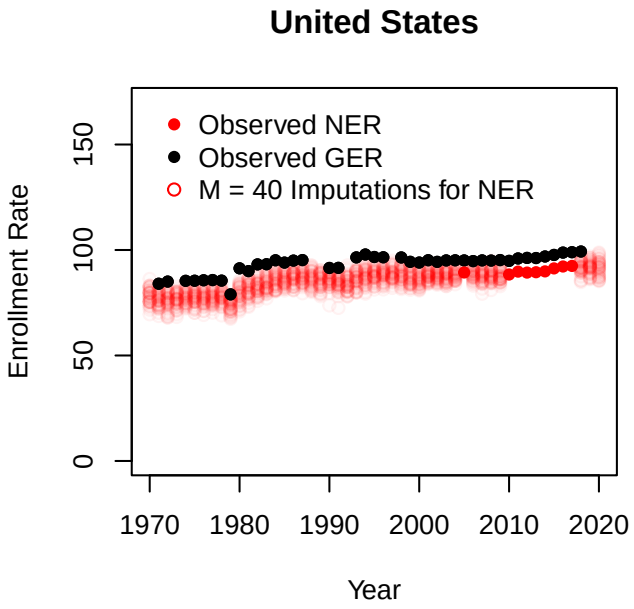

Tuvalu
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

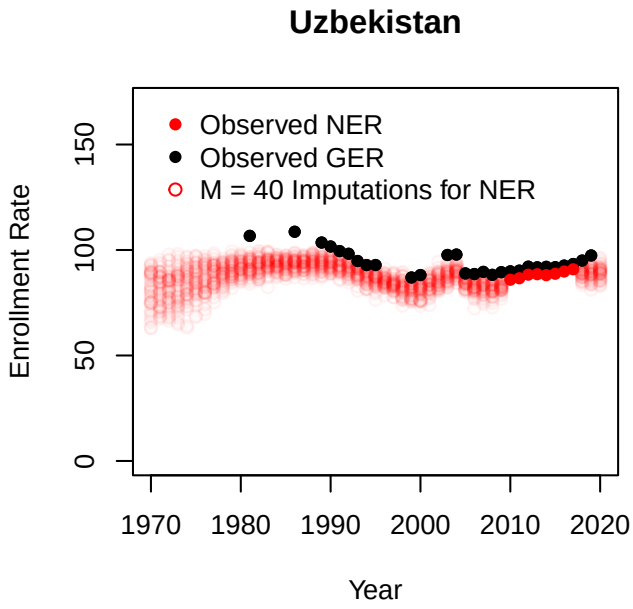

Tanzania
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

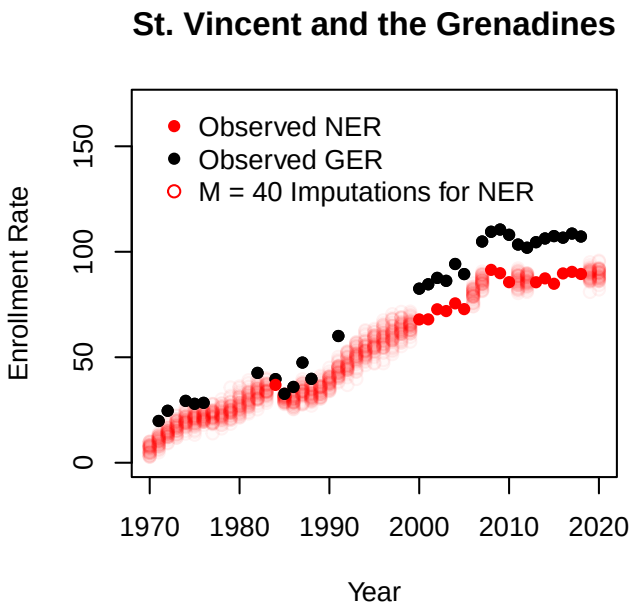

Uganda
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

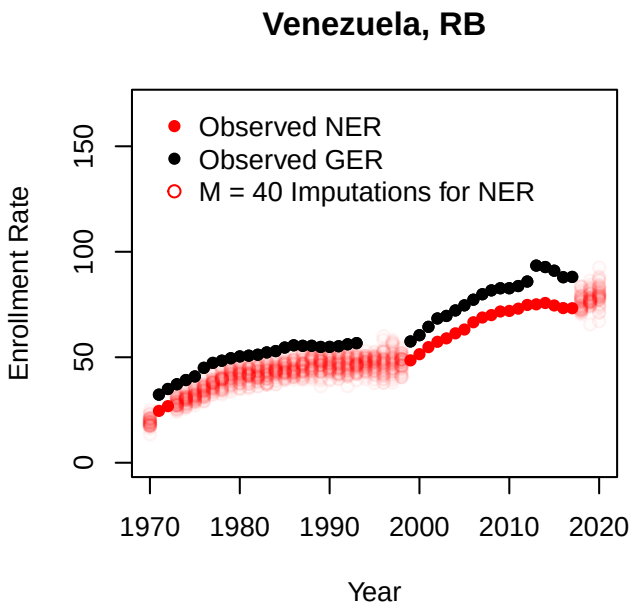

Ukraine
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

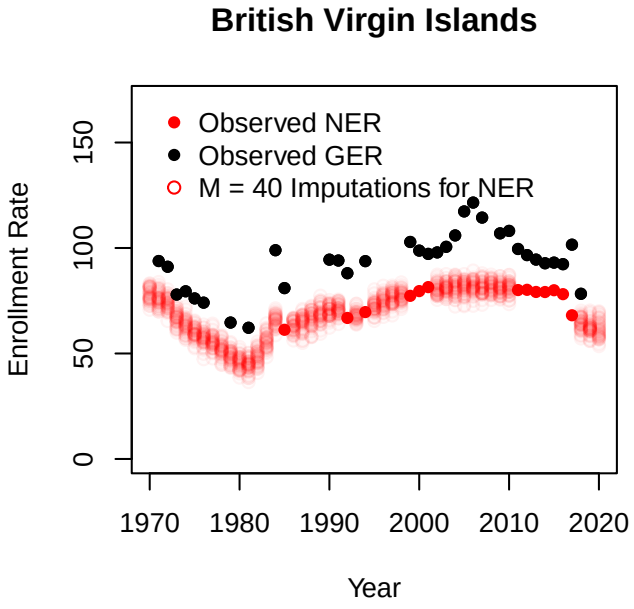

Uruguay
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

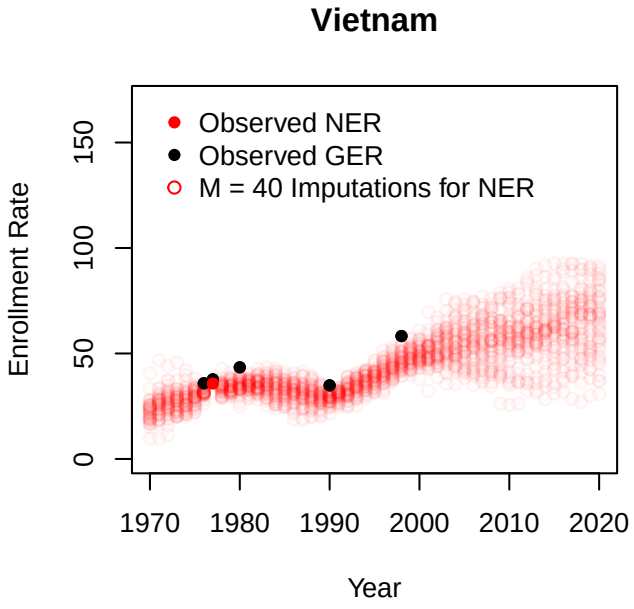

United States
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

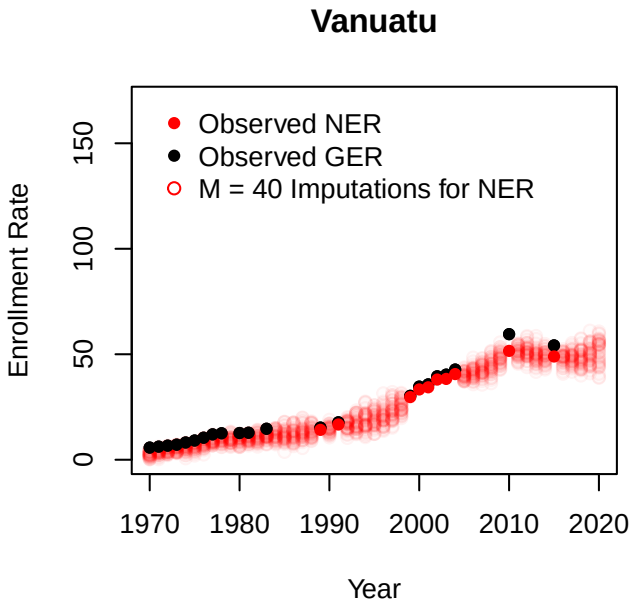

Uzbekistan
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

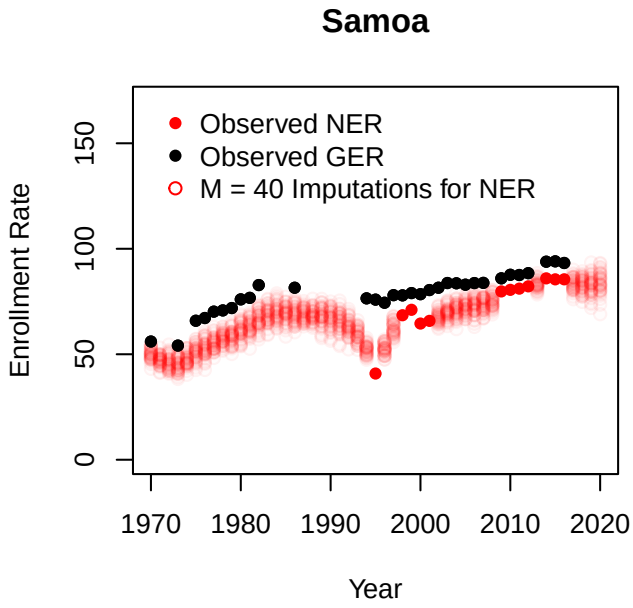

St. Vincent and the Grenadines
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

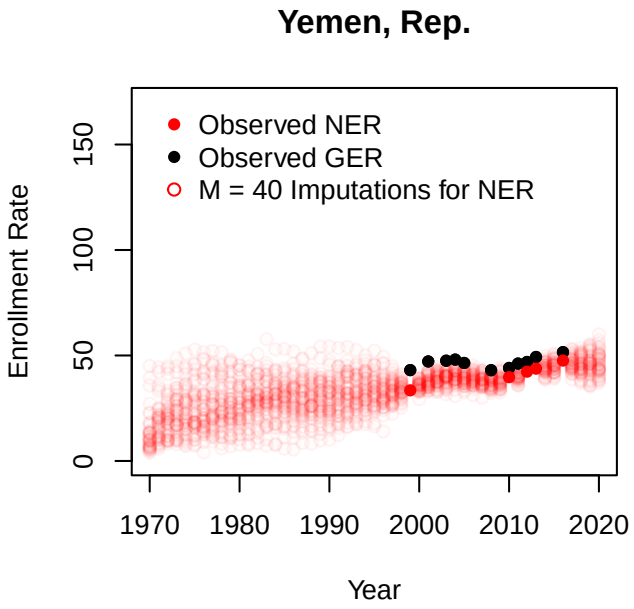

Venezuela, RB
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

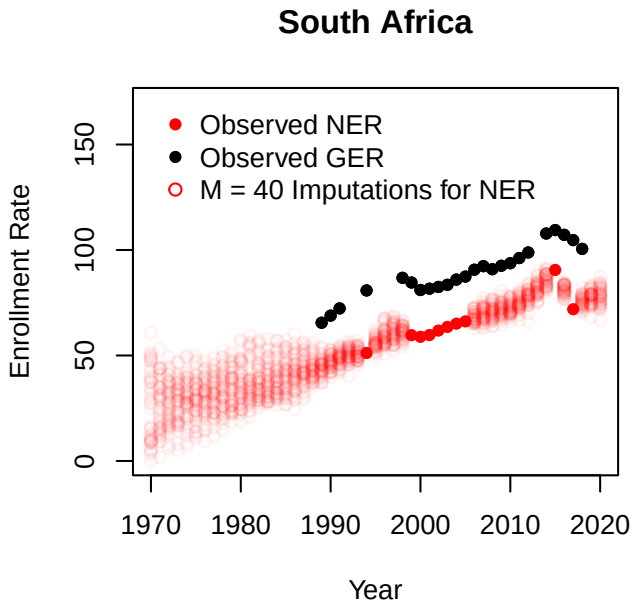

British Virgin Islands
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

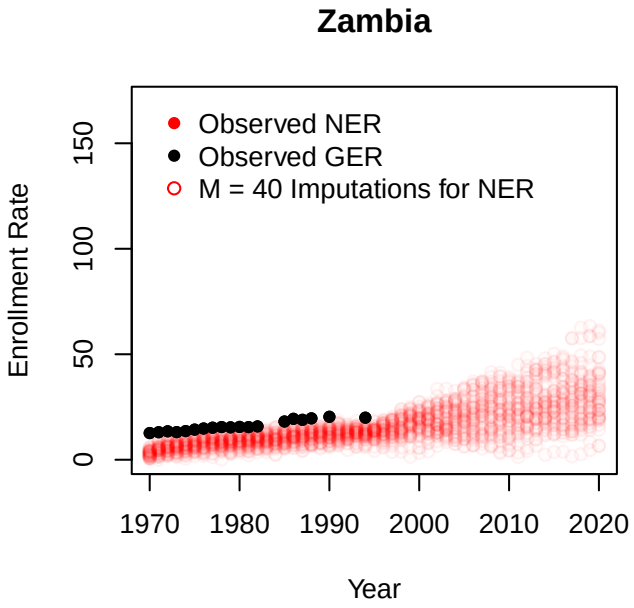

Vietnam
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year

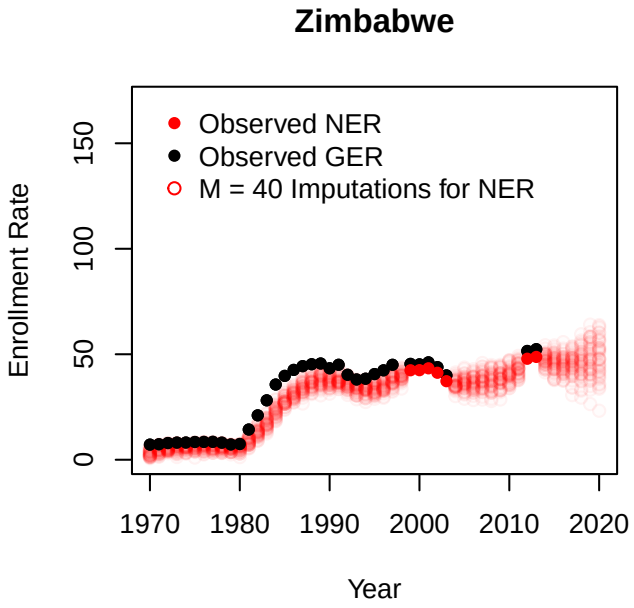

Vanuatu
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year


Samoa
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year


Yemen, Rep.
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year


South Africa
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year


Zambia
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year


Zimbabwe
Observed NER
Observed GER
M = 40 Imputations for NER
Enrollment Rate
0
50
100
150
1970
1980
1990
2000
2010
2020
Year