POPULATION HOMEOSTASIS IN SUB-SAHARAN AFRICA

David de la Croix and Paula E. Gobbi

LIDAM Discussion Paper IRES
2021 / 26

IRES | Louvain Institute of Data Analysis and Modeling in economics and statistics

Population Homeostasis in Sub-Saharan Africa

David de la Croix* Paula E. Gobbi†

Abstract

Global population growth remains one of the major challenges of the twenty-first century. This is particularly true for African countries which have been undergoing their demographic transitions. To investigate whether predicted increasing population density and urbanization can help to stabilize African population, we construct a database for 84 georeferenced Demographic and Health Survey (DHS) samples including 947,191 individuals in sub-Saharan Africa and match each location with gridded population density from NASA. We apply a proportional hazard model to evaluate the quantitative impact of local population density on the transitions from childlessness to motherhood, and from celibacy to marriage. Moving from the 5th to the 95th percentile of population density increases the median age at first birth by 2.2 years. This roughly decreases completed fertility by half a child. The same increase in population density increases the median age at first marriage by 3.3 years. These findings contribute to the understanding of why fertility has not dropped in Africa as fast as expected. One part of the answer is that population density remains low. Yet the total effect of increased density on fertility remains limited and counting on it to stabilize the population would be unrealistic.

Keywords: Fertility, Homeostasis, Africa, Population density.

*IRES/LIDAM, UCLouvain, Belgium and CEPR, London.
†ECARES, Université Libre de Bruxelles, Belgium and CEPR, London.

1 Introduction

Projections by the United Nations (UN) or the International Institute for Applied Systems Analysis (IIASA) show that the global population will peak at the turn of the next century at around 10 billion individuals.[1] The bulk of the increase leading to this peak comes from sub-Saharan Africa (Population Division 2019; Lutz et al. 2014). The African population is expected to double in the next 50 years, and triple by the end of the century. This is one key challenge that the global ecosystem will face in the upcoming decades.

On the one hand, the rise in the world's population strongly depends on how rapid fertility transitions will be. On the other hand, increasing population density is likely to trigger a drop in fertility – a property called population *homeostasis* (Lee 1987). Population homeostasis requires the presence of spontaneous convergence forces that keep population close to a stable, long-run level.[2] A question of major importance is whether these forces are strong enough.

Our contribution here is to quantify the population *homeostasis* property. In detail, we analyze the relationship between population density and fertility in sub-Saharan Africa, in order to better understand whether spontaneous convergence forces are at work. To do so, we combine individual fertility data from Demographic and Health Surveys (DHS) with gridded population density data from NASA. They are based on detailed population data from census administrative units.[3] In DHS data, individuals belong to georeferenced clusters, which allow for mapping population density onto fertility.

Once the local population density is known for each woman, we can run a statistical model relating the key determinants of fertility to the population density of the cluster in which they live. We focus on the starting time of reproduction, using either the age at marriage or the age at first birth. In addition to birth spacing and the age at stopping, which we do not consider here, these two dates are, for any woman, important determinants of her overall completed fertility. Focusing on these dates allows us to assess how fertility reacts to population density *today*, without having to rely on synthetic *tempo* measures of fertility. The latter measures have indeed been shown not to have a strong

[1]To be precise, the peak is not foreseen by the UN within the projection period up to 2100, although it should happen soon thereafter. IIASA's projected population peaks at 9.7 billion as early as 2070. This is happening under the medium variant of the UN and the SSP2 scenario of IIASA.

[2]While homeostatis is a property of a dynamical system, it is related to the idea of the demographic transition (Wilson and Airey 1999). The demographic transition is usually seen as a shift from one stable demographic regime with high fertility and mortality to another stable regime with low fertility and mortality – following some large shock – while the homeostasis property ensures the convergence of demographic variables to some constant levels in the new regime. These different notions are developed in the Appendix.

[3]See `http://sedac.ciesin.columbia.edu/downloads/docs/gpw-v3/` for methodological details.

predictive power for completed fertility, precisely because of the change in the timing of childbearing over time (Bongaarts and Feeney 1998). Hence, we suggest that looking at these two main events, age at marriage and age at first birth, allows us to assess the population *homeostasis* property based on individuals' decisions today.

Compared to the previous literature that has studied the relationship between population density and fertility (Lutz and Qiang 2002; Lutz, Testa, and Penn 2006), we are the first to document the negative effect of population density on fertility using estimates from individual level data. Our main contribution is therefore to show that homeostasis forces are at play for human population, at a fine grid level. More generally, we show that macroeconomic conditions can affect individual fertility behavior in a permanent way. We also challenge the common wisdom that the main driver of the last phase of the demographic transition is family limitation obtained by reducing high order parity progression ratios, a *stopping* strategy. We show that homeostasis also operates through the *entry* into sexual activity (marriage and motherhood).

We also identify the plausible mediating variables responsible for a link between density and fertility. At the individual level, both higher education and health postpone marriage and motherhood. This is not surprising, but it is worth noting that these variables partly capture the effect of population density on fertility. At the collective level, higher education and health in the community to which an individual belongs also have the expected effects on fertility. Importantly, they also capture an additional part of the effect of population density on the probability of entering motherhood (ten percent). This paper thus relates to the more extensive literature on urbanization and economic development in general, and urbanization and demographic change in particular (Dyson 2011; Flückiger and Ludwig 2017).[4]

The magnitude of the overall effect of population density on fertility is sizable, but falls short of what is needed to foster a fertility decline towards the replacement level of the population. Moving from the 5th to the 95th percentile of population density increases the age at first birth by 2.2 years and the age at marriage by 3.3 years. From the observed relationship between the age at first birth and completed fertility, this roughly corresponds to a decrease by half a child at most.

The results are robust to the inclusion of controls for ethnicity, religion, proxies for household wealth, the education of spouses, and contraception knowledge. We show that

[4]See Collier (2017) and Parienté (2017) for discussions on urbanization, infrastructure and productivity in Sub-Saharan Africa. Urbanization has mostly been studied as the migration process of individuals who change location in search of better economic possibilities, but it is also associated with an increase in population density of urban areas, which has been called the internal urban population growth (Fox 2017; Jedwab, Christiaensen, and Gindelsky 2017). This paper naturally relates to this latter, and less explored, aspect of urbanization.

the results are neither driven by selection of individuals, nor by measurement errors in the data.

2 Data

Our main source of data is Demographic and Health Surveys (DHS). We consider every sub-Saharan country for which GPS-coordinates information is available. This amounts to 34 countries. For the majority of these countries, survey data were collected during different DHS phases. We include every "standard DHS" type of survey for phases II and above.[5] The list of countries and DHS phases are provided in Table S.1. We use the individual recode, the household recode, and the GPS dataset. Households are grouped into clusters for which we know the latitude and the longitude from the DHS GPS file. The total number of individuals for each country is shown in the supplementary material, together with the list of variables used, and some descriptive statistics.

From the individual recode, we built a sample consisting of women between 15 and 49 years of age whose cluster of residence is known. We use information on: age, education, partner's education, total number of children ever born, total number of living children, religion, age at first birth, age at first marriage, whether she moved from her place of residence after 14, and ethnicity. From the household recode, we use the information on whether or not the household has electricity or/and a refrigerator. These two variables are used as additional controls to proxy for income. From the GPS dataset, we use the geographical coordinates of each cluster.

The geolocation of DHS samples allows us to combine these data with three sources of geographical data. First, population density raster files are taken from the Center for International Earth Science Information Network (CIESIN) and International Food Policy Research Institute (IFPRI) and The World Bank and Centro Internacional de Agricultura Tropical (CIAT) (2011). They provide information on population density in grids with cell sizes of 30"×30" (approximately 1 km^2). To avoid a possible reverse causality from fertility to population density, we use density in 1990, which is the earliest year available.

The second geographical information that we include is a measure for land productivity. We use one of the caloric suitability indexes developed by Galor and Özak (2016) which has a resolution of 5'×5' (approximately 100 km^2). Galor and Özak (2015) show that the caloric suitability index performs better than conventionally used agricultural suitability data (Ramankutty et al. 2002) in terms of capturing the effect of land produc-

[5]We do not keep DHS data collected during the first phase because these took place prior to our measure for population density.

tivity. We use the raster file for the maximum potential caloric yield attainable given the set of all suitable crops in the post-1500 period. This yield varies across cells depending on their climatic and geographic characteristics, such as elevation, temperature, rainfall, soil quality, terrain ruggedness, steepness, etc.

Finally, as a proxy for income per capita, we use the GDP measures from Ghosh et al. (2010), which are essentially based on nighttime light satellite data. Henderson, Storeygard, and Weil (2012) show that luminosity is a strong proxy of GDP. This proxy has two main advantages. First, it provides a harmonized measure for total economic activity across countries at a disaggregated level. And second, it allows to account for the informal sector, which is often important in developing countries and difficult to include in national statistics. The precision level of the raster is 30"×30"; however, measurement errors at the pixel level are large.[6] Ashraf, Galor, and Klemp (2015) argue in favor of measuring GDP on the basis of a continuum of a larger number of nighttime light pixels. We therefore base our measure on an aggregated 20'×20' raster. To obtain a per capita variable, we divide GDP by our measure of population density taken at the same level of aggregation and discarding pixels with fewer than 0.1 inhabitants per km^2. For every cluster, we impute its GDP as the mean within a circle of 50km in radius.

3 Results

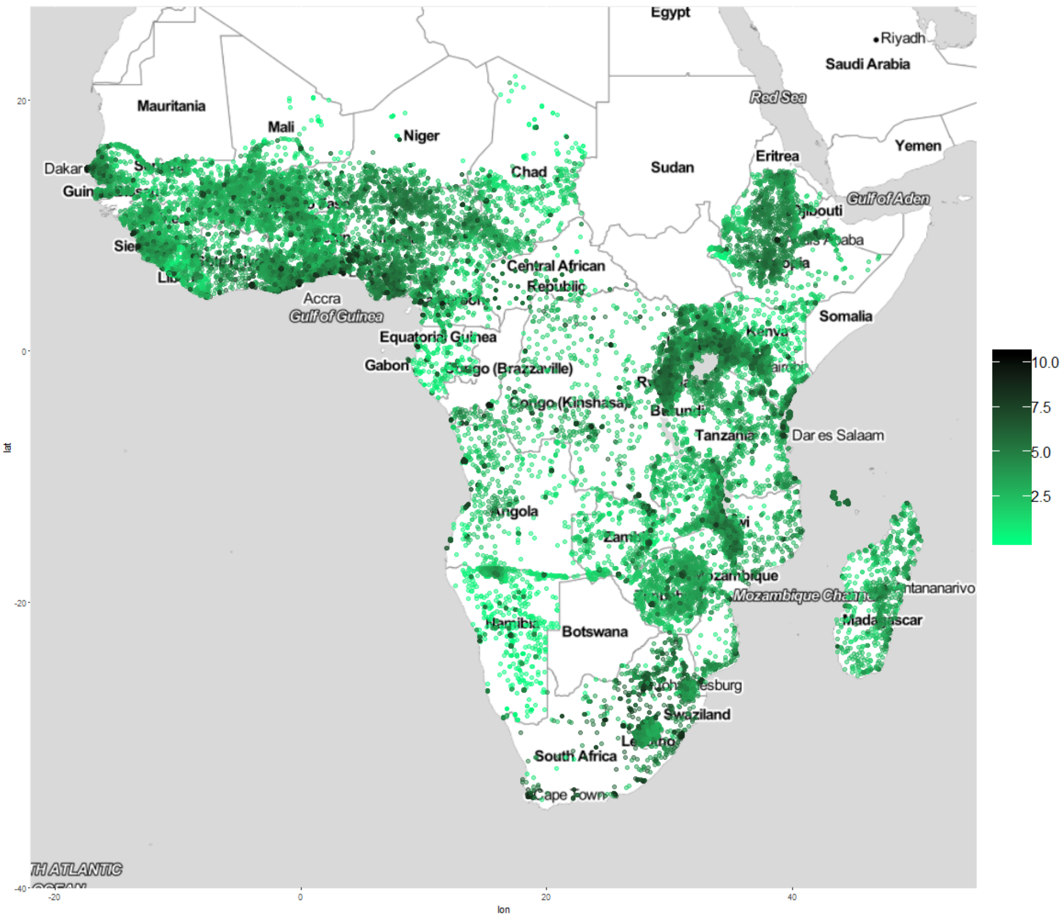
3.1 Demographics in Sub-Saharan Africa

Figure 1 shows the location of all clusters in sub-Saharan Africa from the Demographic and Health Surveys. The shade of green represents the population density in 1990 in these clusters from the gridded population density raster built by NASA.[7] Population density ranges from 0.01 inhabitants per square kilometer in the Karas region (Namibia) to 32,861 in Addis Ababa (Ethiopia). As we have merged all the waves of the DHS, we obtain a comprehensive coverage of Africa, both across countries and across urban/rural areas.

The sample includes 947k women, with an average age of 28. The average total number of children ever born is 2.9. The average age at first birth is 19, and the average age at first marriage is 18. In practice, variations in the number of children per woman depend on the age at marriage and first birth, on the time between each birth (spacing), and/or on when

[6] For example, they can be due to over-glow and blooming. We also check whether one should correct for gas flares, but the measure from Ghosh et al. (2010) seems to have filtered them out.

[7] Center for International Earth Science Information Network (CIESIN) and International Food Policy Research Institute (IFPRI) and The World Bank and Centro Internacional de Agricultura Tropical (CIAT) (2011).



Note: Population density is reported as $ln(1 + \text{population density})$.

Figure 1: Cluster Localization and Population Density.

the last child occurred (stopping). Standard DHS provide the complete history of birth for each woman. The weaknesses of these surveys, as reported in Schoumaker (2014), are particularly relevant to analyzing the spacing between births. We can however still check for the first proximate determinant of fertility - birth and marriage postponement - by studying the determinants of age at first birth and age at first marriage. The supplementary material appendix shows that these two dates are very strong predictors of completed fertility.

The unconditional probability (hazard rate) of becoming a mother and of marrying changes with density. Figure 2 plots the hazard rates[8] as a function of age dividing the sample into four groups of equal size, according to the population density in their area. These are unconditional probabilities, i.e. we do not control for anything but age. Figure 2 displays two salient features. First, there is a postponement in the mean age

[8]Computed in R with the package `muhaz` which estimates a hazard function from right-censored data using kernel-based methods.

at first birth when population density is higher: the probability of becoming a mother peaks at 18 years (220 months), and drops quickly after this peak for the first quartile. In the last population density quartile, the probability peaks over the range between ages 20 and 28 (240-340 months). Second, high density areas have lower risks associated to childbearing at each age. The same description applies to the probability of marrying (right panel).

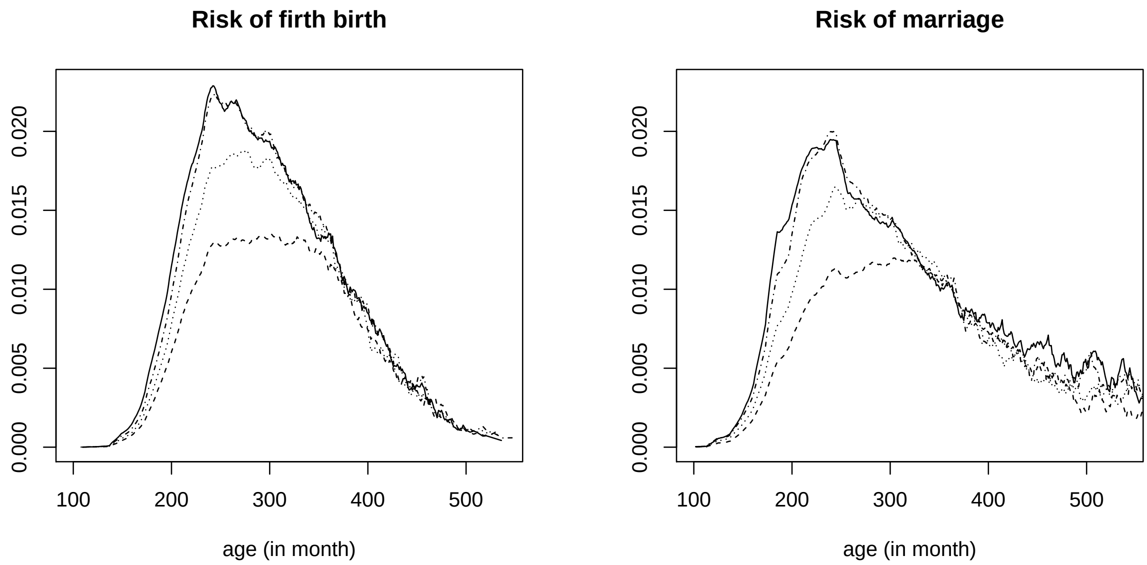


Figure 2: Unconditional Probability of Becoming a Mother (Left) and Marrying (Right) as a Function of Age (in Months) by Population Density Quartile (Q1 - solid line, Q2 - dashed-dotted line, Q3 - dotted line, Q4 - dashed line. Kaplan-Meier estimates.

3.2 Average Effect of Population Density on Population Dynamics

We study how population density affects birth or marriage postponement through a proportional hazard model. The unit of observation is a woman. The probability that woman j living in cluster c will exit childlessness or singlehood at time a, denoted $\lambda_{jc}(a)$, is

$$\lambda_{jc}(a) = \lambda_0(a) \exp\left\{ \tau_1 \ln(1 + \text{density}_c) + \sum_{i=2}^{N} \tau_i X_{jci} + \sum_{i=N+1}^{N_c} \tau_i X_{ci} \right\}. \quad (1)$$

According to Equation (1), the baseline hazard rate, $\lambda_0(a)$, is shifted proportionally by the density in the cluster $\ln(1+\text{density}_c)$, by N individual controls X_{jc}, and by N_c cluster level controls X_c, both listed in the following paragraph. For the age at first marriage and the age at first birth, the hazard rates λ are computed from women's data, where one observes for an individual j the couple $(y_j; I_j)$, where $y_j = \min(t_j; c_j)$ is the minimum between the age at first event t_j (i.e. the survival time) and the age at interview c_j (i.e. the censoring time). The event indicator I_j equals 1 if the event, either a birth or a marriage, has been observed (i.e. $t_j \leq c_j$), and zero otherwise. The estimation uses the Breslow method to handle tied failures. Standard errors are clustered at the DHS cluster level, accounting for the possibility that observations might not be independent within each DHS cluster.

Tables 1 and 2 present the results of the estimated coefficients using either the age at first birth or the age at marriage as dependent variables. Column (1) shows the results for the total effects of population density when we only include survey fixed effects to the model. Column (2) includes a second-order polynomial on individual education. Column (3) adds individual-level covariates. These covariates are the marriage status of the respondent (only when the dependent variable is the birth hazard) and the proportion of children that have died, as observed at the time of the survey. This variable can proxy the overall health of a woman and her income status. Finally, Column (4) adds cluster-level variables that might influence the age at first birth or age at marriage. The first is land productivity, proxied by the amount of calories that the land can provide. This allows us to control for the carrying capacity of each location. Land productivity can positively affect the probability of having a first birth or marrying if a Malthusian-type of argument is at play. The average mortality in the cluster allow us to capture the effect of health institutions. We also control for the average GDP per capita in logs and the average education in the cluster. The coefficients for GDP per capita suggest that richer places are associated with a lower probability of marrying. As shown by Kravdal (2002), average education is an important factor affecting women's birth rates, above their individual education level. Recently, Kebele, Striessnig, and Goujon (2021) also show that average fertility has a negative effect on fertility intentions. This is in line with Beckerian theory, according to which more educated places are associated to economies where the returns to human capital are higher.

The magnitude of the total effect of population density (Column (1)) can be interpreted as the difference between the median age at first birth and the median age at first marriage for a hypothetical individual living in a low- vs. high-density area. We set low-density areas to those in the first decile of the distribution of density (4.7 ind/km^2), and high density areas to those in the tenth decile (3,750.8 ind/km^2). For a given age, the chance of becoming a mother (resp. of being married) in a high-density area is 52.5% (resp. 41.3%) of the chance of becoming a mother in a low-density area. We can also translate these probabilities into median age at first birth and at marriage. In a low-density area, the estimated median age at first birth is 19.0 years. In a high-density area,

Dependent variable:	*Probability of becoming a mother*			
	(1)	(2)	(3)	(4)
ln(1+density)	-0.099***	-0.038***	-0.024***	-0.014***
	(0.001)	(0.001)	(0.001)	(0.001)
education		0.018***	0.046***	0.052***
		(0.001)	(0.001)	(0.001)
(education)^2		-0.007***	-0.008***	-0.008***
		(0.000)	(0.000)	(0.000)
married			1.271***	1.268***
			(0.007)	(0.007)
infant mortality			0.627***	0.583***
			(0.007)	(0.007)
calories				0.014***
				(0.001)
mean mortality				0.577***
				(0.036)
log(GDP per capita)				-0.002
				(0.003)
mean education				-0.011***
				(0.001)
Observations	947,191	947,191	947,191	947,191

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level. All specifications include survey fixed effects.

Table 1: Cox Model Estimates for Age at First Birth

it is 21.2 years (+2.2 years). For marriage, these ages are 17.7 and 21.0 respectively (+3.3 years).

We also identify the plausible responsible mediating variables in Columns (2), (3), and (4). We distinguish between candidates for mediating effects at the individual (Columns (2) and (3)) and collective level (Column (4)). At the individual level, higher education postpones marriage and motherhood. This result is well known in the literature (Kravdal 2002). What matters here is that introducing these variables partly captures the effect of population density on fertility. In detail, sixty percent of the overall effect of population density on the probability of becoming a mother are captured by individual education. This is confirmed when computing the magnitude of the effect of population density at constant education as the difference between the median age at first birth and the median age at first marriage for a hypothetical individual living in a low- vs. high-density area. In a low-density area, the estimated median age at first birth is 19.5 years. In a high-density area, it is 20.3 years (+0.8 years). For marriage, these ages are 18.3 and 19.7 respectively (+1.4 years).

We can compare the effect of population density with the effect of education. First,

Dependent variable:	*Probability of marrying*			
	(1)	(2)	(3)	(4)
ln(1+density)	-0.136***	-0.059***	-0.057***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.002)
education		-0.062***	-0.057***	-0.036***
		(0.001)	(0.001)	(0.001)
$(\text{education})^2$		-0.003***	-0.003***	-0.003***
		(0.000)	(0.000)	(0.000)
infant mortality			0.537***	0.441***
			(0.008)	(0.008)
calories				0.007***
				(0.001)
mean mortality				1.249***
				(0.049)
log(GDP per capita)				-0.010***
				(0.003)
mean education				-0.051***
				(0.001)
Observations	947,191	947,191	947,191	947,191

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level. All specifications include survey fixed effects.

Table 2: Cox Model Estimates for Age at Marriage

using the coefficients of Column (2), a raising density from the first to the last decile (at given education level) has the same effect on the probability of becoming a mother as raising female education from 6 years to 8.9 years.[9] Second, using the coefficients of Column (1) for the total effect of density and of Column (2) for the effect of education, a raising density from the first to the last decile has the same effect as raising female education from 6 years to 12 years.[10]

In Column (3), better health (proxied by the inverse of the ratio between child deaths and total births) also postpones marriage and motherhood. Finally, at the collective level (Column (4)), GDP is negatively associated to both the probability of a first birth and of marrying, although it is only significant for the last. This points towards a Beckerian effect of income on fertility, rather than a Malthusian effect. The no significance might also witness that GDP is poorly measured by satellite lights at the cluster level. Higher education and better health in the community to which an individual belongs have the expected effects on fertility. Importantly, they also capture an additional part of the effect of population density on the probability of entering motherhood. After

[9] Solving $-0.038(8.23-1.74) = 0.018(e-6) - 0.007(e-36)$ for e gives 8.9.

[10] Solving $-0.099(8.23-1.74) = 0.018(e-6) - 0.007(e-36)$ for e gives 12.

accounting for both individual and collective plausible mediating variables, a residual effect of population density on the age at first birth and the age at marriage remains. Non-observable variables, such as the price of space or unobserved income effects, could further mediate the relationship.

3.3 Implications for the demographic transition

We have shown in the previous subsection that the magnitude of the overall effect of population density on fertility is sizable. Here we will show that it however falls short of what is needed to foster a fertility decline towards the replacement level of the population.

First, we provide evidence that both the age at first birth and the age at first marriage are excellent predictors of completed fertility. To illustrate this, Figure 3 shows the relationship between the average number of children ever born for each age at first birth (left panel) and for each age at first marriage (right panel) in months. The average number of children ever born is computed for the subsample of mothers aged 40 and more, being thereby (almost) at the end of their reproductive period. There are 151,190 women in this sample. For each age at first birth and age at first marriage, we then compute the average number of children ever born. For both the age at first birth and the age at first marriage, the relationship with completed fertility is linear over the age range 14-32. The corresponding regression lines are:

$$\text{number of children ever born } = 11.42 - 0.25 \times \text{ age at first birth} \tag{2}$$

$$\text{number of children ever born } = 10.13 - 0.21 \times \text{ age at first marriage.} \tag{3}$$

Hence, postponing birth by one year reduces fertility by one fourth of a child on average.

Second, we can use the relationship between ages at first birth and at marriage and fertility to link population density to the number of children ever born.

From Table 3 we see that moving from the 5th to the 95th percentile of population density increases the age at first birth by 2.2 years (from Column (3)) and the age at marriage by 3.3 years (from Column (5)). From the observed relationship between the age at first birth and completed fertility, this corresponds to a decrease by 0.6 children (from Column (4)). From the observed relationship between the age at marriage and completed fertility, this corresponds to a decrease by 0.7 children (from Column (6)).

One can also use Table 3 to evaluate the effect of population size on population growth in the coming century. Africa's population is expected to triple. Starting from median density, the last two lines show the effect of this tripling on ages at first birth and at marriage, as well as their implication for fertility. Obviously, the expected increase

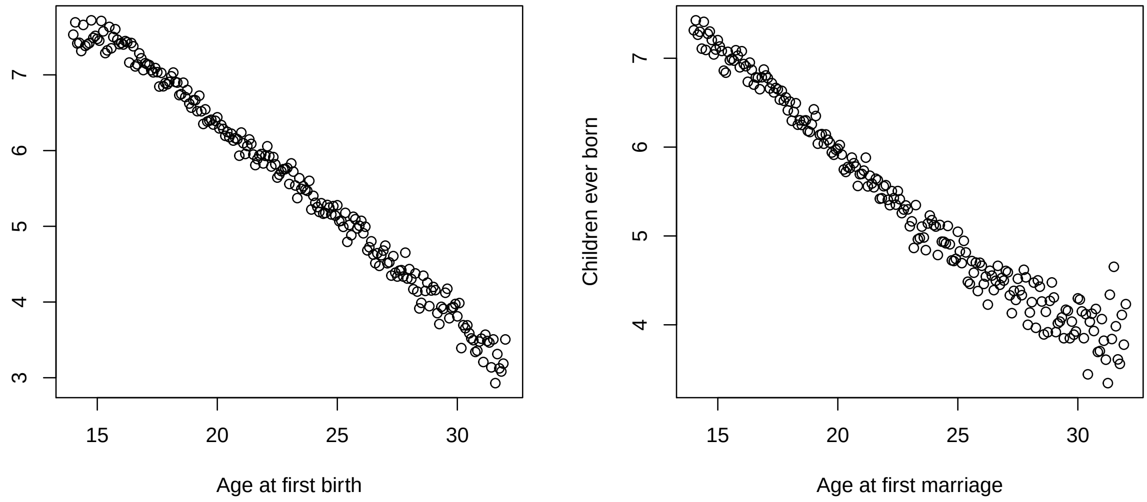


Figure 3: Age at first birth, age at first marriage, and children ever born for women aged 40+

in population density will have a limited effect on fertility in Sub-Saharan countries, although negative. This further suggests that population homeostasis is a slow-moving process. This is in line with recent estimates of population dynamics in a Malthusian economy found by Bouscasse, Nakamura, and Steinsson (2021).

To be more precise, we can compute the half-life of population dynamics implied by our estimates. Regressing population density (Column (2)) on fertility (Column (4)), we get a coefficient of -0.0836. This implies that the dynamics of population are given by $P_{t+25} = 0.9164P_t$ (assuming one generation is 25 years, and mortality is constant). The half life of these dynamics are $ln(0.5)/ln(0.9164) = 7.94$ generations, i.e. 198 years. We can compare this number to the literature. For pre-industrial England, Bouscasse, Nakamura, and Steinsson (2021) find a half-life of 150 years, higher than the previous estimate by Lee and Anderson (2002) of 107 years. Lagerlöf (2019) finds a half life of 356 years for pre-industrial Europe. For the developing world as a whole, De la Croix and Gobbi (2017) find a value of 102 years. For Japan, Sato (2007) estimates a value of 156 years. All these estimates point towards very slow dynamics. Africa is rather on the slow side.

3.4 Robustness

To assess the robustness of our findings, we do three robustness exercises. First, we include other potential determinants of the age at marriage and the age at first birth.

Table 3: Density and Completed Fertility

density quantile (1)	ln (1+ density) (2)	median age at first birth (3)	children ever born from (2) (4)	median age at marriage (5)	children ever born from (3) (6)
0.05	1.74	19.0	6.6	17.7	6.5
0.15	2.77	19.3	6.5	18.1	6.4
0.25	3.44	19.5	6.5	18.3	6.3
0.35	3.99	19.7	6.4	18.6	6.3
0.45	4.51	19.8	6.4	18.8	6.2
0.55	5.01	20.0	6.3	19.1	6.2
0.65	5.59	20.2	6.3	19.4	6.1
0.75	6.15	20.3	6.3	19.8	6.0
0.85	7.09	20.7	6.2	20.3	5.9
0.95	8.23	21.2	6.0	21.0	5.8
median	4.75	19.9	6.4	19.0	6.2
density x3	5.85	20.3	6.3	19.5	6.1

These are religion fixed effects, ethnicity fixed effects, proxies for family income and wealth, the education of spouses, and knowledge about contraception methods. The inclusion of these additional control variables does not alter the significance of the effect of population density. The size of the effect also remains the same overall, except when including the proxies for household income and wealth, which lowers the effect of density further. Second, we look at whether the results might be driven by migration, which could lead to biased estimates due to a selection of individuals into places with higher or lower population density. Dropping the women who changed residence after the age of 14 from the sample leaves the results unaffected. Finally, DHS might suffer from quality issues regarding the reported timing of some events. Dropping countries known to have poor quality data (Schoumaker 2014) amplifies the magnitude of the effect of density. Hence, our results might be seen as providing a lower bound on the true effect. Details are provided in the supplementary material.

4 Discussion and Concluding Remarks

Rather than starting from a unique theory and estimating a structural model derived from it, we estimated a reduced form, whose results can be compatible – or not – with different theories. Several theories originating in different fields have been put forward to understand population dynamics. Our analysis based on individual survey data can be read in light of these theories.

Biological and ecological mechanisms: Populations are simultaneously affected by two classes of mechanisms going in opposite directions (Fowler and Ruxton 2002): some lead to a decrease in fertility with increasing population size (called the *competition effect* by biologists), while others lead to an increase in fertility with increasing population size. An example of the latter is the *Allee effect*: in this instance, the harmful consequences of inbreeding reduce the fitness of a population as its size decreases. Our results clearly indicate that the competition effect is the dominant one for human reproduction in sub-Saharan Africa.

A non-monotonic relationship between fertility and population size has also been described by Lotka and Volterra (Lotka 1925; Volterra 1926) with their predator-prey model. In the original Lotka-Volterra model, it is mortality that channels the link between population density and population growth. Further extensions, as for instance De la Croix and Dottori (2008), show that the same type of interaction may occur through fertility. Instead of being eaten by predators, the preys refrain from procreating. Such a model implies a positive effect of density on the age at first birth and the age at marriage for low levels of density, and a negative effect of density for high levels of density. We test for such effects in the supplementary material. A positive effect is only found on the age at first marriage, when population density increases from the first to the second decile. More importantly, we find that the stabilizing force of population density is stronger when population density is larger (for deciles 8 to 10). This suggests that the demographic transition in sub-Saharan Africa will accelerate once population density has reached a certain threshold.

Demographic mechanisms: The idea that human fertility adjusts to population density precedes the research of biologists and ecologists. Montesquieu (1749) described the view of Greek philosophers on the issue (emphasis added): "In a small and flourishing territory, the number of citizens must soon augment, so as to become *a burden*. This people of consequence omitted nothing which might prevent an *undue increase of children*. Their politics were more immediately confined to the regulation of the number of citizens." The negative effect of population density on fertility was also explained by Malthus (1807). For the latter, when food is expected to become scarce, people limit their fertility. Malthus expressly stressed the role of marriage in regulating fertility. Following the Malthusian theory, other types of scarcities, such as land or housing, lead to the same effect (Ashraf and Galor 2011).

The link between population density and fertility was made explicit by Sadler (1830), who wrote against Malthus *The Law of Population – in disproof of the superfecundity of human beings, and developing the real principle of their increase*. His Law simply states that "The prolificness of human beings, otherwise similarly circumstanced, varies inversely as their numbers." The mechanism by which density influences fertility is the opposite of the Malthusian logic. For Malthus, higher density reduces resources per person, leading to a decline in fertility as a result of preventive (marriage is delayed) and positive checks (mortality increases). In the words of biologists, Malthus refers to the competition effect. For Sadler, on the other hand, affluence increases with population density, as it is purported to in theories of agglomeration externalities since Marshall (1890). When comparing Sadler's and Malthus's theories, both imply that fertility rates should be lower in more densely populated areas, but for different reasons. For Sadler, it is because those areas are richer than others, while for Malthus, it is the opposite.

The link we find between density and age at marriage supports either Sadlerian or Malthusian mechanisms, at least in the versions of the regression in which we do not control for income. When controlling for income (through calories and GDP per capita), under a strict Sadlerian or Malthusian model, density should not matter, because its effect should always go through income. We still find some residual effect of density, which might be because we do not perfectly control for individual income. Moreover, the negative effect of GDP per capita on the risk of marriage reflects the fact that regions with higher income have less marriages than poorer regions, keeping education and health constant across them. This points towards Sadlerian views.

Economic mechanisms: In economics, fertility choices are analyzed following the work of Becker (1993), who claims that the more productive an economy is, the more expensive the time spent on children is, and hence, the lower fertility will be. The Beckerian approach can be "augmented" to account for the effects of population density on fertility by introducing one of the following three features: the housing market, the provision of public infrastructure (education or health system), and an endogenous technology. For the first feature, higher density entails an agglomeration effect and a congestion effect (Sato 2007). The agglomeration effect leads to higher productivity which negatively affects fertility, in line with the Sadler model that we described previously. The congestion effect implies that the price of land and the cost of living are higher, similarly to Malthus's intuitions. Both effects diminish fertility. The second feature introduces the provision of public infrastructure (Boucekkine, de la Croix, and Peeters 2007; Becker, Cinnirella, and Woessmann 2010). When population density increases, it is easier to cover the fixed cost of infrastructure such as schools, and their provision increases (Boucekkine, de la Croix, and Peeters 2007). An increased provision of schools encourages parents to substitute quality for quantity, hence having fewer but better educated children (Becker, Cinnirella, and Woessmann 2010). Hence, higher density leads to more education and lower fertility. The third feature is based on technological progress (Galor and Weil 2000). A denser population increases the pace of technological progress, allowing for faster growth. Higher human capital is therefore required more acutely in the production process in order to deal with fast technical change. The return to education increases, and parents are led to invest more in the quality of their children, at the expense of quantity.

Our results provide strong support for mechanisms linking density to fertility through both health infrastructure and education. Individual education delays birth for education levels above 4 years, because of the quadratic term. Education in the community also leads to birth postponement. Both are plausible mediating variables for population density, because the provision of education is higher in denser places. The same reasoning holds for health. However, all these effects do not appear strong enough to lead to an automatic stabilization of the African population.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement IDED "Inheritance, Demographics, and Economic Development"

References

Akpandjar, George, Conrad Puozaa, and Peter Quartey. 2018. "Explaining fertility variation in rural communities: The role of electricity in Ghana." *Economies* 6 (07): 40.

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review* 101 (5): 2003–41 (August).

Ashraf, Quamrul, Oded Galor, and Marc Klemp. 2015. "Heterogeneity and Productivity." Working papers 2015–4, Brown University, Department of Economics.

Bailey, Martha J., and William J. Collins. 2011. "Did improvements in household technology cause the baby boom? Evidence from electrification, appliance diffusion and the Amish." *American Economic Journal: Macroeconomics* 3 (2): 189–217.

Baron, Reuben M, and David A Kenny. 1986. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of personality and social psychology* 51 (6): 1173.

Becker, Gary. 1993. "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy* 101 (3): 385–409.

Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2010. "The trade-off between fertility and education: evidence from before the demographic transition." *Journal of Economic Growth* 15 (3): 177–204.

Bloom, David E., and P. H. Reddy. 1986. "Age patterns of women at marriage, cohabitation, and first birth in India." *Demography* 23 (4): 509–523 (Nov).

Bongaarts, John, and Griffith Feeney. 1998. "On the Quantum and Tempo of Fertility." *Population and Development Review* 24 (2): 271–291.

Boucekkine, Raouf, David de la Croix, and Dominique Peeters. 2007. "Early literacy achievements, population density and the transition to modern growth." *Journal of the European Economic Association* 5 (1): 183–226 (March).

Bouscasse, Paul, Emi Nakamura, and Jón Steinsson. 2021, March. "When Did Growth Begin? New Estimates of Productivity Growth in England from 1250 to 1870." Working paper 28623, National Bureau of Economic Research.

Center for International Earth Science Information Network (CIESIN) and International Food Policy Research Institute (IFPRI) and The World Bank and Centro Internacional de Agricultura Tropical (CIAT). 2011. *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Density Grid.* NASA Socioeconomic Data and Applications Center (SEDAC). Palisades, NY.

Collier, Paul. 2017. "Africa's prospective urban transition." *Journal of Demographic Economics* 83 (1): 3–11.

De la Croix, David, and Clara Delavallade. 2018. "Religions, Fertility, and Growth in South-East Asia." *International Economic Review* 59 (2): 907–946.

De la Croix, David, and Davide Dottori. 2008. "Easter Island's collapse: a tale of a population race." *Journal of Economic Growth* 13:27–55.

De la Croix, David, and Paula E Gobbi. 2017. "Population density, fertility, and demographic convergence in developing countries." *Journal of Development Economics* 127:13–24.

Dyson, Tim. 2011. "The Role of the Demographic Transition in the Process of Urbanization." *Population and Development Review* 37:34–54.

Flückiger, Matthias, and Markus Ludwig. 2017. "Urbanization, fertility and child education in Sub-Saharan Africa." *Economics Letters* 157 (aug): 97–102.

Fowler, MS, and GD Ruxton. 2002. "Population dynamic consequences of Allee effects." *Journal of Theoretical Biology* 215 (1): 39–46.

Fox, Sean. 2017. "Mortality, migration, and rural transformation in Sub-Saharan Africa's urban transition." *Journal of Demographic Economics* 83 (1): 13–30.

Galor, Oded. 1996. "Convergence?: Inference from theoretical models." *Economic Journal* 106:1056–1069.

Galor, Oded, and Ömer Özak. 2015. "Land productivity and economic development: caloric suitability vs. agricultural suitability." June version.

Galor, Oded, and David N. Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review* 90 (4): 806–828 (September).

Galor, Oded, and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review* 106 (10): 3064–3103 (October).

Ghosh, Tilottama, Rebecca Powell, Christopher D. Elvidge, Kimberly E. Baugh, Paul C. Sutton, and Sharolyn Anderson. 2010. "Shedding light on the global distribution of economic activity." *The Open Geography Journal* 3:148–161.

Greenwood, Jeremy, Ananth Seshadri, and Guillaume Vandenbroucke. 2005. "The Baby Boom and Baby Bust." *American Economic Review* 95 (1): 183–207 (March).

Grimm, Michael, Robert Sparrow, and Luca Tasciotti. 2015. "Does electrification spur the fertility transition? Evidence from Indonesia." *Demography* 52 (5): 1773–1796.

Henderson, J. Vernon, Adam Storeygard, and David. N. Weil. 2012. "Measuring economic growth from outer space." *American Economic Review* 102 (2): 994–1028.

Jedwab, Remi, Luc Christiaensen, and Marina Gindelsky. 2017. "Demography, urbanization and development: Rural push, urban pull and ... urban push?" *Journal of Urban Economics* 98:6–16.

Kebele, Endale, Erich Striessnig, and Anne Goujon. 2021. "The relative importance of women's education on fertility desires in sub-Saharan Africa: A multilevel analysis." *Population Studies.*

Kravdal, Øystein. 2002. "Education and fertility in Sub-Saharan Africa: Individual and community effects." *Demography* 39, no. 2 (May).

Lagerlöf, Nils-Petter. 2019. "Understanding per-capita income growth in preindustrial Europe." *International Economic Review* 60 (1): 219–240.

Lee, Ronald. 1987. "Population Dynamics of Humans and Other Animals." *Demography* 24 (4): 443–465.

Lee, Ronald, and Michael Anderson. 2002. "Malthus in state space: Macro economic-demographic relations in English history, 1540 to 1870." *Journal of Population Economics* 15 (2): 195–220.

Lotka, Alfred James. 1925. *Elements of Physical Biology.* Baltimore: Williams and Wilkins.

Lutz, Wolfgang, William Butz, Samir KC, Warren Sanderson, and Sergei Scherbov. 2014. "Population growth: Peak probability." *Science* 346 (6209): 561–561.

Lutz, Wolfgang, and Ren Qiang. 2002. "Determinants of population growth." *Philosophical Transactions of the Royal Society B: Biological Sciences* 357:1197–1210.

Lutz, Wolfgang, Maria Rita Testa, and Dustin J. Penn. 2006. "Population density is a key factor in declining human fertility." *Population Environment* 289 (2): 69–81 (November).

Malthus, Thomas. 1807. *An Essay on the Principle of Population.* Fourth Edition. London: T. Bensley.

Marshall, Alfred. 1890. *Principles of Economics (8th ed.).* London: MacMilland and Co.

Montesquieu. 1749. *De l'Esprit des Loix.* Amsterdam: Chatelain.

Parienté, William. 2017. "Urbanization in Sub-Saharan Africa and the challenge of access to basic services." *Journal of Demographic Economics* 83 (1): 31–39.

Peters, Jörg, and Colin Vance. 2011. "Rural electrification and fertility - Evidence from Côte d'Ivoire." *The Journal of Development Studies* 47 (5): 753–766.

Population Division. 2019. "World Population Prospects 2019." United Nations.

Potter, Joseph. 1977. "Problems in using birth-history analysis to estimate trends in fertility." *Population Studies: A Journal of Demography* 31 (2): 335–364.

Pullum, Thomas W. 2006. "An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003." Technical Report, Methodological Reports No. 5. Calverton, Maryland: Macro International Inc.

Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney. 2002. "The global distribution of cultivable land: current patterns and sensitivity to possible climate change." *Global Ecology and Biogeography* 11 (5): 377–392.

Rindfuss, Ronald R., and Craig St. John. 1983. "Social Determinants of Age at First Birth." *Journal of Marriage and Family* 45 (3): 553–565.

Sadler, Michael Thomas. 1830. *The Law of Population – in disproof of the superfecundity of human beings, and developing the real principle of their increase.* London: John Murray.

Sato, Yasuhiro. 2007. "Economic geography, fertility and migration." *Journal of Urban Economics* 61:372–387.

Schoumaker, Bruno. 2013a. "Reconstructing Long Term Fertility Trends with Pooled Birth Histories." Paper presented at the XXVII International Population Conference, Busan (South Korea).

———. 2013b. "A Stata Module to Compute Fertility Rates and TFRs from Birth Histories: tfr2." *Demographic Research* 28 (38): 1093–144.

———. 2014. "Quality and Consistency of DHS Fertility Estimates, 1990 to 2012." Technical Report, DHS Methodological Reports No. 12. Rockville, Maryland, USA: ICF International.

Volterra, Vito. 1926. "Fluctuations in the abundance of a species considered mathematically." *Nature* 118:558–560.

Williamson, Stephen D. 2018. *Macroeconomics, 6th.* Pearson.

Wilson, Chris, and Pauline Airey. 1999. "How can a homeostatic perspective enhance demographic transition theory?" *Population Studies* 53 (2): 117–128.

Supplementary material

Homeostasis

In biology, homeostasis is defined as follows:

Homeostasis, from the Greek words for "same" and "steady," refers to any process that living elements use to actively maintain fairly stable conditions necessary for survival. The term was coined in 1930 by the physician Walter Cannon. His book, The Wisdom of the Body, describes how the human body maintains steady levels of temperature and other vital conditions such as the water, salt, sugar, protein, fat, calcium and oxygen contents of the blood. Similar processes dynamically maintain steady-state conditions in the Earth's environment.

From `https://www.scientificamerican.com/article/what-is-homeostasis/` (visited Nov 4 2021)

Homeostasis, any self-regulating process by which biological systems tend to maintain stability while adjusting to conditions that are optimal for survival. If homeostasis is successful, life continues; if unsuccessful, disaster or death ensues. The stability attained is actually a dynamic equilibrium, in which continuous change occurs yet relatively uniform conditions prevail.

From `https://www.britannica.com/science/homeostasis` (visited Nov 4 2021)

Clearly, the biological notion of homeostasis is related to the properties of dynamical systems, which are widely known in economics and in demography. Let us assume population dynamics follow a discrete scalar map:

$$P_{t+1} = \Phi(P_t), \tag{4}$$

such as the one implied by the textbook Malthusian model (Williamson 2018). Then,

Proposition 1 *given P_0, with $\Phi'(\cdot) > 0$ and $\Phi''(\cdot) < 0$,*

- *There is a globally stable steady state defined by $\bar{P} = \Phi(\bar{P})$*
- *Population Homeostasis is satisfied*
- *Population growth is negatively correlated with population density over time.*

Proof: See De la Croix and Gobbi (2017). ■

Proposition 1 is represented in Figure S1. It shows the increasing and concave map Φ, its globally stable steady state, the homeostasis property in blue and the correlation between growth (the fertility rate) and density in dark red.

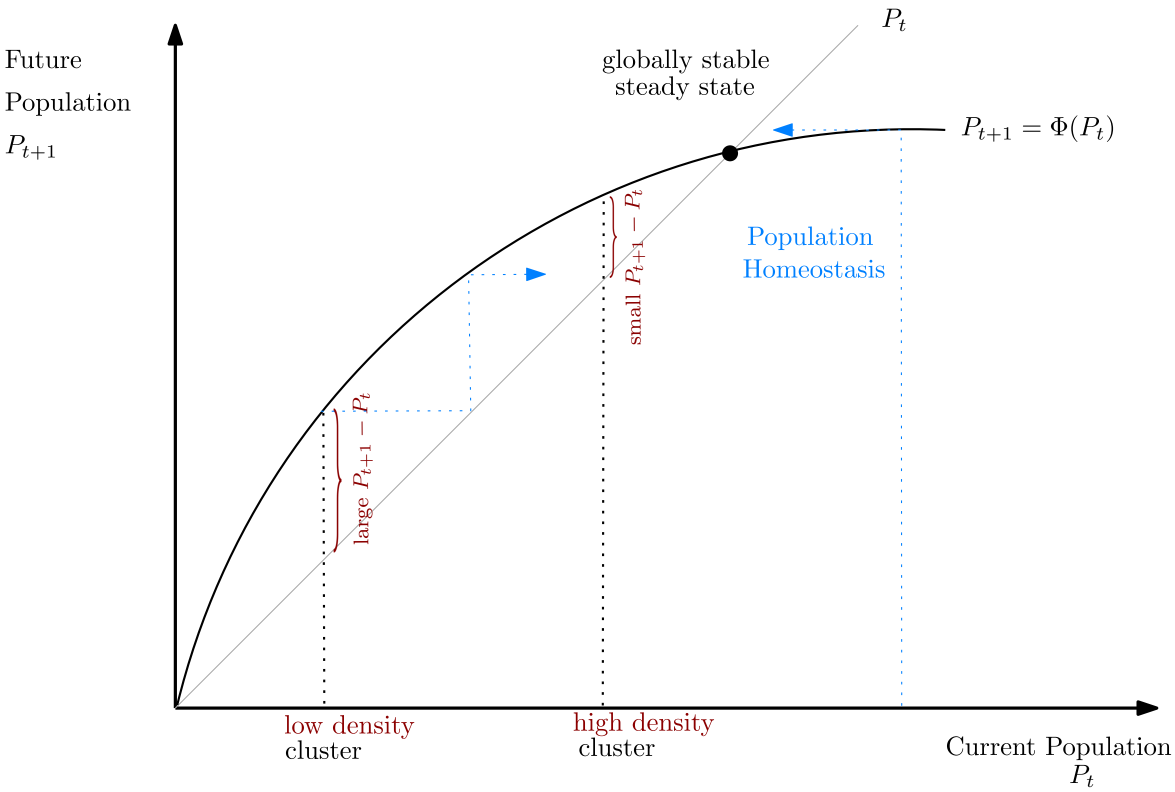


Figure S1: Population Homeostatis

To map the relationship between population density and population growth over time as a relationship across space, one can follow the standard approach provided by growth theory (Galor 1996). Consider a world consisting of different locations, each location isolated from the rest, and following the same law of motion, $\Phi(P_t)$, described in Proposition 1. If each location starts from a different initial condition P_0, then population growth is negatively correlated with population density across space.

Clearly, the assumptions of the above proposition, which ensure existence and global stability of the steady state, are only **sufficient conditions for homeostasis**. Homeostasis can be obtained under other configurations: for example, if population converges to a steady state with damped oscillations, or even if population does not converge to a steady state but rather to a limit cycle. This would be enough to ensure non-explosive behavior.

Note also that the homeostasis property is not incompatible with the idea of a demographic transition. To think about the link between the **demographic transition** and population homeostasis, one can model a simplified demographic transition as an upward shift of the function Φ, following a change in some of its parameters. There will thus be a new steady state with a larger population, and homeostasis will imply that actual population will converge to this new level, displaying a transition during which its growth rate would be high. This is admittedly a very simplistic representation of the demographic transition, which might be completed by incorporating the mortality aspect which is absent from our set-up.

The regression in Column (1) of Table 1 can be seen as an estimation of the linearized version of the dynamics (4) around its steady state:

$$P_{t+1} - \bar{P} = \Phi'(\bar{P})(P_t - \bar{P}) \Leftrightarrow P_{t+1} - P_t = \text{constant} + (\Phi'(\bar{P}) - 1)P_t$$

where the age at marriage and the age at first birth are directly related to the change in population $P_{t+1} - P_t$, P_t is population density, and $(\Phi'(\bar{P}) - 1)$ is the estimated coefficient.

In the other columns, we introduce control variables and what we call **"mediating" variables** (Baron and Kenny 1986). Control variables should be seen as reflecting location-specific parameters determining the position of the function Φ for each location. Mediating variables are variables which are affected by P_t and which affect P_{t+1}. M is a mediating variable if

$$P_{t+1} = \Psi(M(P_t), P_t) \quad = \Phi(P_t).$$

In that formulation, current population has a direct effect on future population and an indirect effect going through M. For example, education is expected to be a mediating variable: in location where density P_t is higher, schools are more accessible. Education is more developed in those areas, which reduces fertility and future population. We acknowledge however that this plausible mediating effect could be partly biased by omitted factors affecting both education and population density.

Data

Table S1 lists every sub-Saharan country for which GPS-coordinate information is available. Descriptive statistics for the data used in the analysis are provided in Table S2.

Notes on the individual recode: In a majority of DHS, eligible individuals include women of reproductive age (15-49). Some countries provide information for older women, but we did not keep these observations in the sample. We drop the observations for which the number of years of education is unknown or is higher than 30. All dates are expressed in Century Month Code (CMC). CMC is the usual way in which dates are coded in DHS. It counts time in terms of months and starts with the value 1 for January 1900. Mortality rates are computed as the difference between the number of children alive and the number of births, divided by the number of births born to a woman in the sample. Marital status is coded as either ever married (includes living with a partner, currently married, divorced, or widowed) or single (never married). Data on religion is available in almost all surveys except for those in Senegal, South Africa, and in DHS Phases VI and VII for Tanzania. Whenever this information is missing, we divide the sample into Muslims and Christians. Christians include women who belong to the Roman Catholic Church, the Evangelical Church, the Anglican Church, Protestants, Seventh-Day Adventists, Pentecostals, Methodists, the Salvation Army, Kimbanguists, the "églises réveillées", Presbyterians, the Apostolic sect, the "Iglesia ni Cristo", the Aglipayan Church (Philippine Independent Church), or those coded as "other Christians" by DHS. Missing information regarding ethnicity is a more common issue across countries (specific surveys have the subscript a in Table S1). This is the case in Angola, Burundi, Comoros, Lesotho, Liberia, Madagascar, Nigeria, Rwanda, Tanzania, Uganda, and Zimbabwe. Except for women whose ethnicity is unknown, every woman belongs to one of the 269 ethnicities documented across all countries. Ethnicities with less than 100 women in a country were not considered.

Notes on geolocation: In order to ensure the anonymity of respondents, urban clusters contain a minimum of 0 and a maximum of 2 kilometers of positional error. Rural clusters contain a minimum of 0 and a maximum of 5 kilometers of error, with a further 1% of rural clusters displaced a maximum of 10 kilometers.[11] To account for this error, we set the density in a cluster to the average density within a 2km radius around the center of this cluster if it is an urban cluster. For rural clusters, we set the radius

[11]DHS do not precisely define the urban-rural variable of the GPS dataset. In each country, they adopt a definition that can depend on the size of the population or on the breadth of infrastructures. See more at: http://dhsprogram.com/What-We-Do/GPS-Data-Collection.cfm

at 5km.[12] Finally, as the raster for the Caloric Suitability Index described above has a lower resolution than the population density raster, we impute the land productivity in each cluster from the value of the index in its given position.

[12]Due to the DHS displacement, two clusters in Uganda (DHS Phase IV) appear to be in Lake Victoria. We give each point the minimal radius so as to have positive population density. This is 13km for one cluster and 33km for the other. A similar problem was fixed for three clusters in Nigeria and one cluster in Tanzania.

Countries		DHS Phases	Years	Nobs	%
Angola	AO	VII[a]	2015-16	14,379	1.52
Benin	BJ	IV	2001	6,218	0.66
		VI	2011-12	16,522	1.74
Burkina Faso	BF	III	1998-99	6,378	0.67
		IV	2003	12,392	1.31
		VI	2010	16,115	1.70
Burundi	BU	VI[a]	2010	9,386	0.99
		VII[a]	2016-17	17,202	1.82
Cameroon	CM	IV	2004	10,599	1.12
		VI	2011	15,401	1.63
Central African Republic	CF	III	1994-95	5,884	0.62
Chad	TD	VII	2014-15	17,693	1.87
Comoros	KM	VI[a]	2012	5,074	0.54
Congo Democratic Republic	CD	V	2007	9,729	1.03
		VI	2013-14	17,399	1.84
Côte d'Ivoire	CI	III	1998-99	3,040	0.32
		VI	2011-12	9,793	1.03
Eswatini	SZ	V	2006-07	4,904	0.52
Ethiopia	ET	IV	2000	15,225	1.61
		V	2005	13,907	1.47
		VI	2011	15,789	1.67
		VII	2016	15,242	1.61
Gabon	GA	VI	2012	8,360	0.88
Ghana	GH	III	1993	4,562	0.48
		IV[d]	1998	4,841	0.51
		IV[d]	2003	5,665	0.60
		V	2008	4,802	0.51
		VII	2014	9,294	0.98
Guinea	GN	IV	1999	6,728	0.71
		V	2005	7,838	0.83
		VI	2012	9,140	0.96
Kenya	KE	IV	2003	8,168	0.86
		V	2008-09	8,421	0.89
		VII	2014	30,929	3.27
Lesotho	LS	IV[a]	2004	6,709	0.71
		VI[a]	2009	7,541	0.80
		VII[a]	2014	6,621	0.70
Liberia	LB	V[b]	2007	6,934	0.73
		VI	2013	9,229	0.97
Madagascar	MD	III[a]	1997	7,026	0.74
Madagascar		V[a]	2008-09	17,074	1.80

Countries		**DHS Phases**	**Years**	**Nobs**	%
Malawi	MW	IV[d]	2000	13,220	1.40
		IV[d]	2004	11,687	1.23
		VI	2010	22,480	2.37
		VII	2015-16	24,562	2.59
Mali	ML	III	1995-96	9,700	1.02
		IV	2001	12,767	1.35
		V	2006	14,455	1.53
		VI	2012-13	10,424	1.10
Mozambique	MZ	VI	2011	13,727	1.45
Namibia	NM	IV	2000	6,731	0.71
		V	2006-07	6,731	0.71
		VI	2013	6,731	0.71
Niger	NI	II	1992	6,503	0.69
		III	1998	7,577	0.80
Nigeria	NG	II[a]	1990	8,723	0.92
		IV	2003	7,571	0.80
		V	2008	33,332	3.52
		VI	2013	38,600	4.08
Rwanda	RW	V[a]	2005	11,177	1.18
		VI[a]	2010	13,671	1.44
		VII[a]	2014-15	13,485	1.42
Senegal	SN	II[c]	1992-93	6,310	0.67
		III[c]	1997	8,563	0.90
		IV[c]	2005	14,272	1.51
		VI[c]	2010-11	15,459	1.63
Sierra Leone	SL	V	2008	7,306	0.77
		VI	2013	16,638	1.76
South Africa	ZA	VII[c]	2016	8,510	0.90
Tanzania	TZ	IV[a]	1999	3,953	0.42
		VI[a,c]	2010	9,760	1.03
		VII[a,c]	2015-16	13,261	1.40
Togo	TG	III	1998	8,521	0.90
		VI	2013-14	9,475	1.00
Uganda	UG	IV[a]	2000-01	6,401	0.68
		V[a]	2006	7,742	0.82
		VI	2011	8,579	0.91
		VII	2016	18,231	1.92

Countries		**DHS Phases**	**Years**	**Nobs**	**%**
Zambia	ZM	V	2007	7,146	0.75
		VI	2013-14	16,347	1.73
Zimbabwe	ZW	IV[a]	1999	5,715	0.60
		V[a]	2005-06	8,852	0.93
		VI[a]	2010-11	8,853	0.93
		VII[a]	2015	9,955	1.05
Total number of individuals				947,191	100.00

[a] refers to surveys for which there is no available information on ethnicity.
[b] refers to surveys for which we use the variable dialect for ethnicity.
[c] refers to surveys for which there is no available information on religion.
[d] Ghana and Malawi had two surveys in Phase IV; we used both.

Table S1: DHS included in the analysis

	N. obs.	Mean	St. Dev.	Min	Max
From the Individual Recode					
Age (in completed years)	947,191	28.387	9.469	15	49
Education (in single years)	947,191	4.734	4.514	0	26
Partner's Education	637,604	5.090	5.054	0	25
Total number of children ever born	947,191	2.889	2.826	0	21
Total number of living children	947,191	2.444	2.359	0	16
Children's mortality rate	947,191	0.089	0.187	0	1
Motherhood rate	947,191	0.729	0.445	0	1
Marriage rate	947,191	0.738	0.440	0	1
Muslim (%)	914,920	0.322	0.467	0	1
Christian (%)	914,851	0.587	0.492	0	1
Age at first birth (in years)	690,406	19.086	3.775	8	48
Age at first birth (in months)	690,406	234.400	45.191	98	586
Age at first marriage (in years)	699,107	18.025	4.216	8	49
Age at first marriage (in months)	699,107	221.582	50.561	96	599
Moved from place of residence after 14 (%)	947,191	0.258	0.438	0	1
Ethnicity	947,164	269 categorical variables			
From the Household Recode					
Has electricity (percent)	936,401	0.301	0.459	0.000	1.000
Has a refrigerator (percent)	907,958	0.138	0.345	0.000	1.000
From CIESIN					
Population density in 1990 (pop. per km^2)	947,191	747.977	1,981.006	0.012	32,860.830
From Galor and Özak (2016)					
Caloric suitability index post 1500	947,191	10.006	2.816	0.000	17.684
From Ghosh et al. (2010)					
GDP per capita	947,191	0.002	0.010	0.00001	0.427

Table S2: Descriptive Statistics

Robustness (details)

Our benchmark specifications for comparison will be those of column (3) for the age at first birth and those of column (6) for the age at first marriage in our main result Table 1.

Omitted variables

Other variables can affect the age at first birth and the age at first marriage than those included in the benchmark model. Here, we look at the robustness of our findings when adding other possible controls: (1) dummies for the ethnicity of the respondent, (2) the individual's religion, (3) whether the household has electricity or a refrigerator, (4) the education of the spouse, and (5) the average knowledge regarding modern contraception. Through different norms and customs, ethnicity and religion can affect fertility choices in general (De la Croix and Delavallade 2018) and also decisions regarding the timing of marriage and of first birth (Bloom and Reddy 1986; Rindfuss and John 1983). Having electricity or a refrigerator can be seen as a proxy for the overall wealth of the household, which could also affect the timing of the two events studied. A more educated partner might also affect the timing of birth, while the average education of men in a female cluster may affect the age at marriage. Finally, knowledge regarding modern contraception could have an effect on the timing of marriage.

We estimate five different regression models. We do not have information regarding religion, having electricity or a refrigerator, or the education of the spouse for every woman. The sample sizes are therefore different in these specifications. The results are provided in Table S3 for age at first birth and in Table S4 for age at marriage. The column "Benchmark" repeats our estimates of columns (3) and (6), for the age at first birth and the age at first marriage respectively, from Table 1.

The sample in column (2) only includes women in countries where there is information available regarding religion. The reference group includes women whose religion is neither Islam nor Christianity. The findings from Tables S3 and S4 suggest that both the age at first birth and the age at first marriage of Christian women are lower than those of this reference group and than those of Muslim women.

Having electricity/a refrigerator negatively relates to both the age at first marriage and the age at first birth (column (3)). The sign of the relationship between electrification and fertility has been a focus of debate in the past two decades. Our findings are in line with those of Greenwood, Seshadri, and Vandenbroucke (2005) that observe and provide a theoretical explanation for the positive effect of electro-domestics on fertility in the context of the American post-war baby boom. However, other authors suggest a negative (causal) relationship, such as Bailey and Collins (2011) for the Amish in the United States, Grimm, Sparrow, and Tasciotti (2015) for Indonesia, or Akpandjar, Puozaa, and Quartey (2018) for rural Ghana. Peters and Vance (2011) find mixed evidence as they show a positive relationship in urban areas of Côte d'Ivoire and a negative sign for the rural areas.

In column (4), only married women are considered. We find no association between the education of the spouse and the age at first birth. However, the age of the spouse is positively related to a woman's age at first marriage. Knowledge regarding modern contraception is positively related to the age at first birth. We do not find a significant relationship with the age at first marriage (column 5).

	Dependent variable:					
	Probability of becoming a mother					
	Benchmark	(1)	(2)	(3)	(4)	(5)
ln(1+density)	-0.014***	-0.014***	-0.013***	-0.009***	-0.010***	-0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
education	0.052***	0.050***	0.053***	0.050***	0.049***	0.050***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$(\text{education})^2$	-0.008***	-0.007***	-0.008***	-0.007***	-0.007***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
infant mortality	0.583***	0.588***	0.583***	0.578***	0.528***	0.583***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
married	1.268***	1.266***	1.307***	1.245***		1.267***
	(0.007)	(0.007)	(0.008)	(0.007)		(0.007)
calories	0.014***	0.013***	0.014***	0.013***	0.014***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
mean mortality	0.577***	0.395***	0.576***	0.602***	0.511***	0.578***
	(0.036)	(0.036)	(0.037)	(0.037)	(0.038)	(0.036)
log(GDP per capita)	-0.002	0.001	-0.004*	0.004	-0.001	-0.004*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
mean education	-0.011***	-0.011***	-0.010***	-0.008***	-0.013***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Muslim			0.009			
			(0.008)			
Christian			-0.023***			
			(0.007)			
refrigerator				-0.073***		
				(0.005)		
electricity				-0.057***		
				(0.005)		
education of spouse					-0.000	
					(0.000)	
mean contraception						0.185***
						(0.014)
Observations	947,191	947,191	914,851	907,071	635,722	947,191
Ethnicity FE	NO	YES	NO	NO	NO	NO

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level. All specifications include survey fixed effects. Column (3) includes married women only.

Table S3: Cox Model Estimates for Age at First Birth: Robustness

	Dependent variable:					
	Probability of marrying					
	Benchmark	(1)	(2)	(3)	(4)	(5)
ln(1+density)	-0.017***	-0.026***	-0.021***	-0.015***	-0.018***	-0.018***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
education	-0.036***	-0.034***	-0.033***	-0.037***	-0.035***	-0.036***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(education)^2	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
infant mortality	0.441***	0.446***	0.444***	0.430***	0.437***	0.441***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
calories	0.007***	0.009***	0.010***	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
mean mortality	1.249***	0.972***	1.267***	1.282***	1.251***	1.248***
	(0.049)	(0.046)	(0.049)	(0.050)	(0.049)	(0.049)
log(GDP per capita)	-0.010***	-0.008**	-0.013***	-0.003	-0.011***	-0.010***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
mean education	-0.051***	-0.040***	-0.044***	-0.049***	-0.067***	-0.052***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Muslim			0.126***			
			(0.009)			
Christian			-0.043***			
			(0.008)			
refrigerator				-0.060***		
				(0.006)		
electricity				-0.022***		
				(0.006)		
education of spouse					0.016***	
					(0.002)	
mean contraception						0.015
						(0.019)
Observations	947,191	947,191	914,851	907,071	931,154	947,191
Ethnicity FE	NO	YES	NO	NO	NO	NO

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level.
All specifications include survey fixed effects.

Table S4: Cox Model Estimates for Age at First Marriage: Robustness

Selection

Density may be correlated with the age at first birth and the age at first marriage because of a selection problem: women with a lower desire for children/marriage may migrate from areas with low population density to areas with high population density. If our estimates are distorted from such a selection bias, then we cannot conclude that there is a causal relationship between population density and fertility, or a built-in stabilizer for population dynamics.

We look at whether this selection problem biases our results by removing from the sample: (1) those we know have moved (keeping those for whom information on the years lived in the place of residence is not available (NA) in the sample), and (2) everyone but those we know did not migrate (we also exclude those for whom we do not have information on migration). We consider a migrant to be a person who arrived in their place of residence when they were between age 15 and their age at the time of the interview.

The results are shown in columns (1) and (2) in Tables S5 and S6. Comparing these results to the Benchmark column, we see that although the sample size is very much reduced after removing migrants, the effect of population density on either the age at first birth or the age at marriage is still significant.

	Dependent variable:		
	Probability of becoming a mother		
	Benchmark	(1)	(2)
ln(1+density)	-0.014***	-0.012***	-0.013***
	(0.001)	(0.001)	(0.002)
education	0.052***	0.052***	0.053***
	(0.001)	(0.001)	(0.002)
$(\text{education})^2$	-0.008***	-0.008***	-0.008***
	(0.000)	(0.000)	(0.000)
infant mortality	0.583***	0.594***	0.584***
	(0.007)	(0.009)	(0.012)
married	1.268***	1.308***	1.370***
	(0.007)	(0.008)	(0.011)
calories	0.014***	0.016***	0.014***
	(0.001)	(0.001)	(0.001)
mean mortality	0.577***	0.703***	0.595***
	(0.036)	(0.042)	(0.058)
log(GDP per capita)	-0.002	0.003	0.015***
	(0.003)	(0.003)	(0.004)
mean education	-0.011***	-0.011***	-0.006***
	(0.001)	(0.001)	(0.002)
Observations	947,191	702,626	327,632

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level. All specifications include survey fixed effects. Column (1): sample of those who did not migrate, keeping those we do not know. Column (2): sample of those we are sure did not migrate.

Table S5: Cox Model Estimates for Age at First Birth – Selection

	Dependent variable:		
	Probability of marrying		
	Benchmark	(1)	(2)
ln(1+density)	-0.017***	-0.014***	-0.017***
	(0.002)	(0.002)	(0.003)
education	-0.036***	-0.046***	-0.049***
	(0.001)	(0.001)	(0.002)
$(\text{education})^2$	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
infant mortality	0.441***	0.515***	0.594***
	(0.008)	(0.009)	(0.013)
calories	0.007***	0.010***	0.005***
	(0.001)	(0.001)	(0.002)
mean mortality	1.249***	1.401***	1.239***
	(0.049)	(0.055)	(0.078)
log(GDP per capita)	-0.010***	-0.008*	-0.005
	(0.003)	(0.004)	(0.006)
mean education	-0.051***	-0.052***	-0.052***
	(0.001)	(0.002)	(0.003)
Observations	947,191	702,626	327,632

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level. All specifications include survey fixed effects. Column (1): sample of those who did not migrate, keeping those we do not know. Column (2): sample of those we are sure did not migrate.

Table S6: Cox Model Estimates for Age at First Marriage – Selection

Measurement error

Misreporting the date of birth or underreporting the number of births are common sources of error in surveys that look at birth history (Schoumaker 2014). These errors are very much linked to the low education levels of respondents (Pullum 2006), and can affect age at first birth in three ways. The first is the "Potter effect," when a woman reports that a birth occurred later than it actually did (Potter 1977). This will likely increase the age at first birth for older women. The second source of error is interviewers or respondents adjusting a date of birth in order to avoid completing the health section of the DHS questionnaire (for children younger than 5 or 3). This will cause a reduction in the average age at first birth for younger women. The last problem is the omission of earlier births, which most likely occurs with older respondents and is likely to increase the average age at first birth in a population.

Columns (1) and (2) of Table S7 respectively show the estimates of the effect of population density on the age at first birth and the age at first marriage after removing the countries classified as having "poor quality" data in Schoumaker (2014) (Table 5) from the sample.[13] Specifically, we removed Burkina Faso, Benin, Cameroon, Chad, Ethiopia, Guinea, Madagascar, Mali, Mozambique, Nigeria, Niger, Uganda, Central African Republic, Liberia, and Sierra Leone from the sample.

By doing so, we drop more than half of the observations. Comparing the results, we see that when the analysis is restricted to these countries, the overall impact of population density on the probability of first birth and marriage is stronger. The direction of the effect of the other covariates remains stable.

[13] Schoumaker (2014) explores the quality of the data using three approaches. The first consists in reconstructing trends in the total fertility rate (TFR) using a Poisson regression, and relying on one survey per country (see (Schoumaker 2013b) for details on this method). The second approach consists in pooling all the surveys conducted in the same country and then reconstructing fertility trends from the pooled dataset (Schoumaker 2013a). The third approach aims to correct birth histories by adjusting or adding births.

	Dependent variable:			
	Probability of first birth		Probability of marrying	
	Benchmark	(1)	Benchmark	(2)
ln(1+density)	-0.014***	-0.022***	-0.017***	-0.035***
	(0.001)	(0.002)	(0.002)	(0.002)
education	0.052***	0.064***	-0.036***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.002)
$(\text{education})^2$	-0.008***	-0.009***	-0.003***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)
infant mortality	0.583***	0.523***	0.441***	0.430***
	(0.007)	(0.011)	(0.008)	(0.011)
married	1.268***	1.109***		
	(0.007)	(0.008)		
calories	0.014***	0.013***	0.007***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
mean mortality	0.577***	0.664***	1.249***	1.124***
	(0.036)	(0.056)	(0.049)	(0.068)
log(GDP per capita)	-0.002	0.009***	-0.010***	-0.004
	(0.003)	(0.003)	(0.003)	(0.004)
mean education	-0.011***	-0.004***	-0.051***	-0.032***
	(0.001)	(0.001)	(0.001)	(0.002)
Ethnicity FE	NO	NO	NO	NO
Observations	947,191	487,581	947,191	487,581

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level. All specifications include survey fixed effects.

Table S7: Cox Model Estimates for Age at First birth and Age at First Marriage – Data Quality

Non-linear effects (details)

To test for non-linear effects of population density on the probability of having a first birth and the probability of marrying, we estimate a model in which we replace the log of density by ten dummy variables corresponding to the ten deciles of density. Table S8 shows that population density and these two probabilities are positively related to density for the lowest deciles of the density distribution. The positive relationship is however only significant for the probability of marrying, between the first and the second decile of the population distribution. From this analysis, we conclude that population density only has a stabilization effect once a certain threshold is reached. Moreover, we also observe that the effect of population density is in general stronger for deciles 8, 9 and 10. This suggests that the demographic transition will accelerate once population density is high enough.

	Dependent variable:							
	Probability of first birth				*Probability of marrying*			
	Bench.	(1)	Bench.	(2)	Bench.	(1)	Bench.	(2)
ln(1+density)	-0.099***		-0.014***		-0.136***		-0.018***	
	(0.001)		(0.001)		(0.001)		(0.002)	
Decile 2		0.014		0.002		0.027**		0.026***
		(0.009)		(0.008)		(0.011)		(0.010)
Decile 3		-0.015*		-0.004		-0.047***		-0.008
		(0.009)		(0.008)		(0.011)		(0.010)
Decile 4		-0.036***		0.004		-0.092***		-0.005
		(0.009)		(0.008)		(0.012)		(0.011)
Decile 5		-0.046***		0.018**		-0.118***		-0.002
		(0.009)		(0.008)		(0.012)		(0.011)
Decile 6		-0.117***		0.002		-0.207***		-0.019*
		(0.010)		(0.008)		(0.013)		(0.011)
Decile 7		-0.170***		-0.006		-0.291***		-0.038***
		(0.009)		(0.008)		(0.012)		(0.011)
Decile 8		-0.308***		-0.043***		-0.459***		-0.062***
		(0.009)		(0.008)		(0.012)		(0.011)
Decile 9		-0.460***		-0.059***		-0.647***		-0.060***
		(0.010)		(0.009)		(0.012)		(0.012)
Decile 10		-0.610***		-0.094***		-0.826***		-0.100***
		(0.010)		(0.009)		(0.012)		(0.012)
Controls	NO	NO	YES	YES	NO	NO	YES	YES
Observations	947,191	947,191	947,191	947,191	947,191	947,191	947,191	947,191

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the cluster level. All specifications include survey fixed effects. Controls include those of column (3) of Table 1.

Table S8: Cox Model Estimates for Age at First birth and Age at First Marriage with Non-linear Effect of Density

INSTITUT DE RECHERCHE
ÉCONOMIQUES ET SOCIALES

Place Montesquieu 3

1348 Louvain-la-Neuve

ISSN 1379-244X D/2021/3082/26