ECONSTOR
Make Your Publications Visible.

A Service of

ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Wazir, Muhammad Asif; Goujon, Anne

Working Paper

Assessing the 2017 Census of Pakistan Using Demographic Analysis: A Sub-National Perspective

Vienna Institute of Demography Working Papers, No. 06/2019

Provided in Cooperation with:
Vienna Institute of Demography (VID), Austrian Academy of Sciences

Suggested Citation: Wazir, Muhammad Asif; Goujon, Anne (2019) : Assessing the 2017 Census of Pakistan Using Demographic Analysis: A Sub-National Perspective, Vienna Institute of Demography Working Papers, No. 06/2019, Austrian Academy of Sciences (ÖAW), Vienna Institute of Demography (VID), Vienna,
https://doi.org/10.1553/0x003cb42c

This Version is available at:
https://hdl.handle.net/10419/207062

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ÖAW

AUSTRIAN
ACADEMY OF
SCIENCES

VIENNA INSTITUTE OF DEMOGRAPHY

WORKING PAPERS

06/2019

ASSESSING THE 2017 CENSUS OF PAKISTAN USING DEMOGRAPHIC ANALYSIS: A SUB-NATIONAL PERSPECTIVE

MUHAMMAD ASIF WAZIR AND ANNE GOUJON

Vienna Institute of Demography
Austrian Academy of Sciences
Welthandelsplatz 2, Level 2 | 1020 Wien, Österreich
vid@oeaw.ac.at | www.oeaw.ac.at/vid

Vienna
Institute of
Demography

Abstract

In 2017, Pakistan implemented a long-awaited population census since the last one conducted in 1998. However, several experts are contesting the validity of the census data at the sub-national level, in the absence of a post-enumeration survey. We propose in this paper to use demographic analysis to assess the quality of the 2017 census at the sub-national level, using the 1998 census data and all available intercensal surveys. Applying the cohort-component method of population projection, we subject each six first-level subnational entities for which data are available to estimates regarding the level of fertility, mortality, international, and internal migration. We arrive at similar results as the census at the national level: an estimated 212.4 million compared to 207.7 million counted (2.3% difference). However, we found more variations at the sub-national level.

Keywords

Census, population projections, reconstruction, Pakistan, Pakistan provinces.

Authors

Muhammad Asif Wazir (corresponding author), United Nations Population Fund, Islamabad, Pakistan. Email: wazir@unfpa.org

Anne Goujon, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography, Austrian Academy of Sciences and World Population Program, International Institute for Applied Systems Analysis.
Email: anne.goujon@oeaw.ac.at

Acknowledgments

This study is based on the publically available data and was not funded. The corresponding author had full access to all data in the study and had the final responsibility. We would like to thank Guy Abel, for his support in analyzing the international migration pattern in Pakistan. We are also grateful to members of the technical group (arranged by UNFPA) and participants in the 19th Annual Population Association of Pakistan conference for their suggestions and feedbacks. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the United Nations Population Fund (UNFPA).

Assessing the 2017 Census of Pakistan Using Demographic Analysis: A Sub-National Perspective

Muhammad Asif Wazir, Anne Goujon

1. Introduction

In 2017, from March to May, the Pakistan Bureau of Statistics (PBS) held the first census of Pakistan in the 21st century, 19 years after the last one conducted in 1998. Census enumeration in Pakistan is more politically motivated than in other countries since it provides the basis to revise the political demarcations, the allocation of national and provincial assembly seats among provinces and regions, and the distribution of inter-provincial resources through the National Finance Commission (NFC) award[1]. Actually, perennial contentions of political parties and provincial governments were one of the main obstacles to census taking in the country. Despite its successful completion, the provisional results of the 2017 census (Pakistan Bureau of Statistics (PBS), 2017) have raised many controversies and concerns, particularly vis-à-vis the results at the sub-national level, mostly for Sindh, Khyber Pakhtunkhwa (KP), and Balochistan[2].

These concerns are aggravated as PBS did not implement a Post-Enumeration Survey (PES), which is an essential component to assess the potential coverage and content errors in the census enumerations. Soon after the census results were published, the majority of the political parties demanded a validation exercise. A year later, this political conflict ended in an agreement that PBS would conduct a validation exercise using five percent of the census blocks of the country. Since then PBS has been unable to conduct the validation exercise of the 2017 census. Performing this validation with a delay of almost two years after the census enumeration would certainly jeopardize the results for two main reasons: 1) PES and/or validation exercise should be carried out within two months after the census (United Nations, 2008) because of administrative and operational issues, and 2) the rapid demographic changes that could have occurred in a year —more importantly the fertility and internal migration—would have influenced the total population size of the provinces.

In the absence of a post-enumeration survey that could either confirm or refute the results, the most sensible alternative is to reconstruct the components of population change

---

[1] The National Finance Commission Award is a series of planned economic program enacted since 1951 to correct financial imbalances between provinces.

[2] The provincial government of the province of Sindh and other political parties in Senate (upper house) have opposed the census results, see https://www.dawn.com/news/1358604 (accessed on 16/11/2018). A former member of governing council of Pakistan Bureau of Statistics has also questioned the census results, see https://www.dawn.com/news/1364371/missing-people-in-census (accessed on 16/11/2018).

(fertility, mortality, and migration) using the several datasets that are available to arrive at an estimate of the Pakistani population for 2017. This reconstructed population could then be compared to the census results and checked for plausibility (see Spoorenberg (2013) in the case of Myanmar and Spoorenberg (2012) for North Korea). In the USA and in Latin American countries, demographic analysis is a standard tool to assess the census coverage at country level (Robinson 2010; Borges & Sacco 2016). Moreover, researchers recommend implementing the analysis at the sub-national level in order to account for demographic heterogeneity (see KC et al. (2018) in the case of India). Such a reconstruction exercise was already performed for Pakistan using several data sources, for the 1998-2013 period (Wazir 2012, Sathar et al. 2014)[3]. This paper proposes to update the earlier work to estimate the population of 2017, using several methodological advancements particularly for mortality and migration, and new data sources that have become available in the meantime.

Using all sample surveys since the 1990s including the population census of 1998, this article undertakes a critical investigation of intercensal demographic changes in Pakistan at sub-national level using retrospective cohort-component population projection method. We consider four provinces (Punjab, Sindh, KP, and Balochistan) and two regions (Islamabad Capital Territory (ICT) and Federally Administrative Tribal Areas (FATA) (see also Annex)[4].

More specifically, the objectives of the study are 1) To conduct a critical investigation of intercensal demographic changes at the sub-national level in Pakistan using all available datasets and 2) by comparison, to assess the validity of the provisional results (as of January 2019) of the 2017 census of Pakistan for provinces and regions.

2. Why are the Much-Awaited Census Results Contested?

There are three distinct issues that could impact the quality of the census data. First, it is a widely accepted practice that the mechanisms to assess the quality of census data should be part of the statistical operations, i.e. pilot census and PES. The case of Pakistan was an exception with neither a pilot census nor a PES, mostly because of time constraints and security issues. Without a validation mechanism, it is challenging to evaluate the content and coverage errors of the recent census data. It also generates several doubts on the census data quality.

Second, there was a confusion about whether the census was conducted on either a *de facto* or *de jure* basis. While the census of Pakistan planned to use the *de-jure* population (recording usual residents, whether or not they are present at the time of the enumeration),

---

[3] These works also entailed population projections up to 2050.

[4] The two regions of Azad Jammu & Kashmir (AJK) and Gilgit Baltistan (GB) were not considered as they lack most demographic data.

the standard questionnaire to collect the *de-jure* population was not followed[5]. The questionnaire does not allow knowing the duration of stay in the place where the person lives permanently or his/her intention to stay there for some time. Without explicit tools, it is difficult to assess the duration and intention of stay, particularly if the person has moved frequently.

Third, the length of the enumeration period, which should be kept short in order to minimize double counting and/or omissions, which can occur due to frequent movements of the population was quite long—more than two months. This might have induced coverage and comparability issues.

3. The 2017 Provisional Census Figures

The provisional results of the 2017 census revealed a population of 207.8 million inhabitants in Pakistan[6], increasing from 132.4 million in 1998—about a 4-million annual increase during the intercensal period (Table 1).

According to the census, the population of Punjab has increased from 73.6 million in 1998 to 110 million in 2017 (an increase of 1.9 million annually). With a 2.1 percent annual intercensal growth rate, the population of Punjab has the lowest growth rate among all provinces, regions and below the national average. Sindh remains the second most populated province with 47.9 million inhabitants in 2017 (from 30.4 in 1998) and 2.4 percent intercensal annual growth rate (or 0.9 million annually). Population growth in KP was substantial: the population grew from 17.7 million to 30.5 million with an annual growth rate of 2.9 percent between 1998 and 2017. Similarly, the population of Balochistan has increased significantly from 6.6 million in 1998 to 12.3 million in 2017, with a growth rate of 3.4 percent— far above the national average. FATA population has increased by 1.8 million from 1998 to 2017. The annual growth rate in FATA stood at 2.4 percent in the intercensal period.

Islamabad Capital Territory (ICT), the federal capital of Pakistan, has recorded the highest population growth in the country with 4.8 percent intercensal annual growth rate. The population of ICT has grown sustainably from 0.8 million to 2 million from 1998 to 2017.

The census has also been implemented in Azad Jammu and Kashmir (AJK) and Gilgit-Baltistan (GB), self-governing territories under Pakistan control. Population growth in AJK was marginal: the population grew from 3 million to 4 million with an annual intercensal

---

[5] The census questionnaire mentioned the following: "Please do not include members of the household, who are absent for education, employment etc". The duration of absence was also not mentioned.

[6] Including an estimated population of 10,418 transgender people.

growth rate of 1.6 percent during 1998-2017. The population of GB was 0.9 million in 1998. However, results of the 2017 census for GB have not yet been published.

Table 1: Demographic profile for provinces and regions in Pakistan, 1998 and 2017

	Total Population (in million)								Annual increase (in million)	Annual growth rate (in %)
Province/Region	1998				2017					
	Female	Male	Total	Sex ratio	Female	Male	Total	Sex ratio	1998-2017	1998-2017
Punjab	35.5	38.1	73.6	107.2	54.0	56.0	110.0	103.5	1.9	2.1
Sindh	14.3	16.1	30.4	112.2	23.0	24.9	47.9	108.6	0.9	2.4
Khyber Pakhtunkhwa (KP)	8.7	9.1	17.7	105.0	15.1	15.5	30.5	102.7	0.7	2.9
Balochistan	3.1	3.5	6.6	114.6	5.9	6.5	12.3	110.6	0.3	3.3
FATA	1.5	1.7	3.2	108.4	2.4	2.6	5.0	104.5	0.1	2.4
ICT	0.4	0.4	0.8	117.0	1.0	1.1	2.0	111.0	0.1	4.8
Total	**63.5**	**68.9**	**132.4**	**108.5**	**101.3**	**106.4**	**207.8**	**105.1**	**4.0**	**2.4**
AJK	1.5	1.5	3.0	101.0	2.0	2.0	4.0	101.0	0.1	1.6
GB	0.45	0.42	0.9	105.0	-	-	-	-	-	-
Pakistan	**65.4**	**70.8**	**136.2**	**108.2**			**211.8**		**4.0**	**2.3**

Source: Pakistan Bureau of Statistics (PBS), Planning & Development Department AJK and GB.

4. Reconstruction: Data Sources and Methodology

This study uses a wide range of sample surveys and censuses conducted since the 1990s. In total, two population censuses (1998 and 2017) and 26 national and provincial/regional representative sample surveys were retained for the analysis (see table 2). We covered six first-level subnational (provinces and regions) entities in our analysis that were presented in Table 1[7] except for AJK and GB. Our approach is bottom-up, conducting retrospective population projections for all sub-national entities separately, and then aggregating them to obtain the population for the whole country.

---

[7] There are three level administrative units in each province, AJK and GB: 1) provincial, 2) divisional and 3) district level administration. Whereas FATA have different administrative setup, including seven agencies and six frontier regions. This analysis is purely focused on the first level administrative units i.e. provinces.

The retrospective projections at sub-national level follow the cohort-component method of population projection (George et al. 2004). The five-year age-sex cohorts are projected starting with $t = 1998$ as the jump-off year, and are subjected to changes regarding their fertility (women), mortality, international and internal migration. The estimates for the population reconstruction are calculated in five-year periods starting from 1998-2002 to 2013-2017. The single year estimates of fertility and mortality are then converted to five-year averages. In the case of international migration, we study immigration and emigration flows. For internal migration, we estimate a complete origin-destination migration matrix (8x8) between the provinces and regions for every five-year period.

To compare the reconstructed population estimates based on the initial population at the time of the census in March 1998 with the 2017 census results, we add one step in the projections to 2018-2022 and compute the age-sex population as of 15 April 2017 using linear interpolation between the two population age distributions[8] (2013-17 and 2018-22). All statistical and demographic analyses are applied in R-statistical programing language and STATA v13 (R Development Core Team, 2018, StataCorp, 2013).

---

[8] We used the "AGEINT" spreadsheet of Population Analysis System (PAS) Software developed by the US census bureau. An R code is developed using the Excel spreadsheet.

Table 2: Data sources for the retrospective projections

Surveys/Census	Rounds	Direct estimates (Full birth histories)		Indirect estimates (Summary birth histories)		International migration	Internal migration	Source
		Childhood Mortality	Fertility	Childhood Mortality	Fertility			
Census	1998 & 2017			x	x			PBS
Pakistan Demographic And Health Survey (PDHS)	1990-91 2006-07 2012-13	X	x					National Institute of Population Studies (NIPS) and ICF
FATA Development Indicators Household Survey	2013-14			x	x			Bureau of Statistics (BoS) FATA
Pakistan Reproductive Health & Family Planning Survey (PRHFPS)	2000-01			x	x			NIPS
Pakistan Fertility and Family Planning Survey (PFFPS)	1996-97			x	x			
Status of women, reproductive health and family planning survey (SWRHFPS)	2003			x	x			
Pakistan Contraceptive Prevalence Survey (PCPS)	1994-95			x	x			
Pakistan Social & Living Standard Measurement Survey (PSLMs)	2005-06 2007-08 2013-14	X	x					PBS
Multiple Cluster Indicator Survey (MICS)	Punjab: 2014, 2011, 2007-08, 2003-04			x	x			BoS, Punjab
	Sindh: 2014, 2003-04			x	x			BoS, Sindh
	Balochistan: 2010, 2004			x	x			BoS, Balochistan
	KPK: 2008			x	x			BoS, KPK

Surveys/Census	**Rounds**	**Direct estimates (Full birth histories)**		**Indirect estimates (Summary birth histories)**		**International migration**	**Internal migration**	**Source**
		Childhood Mortality	**Fertility**	**Childhood Mortality**	**Fertility**			
	FATA: 2007			x	x			BoS, FATA
Labor Force Survey (LFS)	1999-2000 2005-06 2010-11 2012-13 2014-15						x	PBS
UN WPP 2017						x		United Nations Population Division
Database of Guy Abel (2017)						x		(Abel, 2017)
Data Bureau of Emigration & Overseas Employment (BEOE)						x		Govt. of Pakistan

4.1. Smoothing Method: Loess Based Approach for the Estimation of Mortality and Fertility Indicators

We chose the approach to fit loess regression curves to the data using a variety of smoothing parameters to vary the sensitivity to recent data trends.

The basic loess function is:

$\log(M_{it}) = \beta_0 + \beta_1 x_i + \beta_2 z + \varepsilon$…………..(1)

Where M is either infant, under-five mortality or total fertility rate, where $i \epsilon \{province\ and\ region\}$ and $t \sim time$, x is the calendar year and z is an indicator variable taking value 1 if the observed values come from a direct method and value 0 otherwise, and ε is an error term which is assumed to be normally distributed. β_1 is an estimate of the expected change in the given indicator for any single year, and β_2 is a parameter that specifies the completeness of the data for given indicators. We combine all the observation from direct and indirect sources, the loess function is fitted for each province and region separately using weighted least squares regression, with the weights corresponding to each observed values.

Let x_0 be the time point at which a fitted loess curve is required and define $\psi = \|x - x_0\|$ to be the separation between time points $x\ and\ x_0$ and Φ is the maximum value for the separation. The weight function is given by

$\omega = (1 - (\Psi/_{(\Phi)})^3)^3$……….(2)

A weighted ordinary least-squares regression is then applied on combined datasets of given indicators (TFR, U5M or IMR) to create a fitted value at x_0. This process is repeated for every subnational level to generate an entire time trend.

This weighting function is tuned by a single parameter α. This tuning takes two different forms, depending on whether or not α is less than one. When α is less than one, it controls the share of the dataset that is used to estimate the regression coefficient at every point in time. For example, an α of 0.2 means that a point is given zero weight if more than 20% of observation are closer to x_0 (i.e. with smallest values of ω). For $\alpha \geq 1$, all data are included and the weighting function becomes:

$\omega = (1 - (\Psi/_{(\Phi\ \alpha^{\frac{1}{2}})})^3)^3$……….(3)

For a minimum value of α, we ensure that at least 3 data points are always included in the loess regression (as required for the estimation of the variance-covariance matrix associated with the regression coefficients) and examine the fitted loess curve corresponding to this α value. If this α value does not provide a sufficiently smooth fit to the data then we increase α until a sufficiently smooth fit is achieved. The resulting α value is used as α_{min}. For the maximum value of α, we used α_{max}=2.

For each value of α in $\{\alpha_{min}, \alpha_{min} + 0.05, \alpha_{min} + 0.1, \dots 1.0, 1.1, \dots \alpha_{max}\}$, we calculate the weights for each observed indicators (infant and under five mortality rates, and TFR) using eq. (3) and fit the loess function of eq. (1) using weighted least square regression. A 1000 random draws were computed from multivariate normal distribution by the estimated regression coefficient and their variance-covariance matrix. For each of the 1000 random draws, an estimate of given indicators for single year from 1980-2017 is predicted. Finally, we pooled the estimates of given indicators per each α values across the set of α values, a median as well as uncertainty intervals (2.5th and 97.5th centile) were computed. Further details about the Loess methods can be found in Murray et al. (2007).

4.2. Base Year Population by Age and Sex

The post-enumeration survey (PES) of the 1998 census at sub-national level was not available, however demographic analysis at the national level revealed that the last census had most likely suffered from a 2.04 percent undercount (Kemal et al. 2003). Nevertheless, in the absence of any indication of under- or over-count at the sub-national level, we further assume that the coverage of the 1998 census at sub-national level was acceptable. This study uses the micro-level data of 1998 national and population housing census retrieved from IPUMS international data center (Minnesota Population Center, 2017). We modify the data in two ways. First, we correct age heaping by applying cubic splines to smooth the age distribution of each province and region. Second, we adjust the total population from the IPUMS sample to reflect the official figure of the 1998 census (see Table 3) from which it deviates. The spline-smoothed age-sex distribution of each subnational entity is proportionally adjusted based on official census population and used as a baseline for the projections.

For FATA, the age-sex distribution of the population is obtained using the KP age-sex population patterns because of the similarities in the demographic and socio-economic development of both regions.

Table 3: Difference in total population between the IPUMS sample and official 1998 census estimates (provinces and regions)

Provinces/Regions	Total Population (in millions) from IPUMS sample of 1998 census			Difference in the total population (in millions) [Official- IPUMS]		
	Male	Female	Total	Male	Female	Total
Punjab	38.4	35.9	74.3	-0.3	-0.4	-0.7
Sindh	16.3	14.5	30.8	-0.2	-0.1	-0.3
KP	9.4	8.9	18.4	-0.3	-0.3	-0.6
Balochistan	3.6	3.1	6.7	-0.1	-0.1	-0.1
FATA			Not included	-	-	-
ICT	0.5	0.4	0.8	-0.02	-0.02	-0.04
Total	**68.9**	**63.5**	**132.4**	**-1.0**	**-0.9**	**-1.8**

Source: IPUMS International database, data retrieved on January 31, 2018; Pakistan Bureau of Statistics

Note: FATA population is not included in the IPUMS sample.

4.3. The Mortality Transition

There is no comprehensive assessment of the mortality transition in Pakistan, primarily because of the lack of consistent and reliable statistics. Since independence from the British colonial rule, Pakistan has conducted six population censuses: 1951, 1961, 1972, 1981, 1998, and 2017. None of them collected information on the deaths that occurred within the household. Between 1960 and 1980, the Bureau of Statistics had implemented several household sample surveys and recorded the number of deaths to assess mortality levels. However, these surveys have suffered from serious age-heaping and under-enumeration issues and consequently, the mortality rates were unreliable (Irfan 1986, p. 32). In the early 1980s, PBS started participating in the collection of vital statistics through a new type of sample surveys i.e. Pakistan Demographic and Health Surveys (PDHS). These surveys were also challenged because of coverage and reliability issues (Feeney & Alam 2003). However, the authors estimated that the onset of the mortality transition happened in Pakistan in the 1960s — with a substantial increase of the life expectancy from 1963 to 1973, by 6.7 years for women and 7.5 years for men (ibid.: 79-81).

It is uncertain when the mortality transition began in the provinces and at what stage they are currently. The data are not available to come up with a reliable estimate. For instance, in Punjab, the available data seem to point out at an increase in life expectancy at birth from 34.4 years to 57.5 years for women and from 32.9 years to 56.4 years for men between 1952 to 1977 (Rukanuddin & Farooqui 1987). Such an enormous increase in life expectancy within a short period (about 25 years) especially at the onset of the mortality transition – about 0.9-year increase/year for both men and women—seems unrealistic keeping in view the low level of socio-economic, health status and overall demographic situation of the province. In developing countries during the 2nd half of 20th century, life expectancy at birth, on average, rose by less than 0.4 years annually in countries where life expectancy at birth was below 55 (National Research Council 2000, p.126). Therefore, it is highly likely that these estimates for Punjab were over-estimating life expectancy at birth.

4.3.1. Estimation of Sex-Specific Life Expectancy at Birth and Survival Ratios

The process of estimating life expectancy at birth for men and women for each subnational entity is implemented in three steps. First, we estimate infant (${}_1q_0$) and child (${}_5q_0$) mortality rates for both sexes using robust statistical models applied to all available relevant surveys. Second, the child mortality estimates are disaggregated by sex. Finally, we combine those estimates with model life tables to estimate the life expectancy at birth and survival ratios by age. The details of each step are described sequentially below.

4.3.2. Estimates of Infant and Child Mortality Rates

To derive a consistent time series of infant and under-five mortality rates, ${}_1q_0$ and ${}_5q_0$, we use both full birth histories (FBH) and summary birth histories (SBH) data from various national and provincial representative surveys in Pakistan. Estimates of childhood mortality were computed using FBH from three rounds of PDHS for all married women aged 15-49 years based on calendar-years rather than cumulative 5-years preceding the survey[9]. Use of calendar-year is recommended because it brings out variation in child mortality in a clearer and more realistic manner (Pedersen & Liu 2012). In the case of FBH data, only direct estimates are included in the analysis because indirect estimates systematically produce higher estimates than the corresponding direct ones (Silva 2012). Even though PSLMs provide full birth histories data, estimates of childhood mortality from this source were quite low. PSLMs may suffer from several quality issues that were not systematically assessed. Consequently, estimates from PSLMs are based on the four years preceding the survey. Whereas, indirect estimates of childhood mortality are derived from SBH (total number of children ever born and the total number of children surviving to women of reproductive age) data from MICS, 1998 Census, and other nationally representative surveys. The method used to derive indirect estimates from SBH (also known as the Brass method) is described in detail in the United Nations manual[10].

Following the estimation strategy of the UN Inter-agency Group for child mortality (UN IGME 2015), infant mortality rates and under-five mortality rates for both sexes are estimated using all nationally and provincially representative sample surveys and census since the 1990s, adjusting these data for biases if needed. We then apply Loess (locally

---

[9] For few subnational entities, e.g. Balochistan, the sample size is too small. In that case, we truncated the estimates (based on the DHS standard methodology to calculate IMR and U5MR). Thus, in some regions the retrospective estimates are for 3 or 4 years preceding the surveys.

[10] We used 'CEBCS' function of Mortpak– indirect estimation of infant and child mortality from data on children ever born and children surviving.

weighted least squares) regression to derive smoothed single-year estimates of ${}_1q_0$ and ${}_5q_0$ for both sexes at the level of provinces/regions (Hill et al. 2012, Murray et al. 2007). Because of the limited availability of sample surveys in FATA, the regression did not produce robust estimates of ${}_1q_0$ and ${}_5q_0$. Consequently, for this region, we fit a logistic function on the few available observations of ${}_1q_0$ and ${}_5q_0$ (Arriaga et al. 1994, p.351)[11]. The estimates of ${}_1q_0$ and ${}_5q_0$ for four provinces are given in Figure 1. They present a fairly consistent picture of mortality trends and levels showing that the child mortality transition is underway (also in other regions -- data not shown here). Our estimates reveal that, in 2017, the highest under-five mortality rates were found in Punjab followed by Balochistan: 94 and 92.6 deaths per thousand live births, respectively. In Punjab, ${}_5q_0$ fell from a level of 169 deaths per thousand in 1980, whereas it was as high as 191 deaths per thousand live births in Balochistan. In Sindh, ${}_5q_0$ declined from 162 to 84 deaths per thousand live births and in KP from 167 to 69 deaths per thousand live births in KP during 1980-2017.

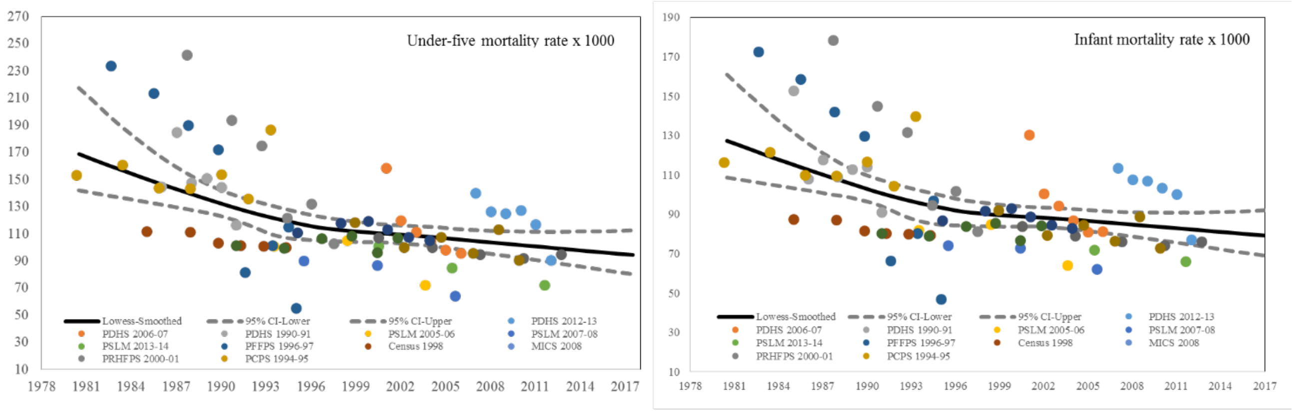
Infant mortality rates follow a similar pattern of decline over time in all four provinces as we have observed in the case of under-five mortality: ${}_1q_0$ declined from 127 to 79 deaths per thousand live-births in Punjab and from 120 to 72 deaths per thousand live births in Sindh between 1980 and 2017. The ${}_1q_0$ in Balochistan stood at the highest level among provinces and declined gradually during the four decades, from 134 to 87 deaths per thousand live births during 1980-2017. Whereas, ${}_1q_0$ in KP has declined substantially from 1980 to 2017, from 126 to 60 deaths per thousand live births.

---

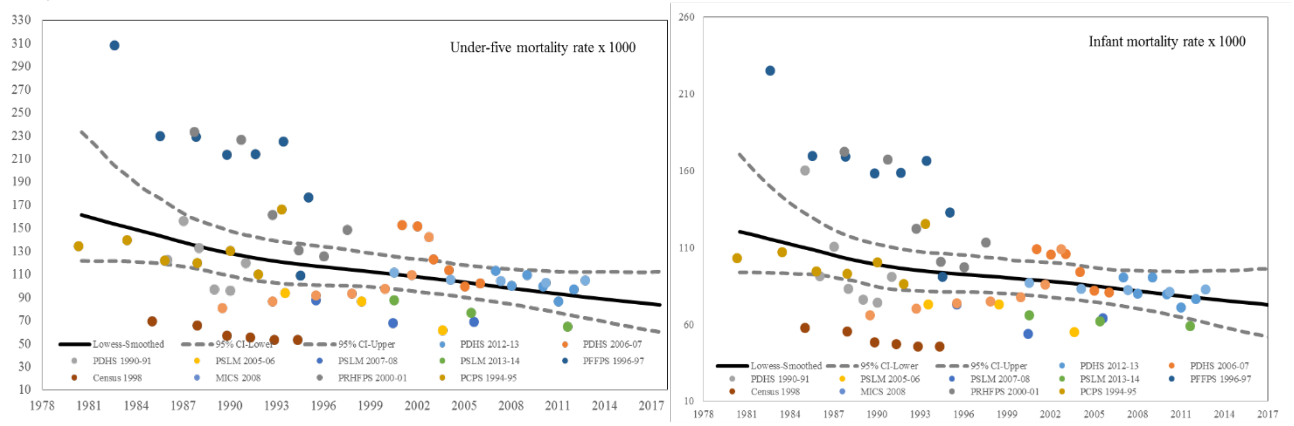
[11] There were few surveys in FATA because of administrative and security reasons. Moreover, the micro-level data for these surveys are not available to the general public. Thus, we used the 'Logistic' interpolation and extrapolation techniques (adopted from Population Analysis Spreadsheets (PAS) developed by the U.S Census Bureau). The advantage of this technique is that it requires few data points along with the values of the lower and upper asymptotes. The values of the asymptotes are grounded on expert knowledge based on the demographic and socio-economic level of the region.

Figure 1: Estimates of $_1q_0$ and $_5q_0$ in the provinces of Pakistan, 1980-2017

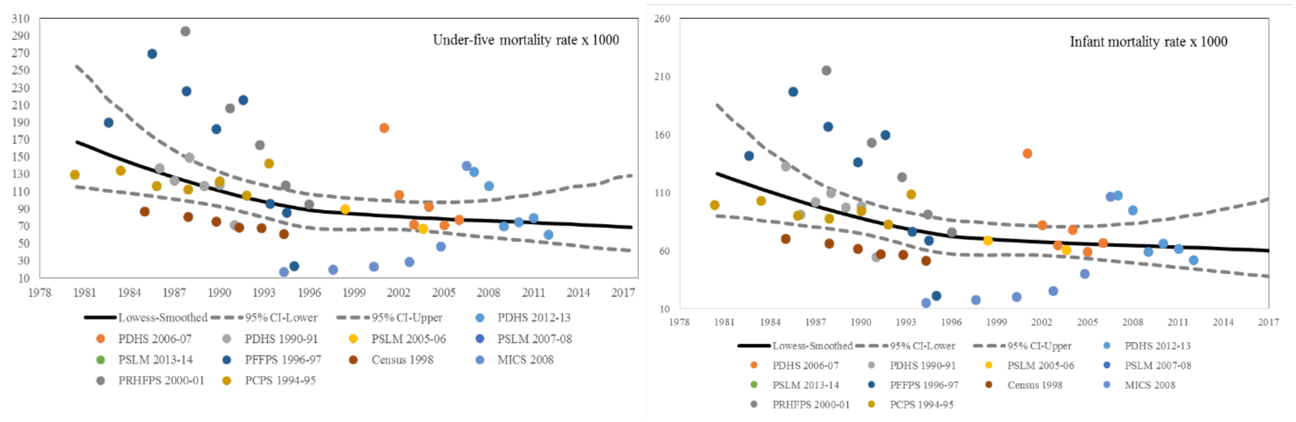
A) Punjab



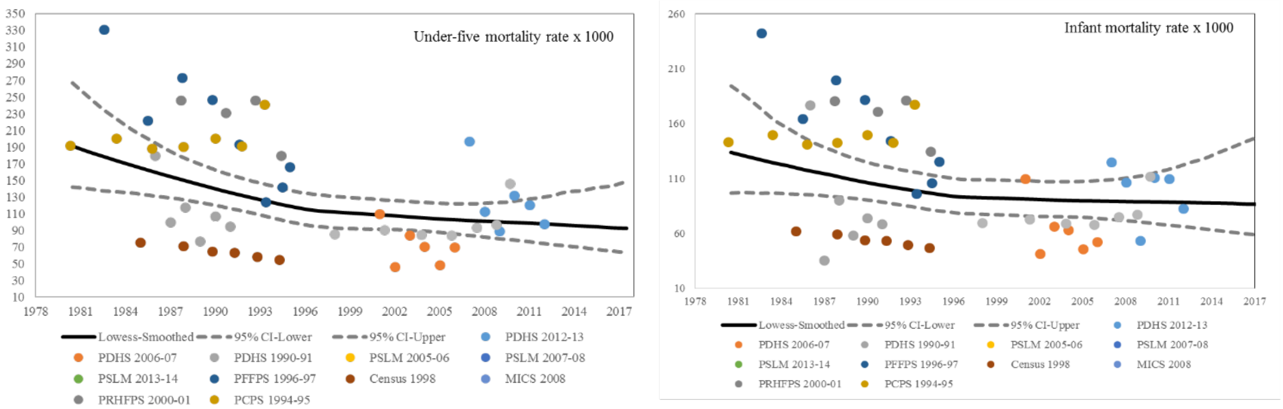
B) Sindh



C) KP



C) Balochistan



Source: Authors' calculations

4.3.3. Life Expectancy at Birth for Men and Women

We derive the disaggregated life expectancy at birth for men and women in two steps. First, we compute estimates of the sex ratio of ${}_{1}q_{0}$ (SR1), ${}_{5}q_{0}$ (SR5) and the sex ratio at birth (SRB) from selected sample surveys from 1998 to 2015, using the average of four years preceding the survey. Because of the two-level disaggregation (sub-national and then by sex), the sample of deaths occurring at ages 1 to 4 years is too small and does not permit to produce adequate estimates of sex ratio for small provinces and regions. However, it is sufficient in the case of Punjab and Sindh.

In the second step, we used a robust linear regression (implemented with 'rlm' function in R MASS package) to derive single-year estimates of sex ratios from 1998 to 2017 for Punjab, Sindh and National level. For other provinces and regions, the average five-year rate of decline for both SR1 and SR5 at the national level is used to derive the estimates of sex ratios. Predicted levels and trends of SR1, SR5, and SRB were applied to estimates of ${}_{1}q_{0}$ and ${}_{5}q_{0}$ for both sexes to produce time series of ${}_{1}q_{0}$ and ${}_{5}q_{0}$ by sex using the formulas presented in (Sawyer 2012, p-11)[12].

The estimates of ${}_{5}q_{0}$ by sex are used to derive $45q_{15}$ for male and females separately using Coale-Demeny West model life tables. For each province and region, we computed abridged life tables by sex using estimates of infant and under-five mortality, and $45q_{15}$[13].

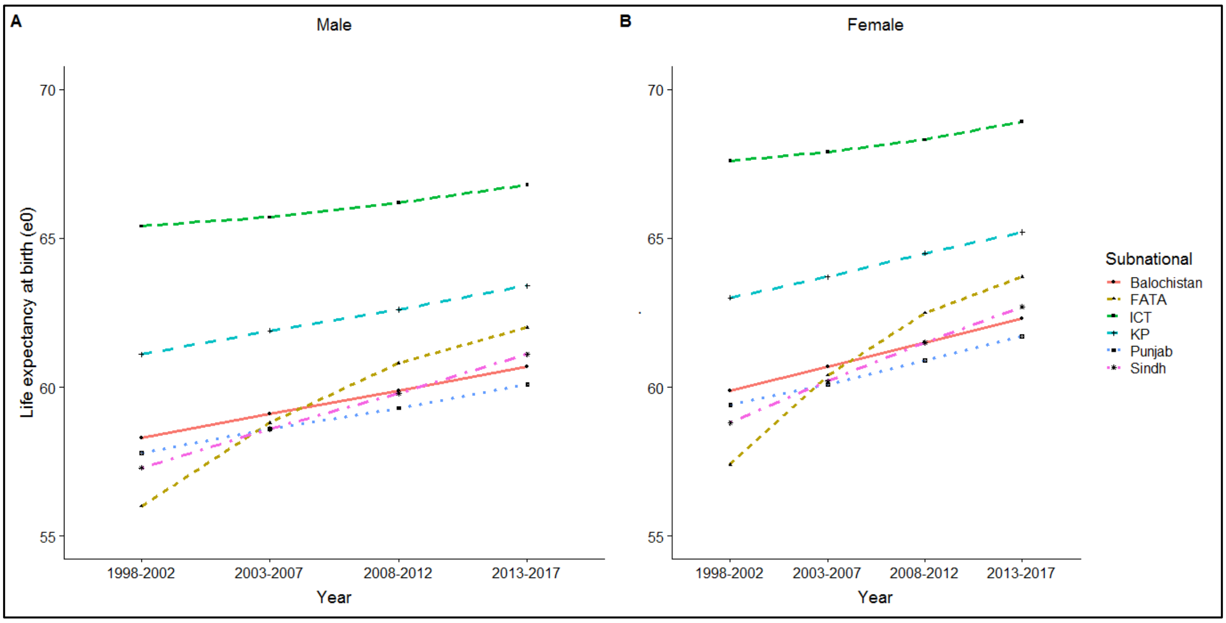
---

[12] The formulas are the following: $5q_{0male_t} = 5q_{0both_t} \times \frac{(1+SRB)}{(SRB+\frac{1}{SR5_t})}$ and $5q_{0female_t} = 5q_{0male_t}/SR5_t$.
Corresponding formulas were applied for infant mortality rates.

[13] To derive the life tables for a province/region in a given period, under-five mortality rate is provided to draw the adult mortality $45q_{15}$ using the 'CORMOR application of Mortpak with West model life tables, because under-five mortality is the preferred indicator to the choice of model life tables due to its robustness. We used estimates of infant, under-five and adult mortality as an input into the 'COMBINE' application of MORTPAK to draw life tables for each province/region in the periods between 1998 and 2017.

The last step of the estimation of e_0 by sex is to adjust the gap of female-male life expectancy at birth (Raftery et al. 2014). Using the mortality data from the Population Division DESA (2017) along with provincial/regional level data of Pakistan, a robust linear regression is implemented to derive the gap in female-male life expectancies at birth for each province and regions, highest in ICT (2.1 year) and lowest in both Punjab and Balochistan (1.6 years). Finally, age-sex specific survival ratios are estimated using the 2017 West model life tables. The final estimates of e_0 of male and female for each province and regions from 1998-2002 to 2013-2017 is presented in Figure 2. Indeed, the mortality transition in all provinces and regions is underway, as is evident from the significant increase in life expectancy at birth for male and female over the last two decades.

Figure 2: Estimated life expectancies at birth of men and women in provinces and regions of Pakistan, 1998-2002 to 2013-2017



Source: Authors' calculations

Estimates of life expectancy at birth depicted in Figure 2 reveal a high level of spatial heterogeneity in Pakistan. Men and women residing in the federal capital, ICT, had a higher life expectancy at birth in 2013-17 (68.9 years and 66.8 years, respectively) than those in other provinces/regions. Among other provinces and regions, life expectancy at birth for women was lowest in Punjab (61.6 years) and highest in KP (65.1 years). The relatively high life expectancy in KP for both men and women despite the low level of socio-economic development could be the effect of favorable environmental conditions (Mariani et al. 2009). Indeed, KP's urbanization and industrialization rates are the lowest in Pakistan, only 19% of the population lives in urban areas according to 2017 census. The gender gap in life expectancy at birth varies slightly among provinces and regions, with the highest value in

ICT (2.1 years). The average gender gap in life expectancy at birth in all provinces is clustered around 1.6 years, except KP with a gap of 1.8 years in 2013-2017.

4.4. Fertility Estimates

4.4.1. Estimating Total Fertility Rates

So far the debate in the demographic literature has addressed fertility transition in Pakistan at the national level, no study has focused on fertility changes at the sub-national level, the main reason being the lack of consistent and reliable data. We attempt here to reconstruct sub-national fertility levels using the range of representative sample surveys and censuses that are available since the 1990s. In the first step, adopting a similar estimation approach as for childhood mortality (see the previous section), we have calculated single-year, calendar based, retrospective total fertility rates (TFR) using the FBH data of all married women aged 15-49 years available from three rounds of PDHS for each province and region[14]. We also use retrospective TFRs from PSLMs, computed for three consecutive periods based on the data for the four years preceding the surveys. Indirect estimates of TFR are also derived from SBH (total number of children ever born of women of reproductive age) data from MICS, the 1998 Census, and other nationally representative surveys, using the Arriaga method (Arriaga 1983).

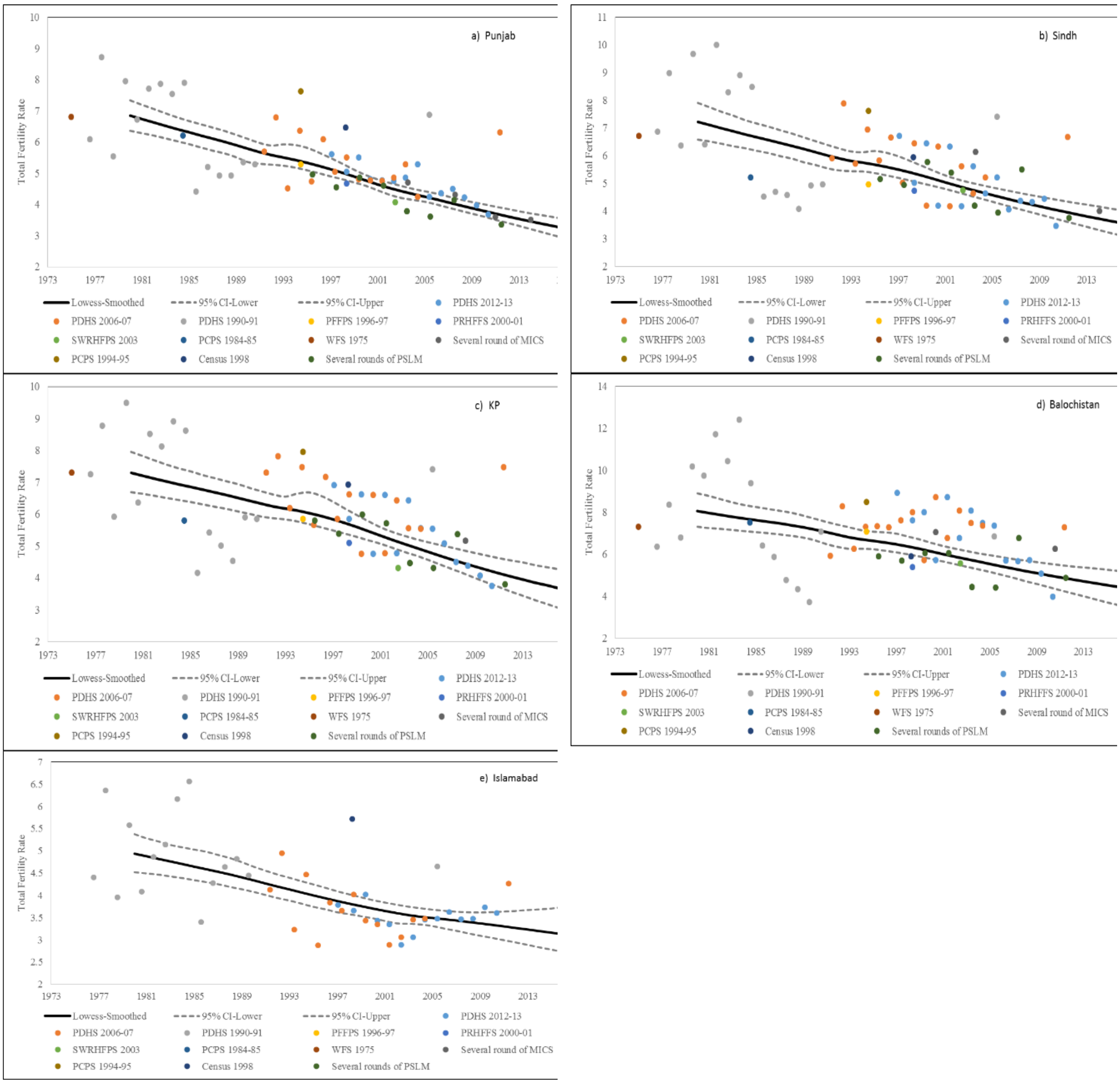
In the second step of the estimation process, province-specific 'Loess' regressions are applied to derive a consistent time series of TFR. For the less statistically developed region of FATA, we used the logistic function (childhood mortality estimation) on available observations of fertility. The total fertility rate in other provinces has also declined steadily, from 6.0 children in 1993-97 to 3.8 children per women during 2013-17 in KP. The fertility in Balochistan stood at the highest level among all provinces and declined gradually during the three decades of observation, from 6.6 to 4.5 children per woman. Fertility is the lowest in 2013-2017 in ICT: 3.2 children per woman – it was 4.0 children in 1993-1997.

Figure 3 plots the estimates of the TFR for selected provinces resulting from the application of diverse methods of fertility estimation. Similarly, to childhood mortality, the trend line gives a fairly consistent picture of fertility declines in the four provinces and ICT. Our estimates show the heterogeneity of fertility at the subnational level.

---

[14] We used tfr2, a Poisson regression-based STATA module for computing fertility rates from birth histories (Schoumaker, 2013).

Figure 3: Estimated total fertility rate (TFR) for provinces and ICT of Pakistan, 1980-2017



Source: Authors' calculations

Over the last three decades, fertility in Punjab declined by 1.9 children, from 5.3 children per woman in 1993-97 to 3.4 in 2013-17. The total fertility rate in Sindh stood at 5.7 children until 1993-97, thereafter fertility continuously decreased and reached a level of 3.7 children per woman in 2013-17. These estimates indicate that fertility in Sindh declined by 2 children per woman during the last three decades.

The total fertility rate in other provinces has also declined steadily, from 6.0 children in 1993-97 to 3.8 children per women during 2013-17 in KP. The fertility in Balochistan stood at the highest level among all provinces and declined gradually during the three decades of observation, from 6.6 to 4.5 children per woman. Fertility is the lowest in 2013-2017 in ICT: 3.2 children per woman – it was 4.0 children in 1993-1997.

4.4.2. Age-Specific Fertility Rates

Estimating age-specific fertility rates in Pakistan at the sub-national level is complicated by the lack of reliable and time-consistent rates. Furthermore, due to random fluctuations in fertility rates over time, estimates based on interpolation between surveys tend to result in irregular age patterns, particularly for the smaller sub-national entities.

We decided to apply a relational model, which relates the age profiles to be fitted or projected to a standard age schedule (Booth 1984), so-called Gompit regression model. It is defined as

$$Y(x) = -\log\{-\log(F(x))\}……….(4)$$

where $F(x)$ is the cumulative relative ASFR at age x. Let $Y_s(x)$ equal Gompit ($F_s(x)$ of the standard fertility schedule. Then a new ASFR schedule $Y(x)$ can be defined in terms of $Y_s(x)$ as

$$Y(x) = \alpha_0 + \beta_0 Y_s(x)………………(5)$$

Using the age-specific fertility patterns of each national and subnational level from all surveys from 1990 to 2012 (eight surveys where ASFR are available), we empirically estimate α_{0i} *and* β_{0i}—where $'i'$ referes to the year of survey—from Eq. (5) using standard age schedule of fertility adopted from (Booth 1984). A cumulative ASFR schedule may be recovered from $Y(x)$ by reversing the above transformations.

The next task is to estimate the unique value of α *and* β by controlling time variation. We aim to relate both parameters to the total fertility rate (TFR). Based on maximum R-square, we find power fit to be best for β and logarithmic fit for α.

$$\beta = 1.8351(TFR)^{-0.526}………..(6)$$

$$R^2 = 0.2513$$

And

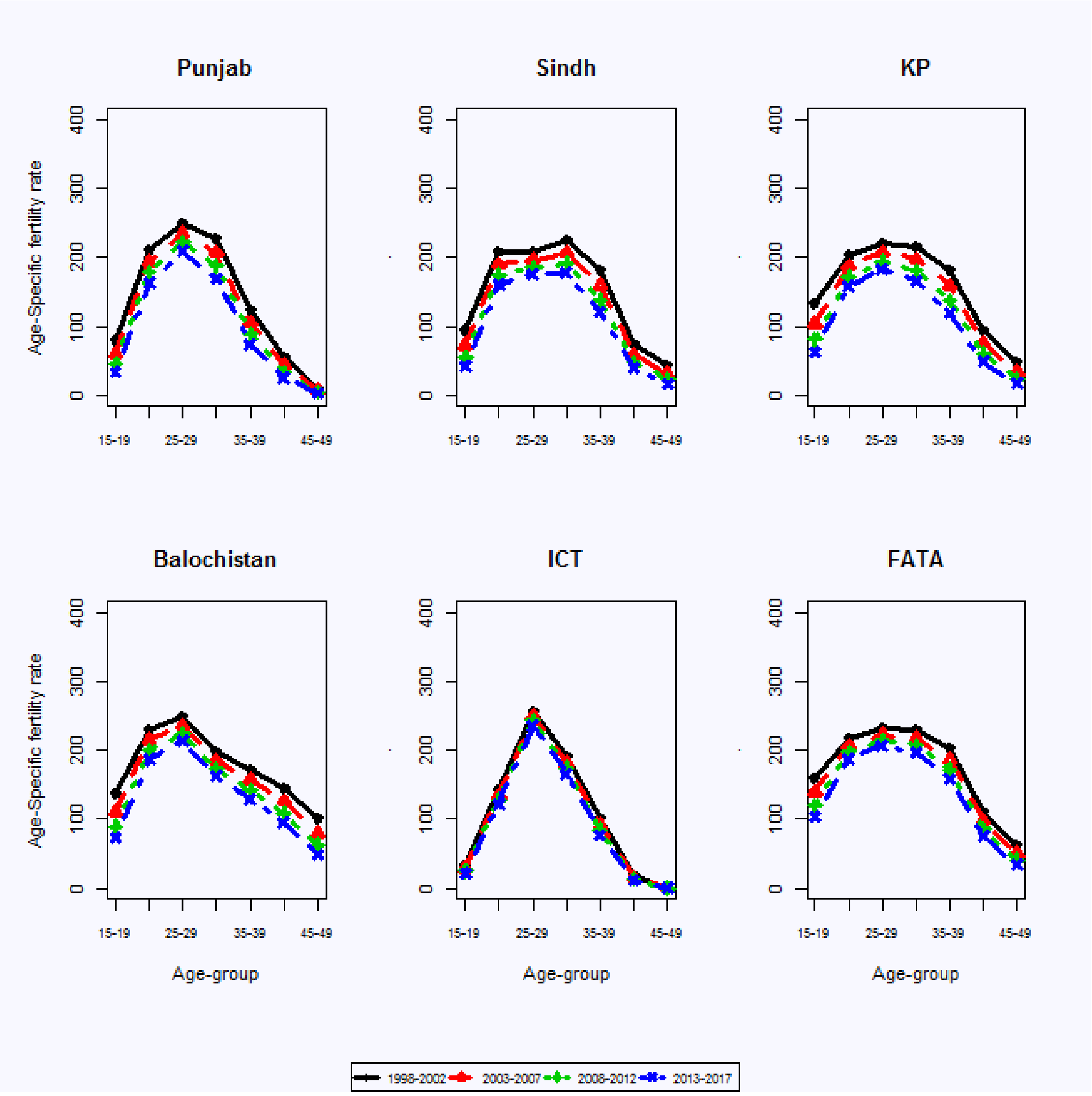
$$\alpha = -0.111\log(TFR) + 0.0382 ……….(7)$$

$$R^2 = 0.0416$$

Using these relationships, the subnational specific ASFRs can be derived for given TFRs in each period.

The age pattern of fertility in Pakistan shows the typical pattern of a country in the middle stage of the demographic transition, with childbearing peaking in the age group 25-29 and declining afterward (Wazir 2012, p-38), and still substantial early childbearing in the age group 15-19 (see Figure 4). Early marriage and childbearing are contributing to high fertility in many high fertility countries (United Nations 2013) since women tend to spend more years in the reproductive span, with increased health risks for both mothers and children, and reduced education and employment opportunities. The age pattern is slightly more concentrated in the age group 25-29 compared to 1998-2002. This phenomenon is especially visible in Sindh. However, the stagnation in the overall age-specific pattern reflects the sustained preference of Pakistani for large family size (Wazir 2018).

Figure 4: Estimated Age-specific fertility rates for four provinces and two regions of Pakistan, from 1998-2002 to 2013-2017



Source: Authors' calculations

4.5. International Migration

In-depth analysis of international migration is difficult in Pakistan because existing data are incomplete and do not effectively capture contemporary intensities and patterns. Moreover, the limited available data in Pakistan on international migration are not compatible with international definitions.[15]

Some information is available about the stock of Pakistanis residing in other countries—that shows the magnitude of out-migration from the country. According to estimates from the United Nations in 2017, approximately 5.9 million Pakistani were residing in another country in 2015 (about 2.7 million Pakistani in Gulf countries). They were 3.4 million in 2000 (Population Division, DESA 2017)[16]. Official data from the Bureau of Emigration and Overseas Employment (BEOE) of Pakistan mention about 5.8 million Pakistanis who migrated legally for overseas employment in Gulf countries during 2005-2015. Large-scale labor migration of Pakistanis to the Gulf region started with the oil boom in the late 1970s and continued in the context of the volatile political situation in Pakistan combined with issues related to poverty and unemployment. Gulf countries also have the advantage of geographical and religious proximity. However, official data do not fully capture the total overseas migration because highly skilled migrating Pakistani who are migrating, especially to non-GCC countries, usually do not register with government agencies. It was estimated that only 2 percent of Pakistani workers going abroad for employment formally went through the registration procedure with the overseas employment corporation (Ministry of Overseas Pakistanis and Human Resource Development 2016). Since the early 1990s, there has been a growing trend among well-off urban and rural families to send their male children for educational purposes, most commonly to the United States, United Kingdom, and other European countries. These men often established themselves overseas and are later joined by their dependents (Gazdar 2003).

While there is a lack of official data, the number of immigrants to Pakistan is estimated to be very low, except for the continuous influx of refugees from Afghanistan, since the early 1980s (Soviet invasion). Estimates from the 2005 Afghan population census revealed that Pakistan hosted some 3 million Afghans, the largest Afghan diaspora in the world, and the majority of this Afghan population (about 88 percent) were residing in KP and Balochistan (Ministry of States & Frontier Regions, Population Census Organization, UNHCR 2005). The 2005 census also showed that about 89 percent of the Afghans who settled in camps and 78 percent who settled in non-camps areas were not willing to return to Afghanistan, in tandem with the deterioration of living conditions in their country. Because of the language and cultural similarities, many Afghan refugees have become a

---

[15] Migration is generally measured using one of the two conceptual frameworks: as *transition* or as *movement*. Traditionally, the measurement of *transitions* involves the comparison of a person's location at two points in time within a given time period, whereas the measurement of *movements* involves the counting of each change of address that occurs within a given time period. One of the important parameters necessary to compute age-standard migration intensities is the measurement of population at risk, which is challenging in the case of Pakistan.

[16] These estimates present only the legal international migration. There is no comprehensive estimate of asylum seekers, refugees and irregular migrants from Pakistan. Estimates point at approximately 300,000 people migrating from Pakistan each year using irregular channels (Cibea, et al. 2013).

permanent resident in Pakistan. According to UN estimates, in 2012, about 2.3 million Afghanis had settled permanently in Pakistan, and 1.6 million still had refugee status[17].

Other immigrant groups in Pakistan—though smaller in number—originate from India, Bangladesh, Myanmar, Middle Eastern countries, Central Asian states, and China. The majority of immigrants registered as foreign workers or skilled technical experts under bilateral agreements. Recently, with increasing investment under the China-Pakistan Economic Corridor, labor force migration from China has increased, mainly to Gilgit Baltistan, Punjab, and Balochistan. The number of Chinese engineers working in Pakistan has increased from 3,000 in 2008 to 13,000 in 2011 (Cibea et al. 2013).

However, migrant stock data do not capture migration flows and particularly the timing and duration of migration movements. For our purposes, we would require consistent international migration flows (both in-and-out) over five-year periods from 1998 to 2017. In a series of pivotal studies, Abel developed a methodology to derive global estimates of origin-destination migration flows by using a variety of migrant stock data (Abel 2013; Abel & Sander 2014; Abel 2017). We used Abel's five-year estimates of total emigration and immigration flows from 1990 to 2010 (Abel 2017), adjusting it with the yearly registered emigrants from BEOE official data from 1990 to 2016. A smooth spline is applied to reconstruct the annual total emigration and immigration flows during the period of interest. Next, estimated annual migration flows are distributed among the provinces and regions using the five-year average proportional share calculated from BEOE official data. For gender division of migration, we used the average of South-Asian migrant flows by gender (i.e. 60 percent male and 40 percent female) in all provinces and regions (Abel 2017), which we assumed to be constant through the reconstruction period. The age pattern of international migration is obtained using the seven-parameter model of migration (Rogers & Castro 1981), adopting the parameter estimates of young standard age-profiles developed by Raymer & Rogers (2006).

4.6. Internal Migration

For the sub-national demographic analysis, internal migration plays an important role in shaping the demographic landscape, influencing the social-cultural and economic aspects of both origin and destination areas, particularly in Pakistan. However, little progress has been made to quantify the spatiotemporal dynamics of internal migration and its impact on population growth in the Pakistani context. According to stock data from the 1998 census, about 4 million people had internally migrated within and between the provinces and federal capital during the ten years preceding the census. Out of these four million, approximately 2.1 million (53.4 percent) had migrated between the four provinces and Islamabad (Karim & Nasar 2003, p. 169): about 44 percent had migrated to Punjab (0.95 million), 36% to Sindh (0.77 million). Unfortunately, the 2017 census did not capture any

---

[17] Data are available from https://esa.un.org/miggmgprofiles/indicators/files/Pakistan.pdf [accessed on 20.11.2018]

information on internal migration. Thus, it needs to be reconstructed using other indirect methods.

4.6.1. Finding Migrants in the Labor Force Surveys

Recently, labor force surveys (LFS) and other household surveys have been frequently used to estimate internal migration flows in the absence of direct data in different settings (Willekens et al., 2017; Marti & Rodenas 2007; Wisniowski 2017). LFS that has been conducted yearly since 1963 by PBS include questions about the place of birth, previous place of residence and duration of stay in the current place. While these data have not yet been used to quantify internal migration flows in Pakistan, we hereby assume that they are a suitable source with large samples and long time series.

In this section, we present the method for estimating bilateral migration flows from survey data. We used data from five rounds of LFS conducted from 1999 to 2015, which were using a multistage, stratified sampling design to collect data on the socio-demographic characteristics of the population including internal migration. These five datasets are nationally representative and were chosen to reflect the trend over the 15-year period. LFS measured migration using the transition conceptual framework, i.e. comparison of a people's location at one point in time with that at another point (Rees et al., 2000).

The United Nations introduced the threshold to distinguish long-term migration (intention to stay for at least one year) to short-term migration (visiting for less than one year). In order to increase the sample size, in the case of Pakistan, we considered short-term migration to refer to people who had been residing in a place during the last four years or less coming from another place. A limitation is of course that a person who makes multiple transitions during the intermediate period is hard to identify because the questionnaire does not record such information. Bilateral migration flows between the provinces and regions for the population of 10 years of age and over—i.e. 8x8 origin-destination migration matrix—are computed by applying sample weights from each LFS round using STATA v13 (StataCorp 2013). Because LFS are not implemented in FATA, missing values have been generated predominantly for the destination of migrants to this region. Even though the FATA region is mostly origin of internal migrants (either in the form of labor migration to more developed provinces or internally displaced persons due to unrest), some skilled labor immigration to this region has also been observed. To overcome the lack of data, we assigned, conservatively, about one percent of the total sample of migrants to migrate to FATA.[18]

---

[18] Initially, we have tried several imputing techniques of missing values of the origin-destination matrix of each LFS survey, such as 'predictive mean matching', and 'weighted predictive mean matching' and random imputations. All these function are available in R-package "mice v3.0.0". However, these methods were not producing reliable estimates of destination and origin migrants.

Moreover, the coverage for internal migrants in the LFS might low due to several issues: non-response, imperfect coverage of the migrant population in the sampling frame, undercount due to the emigration of entire households, and non-sampling variability (Wisniowski (2017)). Moreover, KP and FATA have hosted a significant number of refugees and the internally displaced person that might not be counted in the LFS (UN Office for the Coordination of Humanitarian Affairs (OCHA) 2017).

Computing origin-destination migration rates from LFS is problematic because migration rates are considered as occurrence-exposure rates that are obtained dividing the total number of migrants by the population at risk, weighted by the duration of exposure (Willekens et al. 2017), which we do not have. Instead of using migration rates, we adopted the following approach to compute migration flows. The share of migrants to the total sample size is computed as $p_m = n_m/n$, where m denotes the migrant sample size and n is the total sample size. Using the inter-census population growth rate of 1998-2017, the total population of each province and region (P_{it-1}) is exponentially interpolated for the year $t-1$, where t is the year of each LFS. Next, the total number of migrants is computed as $\mathrm{T}_t = p_m \times P_{it-1}$.

Next, using the origin-destination matrix computed above, the ratio of migrant count $\mu_{ik} = n_{ik}/n_m$ is computed, where n_{ik} indicates the marginal sum of the matrix, $k \in \{Origin, Destination\}$ and $i \in \{province\ and\ region\}$. Finally, internal migration flows are obtained by multiplying the total population of the appropriate year as $y_{ikt} = \mu_{ik} \times \mathrm{T}_t$. With known margins y_{ikt} the origin-destination migration matrix is further filled proportionally.

4.6.2. A Statistical Model of Migration Flows

Migrants usually constitute a tiny fraction of the total sample size of LFS: less than 1 percent from 1998 to 2014, except in 2014-15 round in which the migrant's sample was 3.6 percent. To surmount this shortcoming, we build a model to adjust for statistical bias, and derive data for years without LFS, using indirect variables in a Poisson regression model. To assess whether the volume of migration flows among origin and destination regions differ, we introduced several explanatory variables in the model including the population size, the area of origin and destination, along with the distance between the capital cities of provinces and regions. To capture the change in migration flows over time, we also introduced the year as an independent variable.

To summarize, y_{ijt} is the expected migration flow from origin i to destination j, P_i is the population in the region of origin i and P_j is the population in the region of origin j, d is the distance between the capital city in origin i and destination j, t is the year and Z the dummy variable for $i, j = 1,2,3 \dots ,8$.[19] ε_{ij} is the random residual. In a first step, we included in the

---

[19] In order to compute the distance, we adopt the formula presented in Cohen et al. (2008). We compute the longitude and latitude of each capital for each province and region using public sources

model the $area$ (in km²) of the origin and destination regions to estimate the coefficients using a generalized linear model (GLM). Then a stepwise regression algorithm (stepAIC of R MASS package) was applied to the starting model, in which we eliminated areas of both origin and destination and retained all other explanatory variables. The final model is presented in Eq. 1. All variables in Eq.1 are highly significant and pseudo-R^2 for the final model was 0.809.[20]

$$y_{ijt} = \exp\left(\text{Intercept} + \alpha_1 P_i + \alpha_2 P_j + \varphi d_{ij} + b_1 z_i + b_2 z_j + \delta t + \varepsilon_{ij}\right) \text{ ....(1)}$$

Bilateral Patterns

To depict the pattern of estimated bilateral migration flows over time, two circular migration plots are shown in Figure 5 for 1993-1997 and 2013-2017. Plots were created in R using *migest* package (Abel 2018). The direction of the flow is indicated by the arrowhead. The number on the outer section axis indicates the size of migration flows (in 000s).

In 1993-97, Punjab was the destination of most migrants from other parts of the country, with a net gain of about +0.19 million migrants. The net gain has increased significantly during the intercensal period and reached +0.42 million in 2013-17. Most migrants to Punjab were coming from Sindh, KP, and FATA in both periods. In 2013-17, Balochistan has also become a major sending province to Punjab. The largest out-migration flows from Punjab were to Sindh, KP, and ICT.

Changes in the size of migration flows in Sindh revealed an interesting pattern. The net gain of internal migration reversed during the intercensal period, from a tiny negative number (about 3,000) in 1993-97 to +0.34 million in 2013-17. The most visible pattern is that net gains in Sindh have surpassed that of Punjab most likely due to the improved security and economic situation in Sindh. In the last two decades, significant flows of immigration to Sindh were originating from Punjab, KP, and FATA, whereas, the destination of the Sindhi migrating population was mostly to Punjab, KP, and ICT.

KP and FATA have experienced significant out-migration flow because of the undergoing armed conflict during the period: In KP, we estimate a net loss of -40,000 in 1993-97 to a gain of +0.2 million in 2013-17. The large influx of migrants to KP in the last period is primarily due to armed conflicts in FATA and other northern areas of the country. During the last two decades, the majority of inflows to KP came from Punjab, Sindh, and FATA, whereas, the destination of KP migrant population was mostly to Punjab, Sindh, and

---

(such as NASA Long-Lat and Map Coordinates). Further, the longitude and latitude values were converted to radiant and the distance (in km) is computed for each pair of origin-destination.

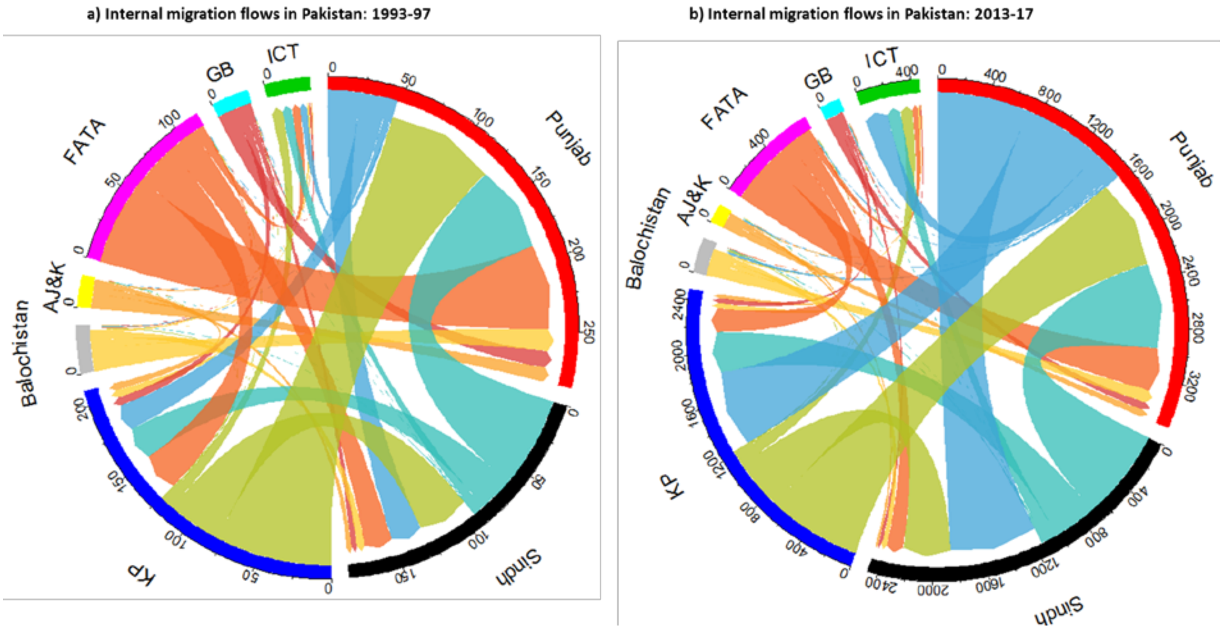
[20] We also perform an over-dispersion test of the model in Eq.1 using *dispersiontest* from "AER" R package. There is clearly evidence of over-dispersion in the model. In presence of over-dispersion, the obvious solution is to use quasi-Poisson or negative-binomial distribution. However, the quasi-Poisson regression model results in similar fitted values as the Poisson regression model and the distance variable becomes insignificant. Therefore, we prefer to use the original Poisson model.

ICT. It is important to note that this migration flows in KP and FATA do not fully capture the refugee population from Afghanistan.

The net loss of population due to migration in the FATA region has increased from -0.1 million in 1993-97 to -0.8 million in 2013-17, mostly to regions such as Punjab, followed by Sindh and KP. A small fraction of FATA population was also migrating to ICT.

Large flows, mostly labor migrants, from other provinces to ICT are also estimated, to reach 0.7 million in 2013-17. While the flows for Balochistan and other regions are contributing a tiny proportion to the total number of internal migrants.

Figure 5: Estimated bilateral internal migration flows in Pakistan, 1993-97 and 2013-17 (flows in 000s)


a) Internal migration flows in Pakistan: 1993-97

b) Internal migration flows in Pakistan: 2013-17

Source: Authors' calculations

5. Results of Retrospective Population Projections

We used the above-mentioned estimates of fertility, mortality, and migration (both international and internal) to project forward population by age and sex at sub-national level using the 1998 census as the base-year. By comparing the reconstructed 2017 population with the official results of the 2017 census population using a goodness-of-fit-statistic computed as the average absolute deviation (AAD)[21], it emerges that the

---

[21] Goodness-of-fit statistics –Average absolute deviation (AAD) = $\frac{p^{Obs}-p^{Proj}}{p^{Obs}} \times 100$

Note: Obs. denotes the official census total population size, while Proj denotes the projected population. A positive value of AAD indicates census estimates were higher compared the reconstructed total population of each province and region significantly differs compared to the official figure.

5.1. Total Population

Figure 6 illustrates the difference in population between the reconstructed 2017 population and the enumerated population by the census (A) and the AA (B). The two exercises lead to similar results at the national level. The reconstructed population (excluding AJK and GB) is 4.7 million larger than the census one, hence a marginal difference of 2.3 percent (see Figure 6).

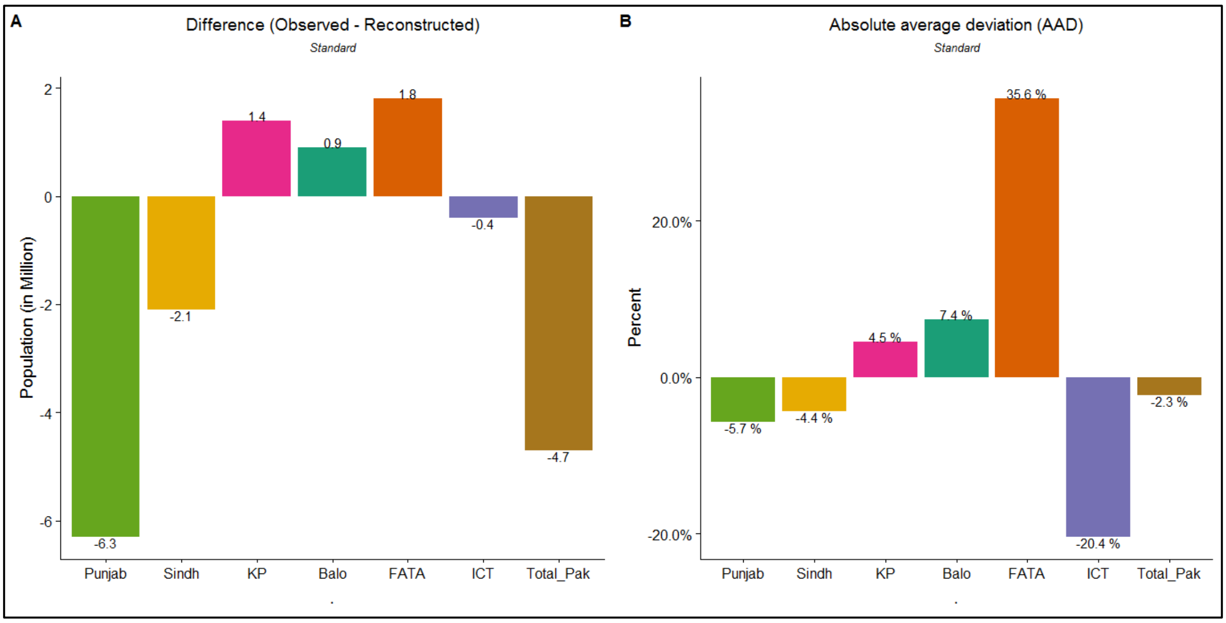
However, we find more considerable variations at the sub-national level. For the two largest provinces—Punjab and Sindh— as well as for the two smallest regions –ICT and FATA—the reconstructed population is much larger than the census population. The difference between census estimates and the reconstruction in the capital city for which coverage should be logically better could be due to flawed migration estiamtes. Another explanation could be that people want to be counted at their region of origin because of several reasons, e.g. identity, political, social, etc. The reverse is shown for KP, Balochistan, and FATA, which have a lower population in the reconstruction compared to the census. The retrospective population projections indicate that the total population of Punjab and Sindh would be 116.3 million and 50 million in 2017, a difference of respectively 6.3 (AAD: 5.7%) and 2.1 million (AAD: 4.4%) with the official census estimates. For ICT and FATA, though the difference is trivial in absolute number (0.4 million and 1.8 million, respectively), AADs are significantly higher, compared to Punjab and Sindh.

As mentioned, the retrospective population projections point at the possibility of over-enumeration of the population of KP and Balochistan, The reconstructed total population for KP and Balochistan are 29.1 million and 11.4 million in 2017, a difference of about 1.4 and 0.9 million, respectively. The values of AAD for KP and Balochistan stood at 4.5% and 7.4%, indicating a significant difference from the census figures. The case of FATA needs to be treated with special caution since most reconstruction parameters are estimated, including the base year age and sex structure of the population. While the difference between enumerated and reconstructed population is 1.8 million, due to the small denominator, the value of AAD is the highest among all provinces and regions in Pakistan, i.e. 35.6%.

---

reconstructed results —pointing at the possibility of an overestimation by the census; negative values show that the census estimates were lower compared the reconstructed results —pointing at the possibility of an underestimation by the census.

Figure 6: Difference of the 2017 census results to the retrospective population projections (A) and AAD (B) for Pakistan, by provinces and regions



Source: Authors' calculations. Note: Total difference for Pakistan excluded the AJK and GB.

To assess the sensitivity of the reconstruction to the demographic components of population growth during the intercensal period, in addition to the standard variant (that refers to the retrospective population projections above mentioned) we also ran two additional population reconstruction scenarios: 1) without internal migration, and 2) without internal and international migration – just considering natural increase. The difference between the standard and without internal migration projection variant shows the effect of internal migration. Without internal migration, the population of Punjab and Sindh would have reached 114.1 and 49.6 million in 2017, a difference of 2.2 million in Punjab and 0.4 million in Sindh from the standard variant (Table 4). These results also reveal that internal migration is responsible for most the shrinking in the FATA population (-2.4 million between the variant without internal migration and the standard variant) in 2017. The decrease in population due to internal migration in KP and Balochistan in 2017 is 0.1 and 0.5 million. Thus, the reconstruction is very sensitive to estimates of internal migration in some provinces/regions: Punjab, FATA, and ICT.

Table 4: Total population of Pakistan at the sub-national level in 2017 according to the census and to three population projection scenarios (in millions)

Provinces/Regions	Total Population in 2017			
	Census population (official)	Population reconstruction (standard variant)	Without internal migration	Natural increase
Punjab	110.0	116.3	114.1	117.9
Sindh	47.9	50.0	49.6	50.3
KP	30.4	29.1	29.2	31.3
Balochistan	12.3	11.4	11.9	12.0
FATA	5.0	3.2	5.6	6.0
ICT	2.0	2.4	1.2	1.3
Pakistan	**207.7**	**212.4**	**211.5**	**218.8**

Note: Population reconstruction refers to the retrospective population projections of 2017 and serves as the reference to assess the demographic effect of other population scenarios. Source: Authors' calculations.

The impact of international migration is shown in the difference between the natural increase variant and the without-internal migration variant. A significant number of people in Punjab and KP have migrated to other countries in 2017, about 3.8 million and 2.1 million respectively. Similarly, all other provinces and regions have minor net negative international migration (see table 4).

Table 5: Total population of Pakistan by sex at the sub-national level in 2017 according to the census and to the standard population projection scenario (in millions)

Provinces/Regions	Total Population in 2017					
	Census population (official)	Difference (Obs. – Reconst.)	AAD (%)	Census Population (Official)	Difference (Obs. – Reconst.)	AAD (%)
	Male			Female		
Punjab	56.0	-3.6	-6.5%	54.0	-2.6	-4.9%
Sindh	24.9	-0.8	-3.1%	23.0	-1.3	-5.8%
KP	15.5	0.7	4.3%	15.1	0.8	5.2%
Balochistan	6.5	0.5	7.9%	5.9	0.4	6.9%
FATA	2.6	0.8	30.9%	2.4	1.0	40.6%
ICT	1.1	-0.1	-13.3%	1.0	-0.3	-28.3%
Pakistan	**106.4**	**-2.6**	**-2.4%**	**101.3**	**-2.1**	**-2.0%**

Source: Authors' calculations

Table 5 shows the difference in population between the reconstructed 2017 population and the enumerated population from the census by sex. The results at the national level revealed that the reconstructed population for male is 2.6 million larger than the census and 2.1 million for female. Hence a marginal difference of 2.4 percent for male and 2 percent for female

However, we find more significant variations by sex at the sub-national level. In Punjab, the census population is much smaller than the reconstructed population for both men and women, 3.6 million for male (AAD: -6.5%) and 2.6 million for female (AAD: -4.9%). The retrospective population projections indicate that the difference between census and reconstructed population in Sindh is 0.8 million for male and 1.3 million for female in 2017, respectively. Similarly, the difference in KP is 0.7 million for male and 0.8 for female in 2017. For ICT and Balochistan, though the difference for both male and female population is trivial in absolute number (for male: 0.1 million and 0.5 million, and for female: 0.3 million and 0.4 million, respectively), AADs are significantly higher, compared to the other provinces.

5.2. Counterfactual Population Projections

The intercensal analysis and reconstruction show the different pace of demographic change at the sub-national level.

We used counterfactual population projections (CPP) to estimate the impact of varying the pace of change of several determinants of population dynamics on population size. Two scenarios are implemented -- rapid transition (1) and stalled transition (2) – with alternative assumptions about the four components of sub-national population change as shown below. The estimates of fertility and mortality at the sub-national level in Pakistan for 1998-2002 and 2013-17 under standard variant are presented in Table 6.

1) **Fertility:**
    a. *Rapid transition:* TFR is half a birth lower than the standard fertility estimates in each reconstruction period
    b. *Stalled transition:* TFR is half a birth higher than the standard fertility estimates in each reconstruction period.
2) **Mortality:**
    a. *Rapid transition:* Life expectancy at birth is two years higher than the standard mortality estimates for men and women in each reconstruction period
    b. *Stalled transition:* Life expectancy at birth is two years lower than the standard mortality estimates for men and women in each reconstruction period.
3) **International migration:**
    a. *Rapid transition:* assumes 25% higher immigration & emigration number from standard international migration estimates for men and women throughout the reconstruction period.
    b. *Stalled transition:* assumes 25% lower immigration & emigration number from standard international migration estimates for men and women throughout the reconstruction period.
4) **Internal migration:**

a. *Rapid transition:* assumes 25% higher in-migration & out-migration number from standard internal migration estimates for men and women throughout the reconstruction period.

b. *Stalled transition:* assumes 25% lower in-migration & out-migration number from standard internal migration estimates for men and women throughout the reconstruction period.

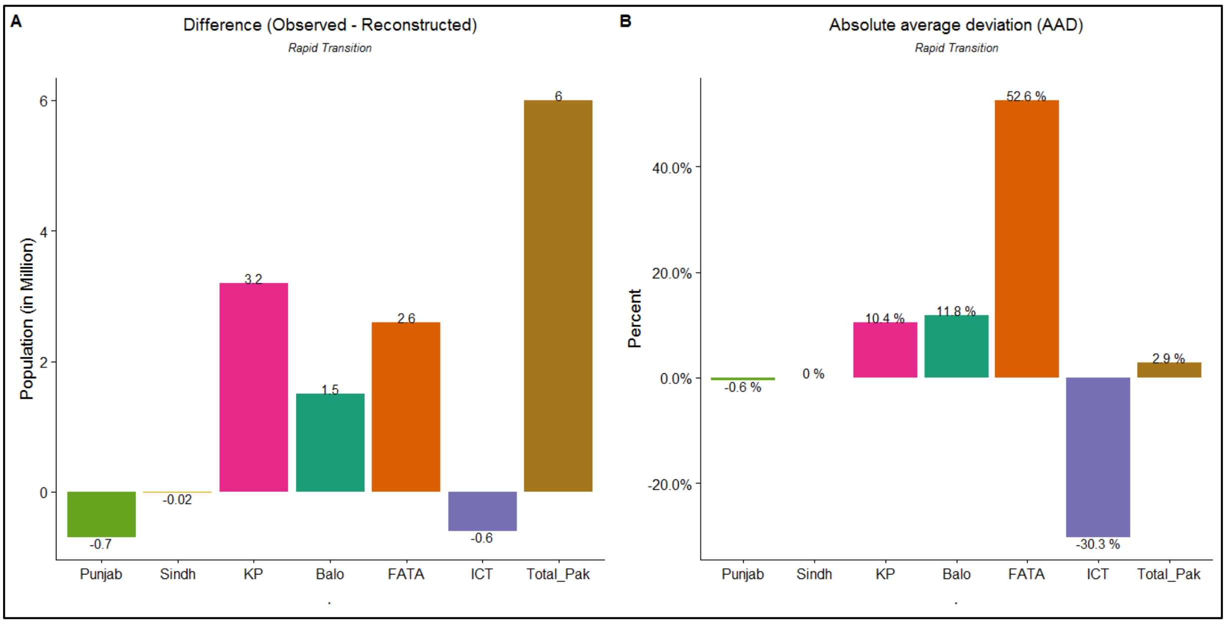
Table 6: Estimates of fertility and mortality at sub-national of Pakistan under the standard, rapid and stalled variants, 1998-2002 and 2013-17

Scenario	Period	Indicator	Punjab	Sindh	KP	Balochistan	FATA	ICT
Standard Variant	1998-2002	TFR	4.8	5.2	5.5	6.1	6.0	3.7
		Life expectancy at birth - female	59.3	58.7	62.9	59.8	57.2	67.6
		Life expectancy at birth - male	57.7	57.2	61.1	58.2	55.7	65.4
		Net international migration (000s)	-365.5	-56.4	-176.3	-8	-52.9	-3.5
		Net internal migration (000s)	249.3	-2.9	-48.8	-33.8	-156.6	47.7
	2013-2017	TFR	3.4	3.7	3.8	4.5	4.8	3.2
		Life expectancy at birth – female	61.6	62.6	65.1	62.2	63.6	68.9
		Life expectancy at birth – male	60.0	60.9	63.3	60.6	61.8	66.8
		Net international migration (000s)	-904.1	-176	-509.7	-22.5	-92.1	-16.4
		Net internal migration (000s)	360	186.6	54.9	-154.1	-707.7	480.5
Rapid Variant	1998-2002	TFR	4.3	4.7	5.0	5.6	5.5	3.2
		Life expectancy at birth - female	61.3	60.7	64.9	61.8	59.4	69.6
		Life expectancy at birth - male	59.7	59.2	63.1	60.2	57.7	67.4
		Net internal migration (000s)	311.6	-3.7	-61.1	-42.3	-195.3	59.6
	2013-2017	TFR	2.9	3.2	3.3	4.0	4.3	2.7
		Life expectancy at birth - female	63.6	64.6	67.1	64.2	65.6	70.9
		Life expectancy at birth - male	62.0	62.9	65.3	62.6	63.8	68.8
		Net internal migration (000s)	450	233.3	68.7	-192.6	-884.6	600.6
Stalled Variant	1998-2002	TFR	5.3	5.7	6.0	6.6	6.5	4.2
		Life expectancy at birth - female	57.3	56.7	60.9	57.8	55.2	65.6
		Life expectancy at birth - male	55.7	55.2	59.1	56.2	53.7	63.4
		Net internal migration (000s)	187	-2.2	-36.6	-25.4	-117.5	35.8
	2013-2017	TFR	3.9	4.2	4.3	5.0	5.3	3.7
		Life expectancy at birth - female	59.6	60.6	63.1	60.2	61.6	66.9
		Life expectancy at birth - male	58.0	58.9	61.3	58.6	59.8	64.8
		Net internal migration (000s)	270	140	41.2	-115.6	-530.8	360.4

Source: Authors' calculations

Results for the Rapid transition scenario are depicted in Figure 7. This scenario leads to similar population estimates as those of the census for Punjab and Sindh, indicating a rapid fertility and mortality transition in these two provinces. The difference with the census is larger in the other provinces and regions, similarly to what was observed in the standard variant (see Figure 6).

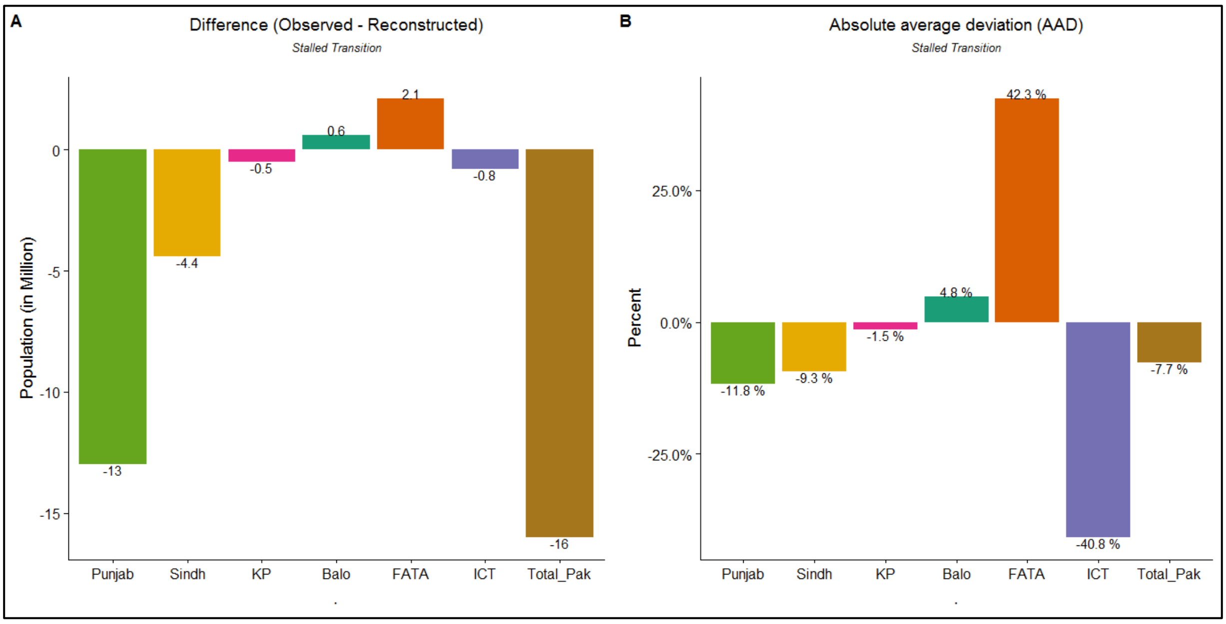
Figure 7: Difference of the 2017 census results to the retrospective population projections (A) and AAD (B) for Pakistan, by provinces and regions according to Rapid Transition Scenario



Source: Authors' calculations

Alternatively, the stalled transition scenario that assumes a slower transition pace compared to the standard variant provides a better fit for the official census results for KP and Balochistan.

Figure 8: Difference of the 2017 census results to the retrospective population projections (A) and AAD (B) for Pakistan, by provinces and regions *according to Stalled Transition Scenario*



Source: Authors‘ calculations

6. Study Limitations

This study has some limitations, which have to be pointed out. First, there is the inherent weakness of demographic analyses which do not measure census coverage as PES would.

The second constraint is that the reference census date to compare the census and reconstructed results is not accurately specified, because of the long duration of the enumeration period of the census. The census enumeration was divided into two phases and completed in 70 days: the first phase was from 15 March to 13 April 2017, and the second phase from 25 April to 24 May 2017. We found that the net undercounts at the national level varied from 5.0 million on 15 March 2017 to 5.6 million on 25 May 2017 (compared to 4.9 million in a standard variant, see Figure 6). Because of the small distortion in the net undercount, we used 15 April 2017 as a reference date.

Third, we further assume that the coverage of the 1998 census at sub-national level was in acceptable range. This is also a possible limitation because the 1998 census figures at sub-national levels may also be affected by issues of under- or over-count.

Finally, we have collected many surveys to come up with the demographic components of intercensal change, and we rely on their quality and representativeness that may not be optimal. Moreover, in many cases, we used indirect methods to come up with the estimates and this may induce some uncertainties/errors in the data. This is particularly the case for the internal migration estimates that could bias the intercensal results, especially since we noticed that the results of the reconstruction are highly sensitive to the level of internal migration

7. Conclusion

The census provides a unique opportunity to get reliable and comprehensive estimates of various socio-demographic variables. The mechanisms to assess the quality of the 2017 census in Pakistan were not part of the statistical operation. In this context, the demographic analysis provides a tool to check potential concerns with the quality of the census data. Utilizing all available empirical evidence, we derive estimates of fertility and mortality from full birth and summary birth histories by applying robust statistical models. We also derive complete origin-destination internal migration flows that are estimated indirectly using sample surveys. Using all these components of population change including international migration, the 2017 population by age and sex of the country at subnational level is prospectively reconstructed using the 1998 census data as base-year.

The reconstructed estimates of mortality and fertility give a consistent picture of demographic developments in each province and region over the last three decades. Though the historical trend is not available, mortality continued falling since the 1980s in all provinces. Child mortality was the first to drop, followed somewhat later by declines in infant mortality with the varying pace at subnational level: the pace of decline (both in under-five and infant mortality) in Punjab and Sindh was slowest compared to other regions and provinces. There was also a more gradual increase in life expectancy at birth for male and female alike in all subnational entities with varying pace: Slow in Punjab, Sindh, and Balochistan and rapid in other provinces and regions. The decline in fertility followed in the 1990s and 2000s, across all provinces and regions. However, fertility is still quite high at the sub-national level, above three children per woman in 2017. Although this study was largely descriptive and data-driven, the social and economic factors that contributed to mortality and fertility decline in all provinces and regions clearly deserve in-depth analysis in their own right.

International and internal migration flow emerge as important sources of population change within and between the sub-national levels in Pakistan. Despite the recognized importance of migration flows, remarkably little progress has been made to estimate them reliably and to measure their impact on population dynamics and socio-economic development. For international migration, we have sought to address the issue by harnessing national and international datasets and estimate the consistent series of in-and-out migration flows from each subnational level. Similarly, we have developed a robust statistical model to estimate internal migration flows between sub-national levels based on data from LFS during the last two decades. While not without limitations (see section 6), this method reveals the expected variation in the incidence of internal migration: Punjab, Sindh, and ICT as net-receiver during the last 15 years.

Our analysis showed three main interesting findings. First of all, that at the national level, the standard reconstruction leads to results close to that of the census. There are more differences at the sub-national level with the standard variant leading to larger populations in Punjab, Sindh, and ICT compared to the census and lower population in KP, Balochistan, and FATA. Second, we show that the reconstruction is highly sensitive to internal migration. Finally, the two counterfactual projections show that while Sindh and Punjab seemed to have followed a rather fast demographic transition, it is not the case of the other regions.

This analysis has emphasized the importance of independent analysis of census data using demographic analysis, particularly in low-income countries. While the analysis pointed out the possibility of under- and over-coverage at the sub-national level, there are too many uncertainties in the data used for the reconstruction to certain about it. However, this analysis should prompt the government and the international community to prepare for the next census to ensure the high quality of the enumeration procedure. In particular, the census should use both ***de-jure and de-facto*** population count within a short enumeration period (ideally 3-5 days), collect the information on the previous place of residence and duration of stay at current residence (status of internal migration) and ensure that a **Post-enumeration Survey (PES)** and **Demographic Analysis** are regular features of the census operations.

References

Abel, G., 2013. Estimating global migration flow tables using place of birth data. *Demographic Research,* Volume 28, pp. 505-546.

Abel, G., 2017. Estimates of global bilateral migration flows by gender between 1960 and 2018. *International migration review.*

Abel, G., 2018. *migest: Methods for the Indirect Estimation of Bilateral Migration. R package version 1.7.4..* s.l.:R Core Development Team.

Abel, G. & Sander, N., 2014. Quantifying global international migration flows. *Science,* Volume 343, pp. 1520-1522.

Arriaga, E., 1983. *Estimating fertility from data on children ever born by age of mother,* Washington, D.C.: United States Bureau of the Census.

Arriaga, E., Johnson, P. & Jamison, E., 1994. *Population Analysis with Microcomputers.* s.l.:U.S. Census Bureau, International Programs Software.

Booth, H., 1984. Transforming Gompertz's Function for Fertility Analysis: The Development of a Standard for the Relational Gompertz Function. *Population Studies,* 38(3), pp. 495-506.

Borges, G. & Sacco, N., 2016. *Has census coverage in Latin America improved with time?,* Sao Paulo: Unicamp, Nepo, IUSSP, Campinas.

Cibea, A. et al., 2013. *A silk routes partnership for migration: Pakistan migration country report,* Vienna: International Centre for Migration Policy Development.

Cohen J.E., Roig, M., Reuman, C. & GoGwilt, C., 2008. International migration beyond gravity: A statistical model for use in population projections. *PNAS,* 105(40), pp. 15269-15274.

Feeney, G. & Alam, I., 2003. Fertility, Population growth and accuracy of census enumeration in Pakistan: 1961-1998. In: A. Kemal, M. Irfan & N. Mehmood, eds. *Population of Pakistan: An Analysis of 1998 Population and Housing Census.* Islamabad: Pakistan Institue of Development Economics (PIDE) and UNFPA, pp. 73-88.

Gazdar, H., 2003. *A review of migration issues in Pakistan,* Karachi: Collective for Social Science Research.

George, M., Stanley, K., Swanson, D. & Tayman, J., 2004. Population Projections. In: S. a. Swanson, ed. *The Methods and Materials of Demography.* s.l.:Elsevier, Academic Press, pp. 561-601.

Hill, K. et al., 2012. Child Mortality Estimation: Accelerated Progress in Reducing Global Child Mortality, 1990-2010. *PLoS Medicine,* 9(8).

Irfan, M., 1986. *Mortality and health issues: mortality trends and patterns in Pakistan,* Bangkok: United Nations ESCAP.

Karim, M. & Nasar, A., 2003. Migration patterns and differentials in Pakistan: Based on the analysis of 1998 census data. In: A. Kemal, M. Irfan & N. Mehmood, eds. *Population of Pakistan: An Analysis of 1998 Population and Housing Census.* Islamabad: Pakistan Institute of Development Economics (PIDE), UNFPA.

KC, S., Wurzer, M., Speringer, M. & Lutz, W., 2018. Future population and human capital in heterogeneous India. *Proceedings of the National Academy of Sciences (PNAS) of the United States of America,* pp. 2-6.

Kemal, A., Irfan, M. & Mahmood, N., 2003. *Population of Pakistan: An analysis of 1998 Population and Housing Census.* Islamabad(ICT): Pakistan Institute of Development Economics (PIDE) & UNFPA.

Mariani, F., Perez-Barahona, A. & Raffin, N., 2009. *LIfe expectancy and the Environment,* Bonn, Germany: The Institute for the Study of Labor (IZA).

Marti, M. & Rodenas, C., 2007. Migration Estimation Based on the Labour Force Survey: An EU-15 Perspective. *International Migration Review,* Volume 41, pp. 101-126.

Ministry of Overseas Pakistanis and Human Resource Development, 2016. *Labor Migration from Pakistan: 2015 status report,* Islamabad: Govt. of Pakistan.

Ministry of States & Frontier Regions; Population Census Organization; UNHCR, 2005. *Census of Afghans in Pakistan 2005,* Islamabad: Ministry of States & Frontier Regions; Population Census Organization; and UNHCR.

Minnesota Population Center, 2017. *Integrated Public Use of Microdata Series, International: Version 6.5 [dataset].* Minneapolis, MN: University of Minnesota.

Murray, C. et al., 2007. Can we achieve Millennium Development Goal 4? New analysis of country trends and forecasts of under-5 mortality to 2015. *Lancet,* 370(9592), pp. 1040-54.

National Research Council, 2000. *Beyond Six Billion.* John Bongaarts; Rodolfo A. Bulaato ed. Washington D.C.: National Research Council.

Pakistan Bureau of Statistics (PBS), 2017. *Provisional Summary Results of 6th Population and Housing Census,* Islamabad: Pakistan Bureau of Statistics, Ministry of Statistics, Islamabad, Islamic Republic of Pakistan.

Pedersen, J. & Liu, J., 2012. Child Mortality Estimation: Appropriate time periods for child mortality estimates from full birth histories. *PLOS Medicine,* 9(8), pp. 19-91.

Population Division, DESA, 2017. *Trends in International Migrant Stock: Migrants by Destination and Origin (United Nations database, POP/DB/MIG/Stock/Rev.2015).,* New York: Population Division, Department of Economic and Social Affairs.

Population Division, DESA, 2017. *World Population Prospects: The 2017 Revision,* New York: United Nations, Department of Economic and Social Affairs, Population Division.

R Development Core Team, 2018. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A., Lalic, N. & Gerland, P., 2014. Joint probabilistic projection of female and male life expectancy. *Demographic Research,* Volume 30, pp. 795-822.

Raymer, J. & Rogers, A., 2006. *Applying Model Migration Schedules to Represent Age-Specific Migration Flows,* Boulder: Institute of Behavioral Sciences.

Rees, P., Bell, M., Duke-Williams, D. & Blake, M., 2000. Problems and solutions in the measurement of migration intensities: Australia and Britain compared. *Population Studies,* Volume 54, pp. 207-222.

Robinson, J., 2010. *Coverage of Population in Census 2000 based on Demographic Analysis: The History Behind the Numbers,* Washington, DC: U.S. Census Bureau.

Rogers, A. & Castro, J., 1981. *Model migration schedules,* Vienna: International Institute for Applied System Analysis.

Rukanuddin, A. & Farooqui, N., 1987. *The State of Population of Pakistan.* Islamabad: National Institute of Population Studies (NIPS).

Sathar, Z., Wazir, M. & Sadiq, M., 2014. *Prioritizing family planning for achieving provincial maternal child health and development goals,* Islamabad: Population Council.

Sawyer, C., 2012. Child Mortality Estimation: Sex differences in childhood mortality since the 1970s. *PLoS Medicine,* 9(8).

Schoumaker, B., 2013. A Stata module for computing fertility rates and TFRs from birth histories: tfr2. *Demographic Research,* Volume 28, pp. 1093-1144.

Silva, R., 2012. Child Mortality Estimation: Consistency of under-five mortality rates estimates using full birth histories and summary birth histories. *PLOS Medicine,* 9(8), pp. 32-45.

Spoonrenberg, T., 2013. Demographic Changes in Myanmar since 1983: An examination of official data. *Population and Development Review,* 39(2), pp. 309-324.

Spoorenberg, T., 2012. Demographic Changes in North Korea: 1993-2008. *Population and Development Review,* 38(1), pp. 133-158.

StataCorp, 2013. *Stata Statistical Software: Release 13.* s.l.:College Station, TX: StataCorp LP..

Statistic Section, P., 2018. *Azad Jammu & Kashmir Statistical Year Book 2017,* Muzaffarabad: Planning & Development Department, Govt. of AJ&K.

U.S. Bureau of Census, 1985. *Evaluating Censuses of Population and Housing,* Washington, D.C.: U.S. Bureau of Census.

UN IGME, 2015. *Levels and trends in child mortality, Estimates developed by UN Inter-agency group for child mortality estimation,* New York: United Nations Children's Fund, World Health Organization, The World Bank, and United Nations Population Division.

UN Office for the Coordination of Humanitarian Affairs (OCHA), 2017. *Pakistan: KP and FATA - Areas of Displacement, Hosting and Returns as of 31 October 2017.* Islamabad: UNOCHA.

United Nations, 2008. *Principles and recommendations for population and housing censuses Revision 2,* New York: Statistics Division, DESA.

United Nations, 2013. *Adolescent Fertility since the International conference on population and development (ICPD) in Cairo,* New York: United Nations.

Wazir, M., 2012. *Evolution of population and education in Pakistan: Analysis and projections by provinces,* Vienna: University of Vienna.

Wazir, M., 2018. Fertility Preferences in Pakistan. In: S. Gietel-Basten, J. Casterline & M. Choe, eds. *Family Demography in Asia: A Comparative Analysis of Fertility Preferences.* London: Edward Elgar.

Willekens, F., Zinn, S. & Leuchter, M., 2017. Emigration rates from sample surveys: An application to Senegal. *Demography,* Volume 54, pp. 2159-2179.

Wisniowski, A., 2017. Combining Labour Force Survey data to estimate migration flows: The case of migration from Poland to the UK. *Journal of the Royal Statistical Society: Series A,* Volume 180, pp. 185-202.

Annex

Description of Sub-National Entities

Pakistan consists of four provinces (Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan), Islamabad Capital Territory (ICT) a federal capital, and Federally Administered Tribal Areas (FATA) a semi-autonomous tribal region in northwestern Pakistan and directly governed by Pakistan's federal government. Through 25th constitutional amendment on 31 May 2018, FATA was officially merged with Khyber Pakhtunkhwa. Azad Jammu and Kashmir (AJK) are self-governing territories administrated by Pakistan through the Ministry of Kashmir Affairs and Gilgit Baltistan. Gilgit Baltistan (GB) (formally known as "Northern Areas") is an administrative territory in Pakistan and a candidate for the fifth province. GB was granted limited autonomy and renamed GB via self-governance order 2009. United Nations referred AJK and GB as "Pakistan-administered Kashmir". AJK and GB have not represented in the Parliament of Pakistan. However, PBS has implemented the census operations in all these provinces and regions including AJK and GB.

VIENNA INSTITUTE OF DEMOGRAPHY

Working Papers

Yegorov, Yuri, Dieter Grass, Magda Mirescu, Gustav Feichtinger and Franz Wirl, *Growth and Collapse of Empires: A dynamic Optimization* Model, VID Working Paper 05/2019.

Chisumpa, Vesper H., Clifford O. Odimegwu and Nandita Saikia, *Mortality in Sub-Saharan Africa: What is Killing Adults Aged 15-59 Years in Zambia?*, VID Working Paper 04/2019.

Striessnig, Erich and Jayanta Kumar Bora, *Under-Five Child Growth and Nutrition Status: Spatial Clustering of Indian Districts*, VID Working Paper 03/2019.

Speringer, Markus, Anne Goujon, Samir K.C. and Michaela Potančoková, Claudia Reiter, Sandra Jurasszovich and Jakob Eder, *Global Reconstruction of Educational Attainment, 1950 to 2015: Methodology and Assessment*, VID Working Paper 02/2019.

Testa, Maria Rita and Danilo Bolano, *Intentions and Childbearing in a Cross-Domain Life Course Approach: The Case of Australia*, VID Working Paper 01/2019.

Goujon, Anne, Claudia Reiter and Michaela Potančoková, *Religious Affiliations in Austria at the Provincial Level: Estimates for Vorarlberg, 2001-2018*, VID Working Paper 13/2018.

Spitzer, Sonja, Angela Greulich and Bernhard Hammer, *The Subjective Cost of Young Children: A European Comparison*, VID Working Paper 12/2018.

Jatrana, Santosh, Ken Richardson and Samba Siva Rao Pasupuleti, *The Effect of Nativity, Duration of Residence, and Age at Arrival on Obesity: Evidence from an Australian Longitudinal Study*, VID Working Paper 11/2018.

Nitsche, Natalie and Sarah Hayford, *Preferences, Partners, and Parenthood: Linking Early Fertility Desires, Union Formation Timing, and Achieved Fertility*, VID Working Paper 10/2018.

Riederer, Bernhard, *Experts' Expectations of Future Vulnerability at the Peak of the "Refugee Crisis"*, VID Working Paper 9/2018.

Cukrowska-Torzewska, Ewa and Anna Matysiak, *The Motherhood Wage Penalty: A Meta-Analysis*, VID Working Paper 8/2018.

Nitsche, Natalie and Hannah Brückner, *High and Higher: Fertility of Black and White Women with College and Postgraduate Education in the United States*, VID Working Paper 7/2018.
