J Popul Econ (2012) 25:871–897
DOI 10.1007/s00148-011-0368-6

ORIGINAL PAPER

AIDS, "reversal" of the demographic transition and economic development: evidence from Africa

Sebnem Kalemli-Ozcan

Received: 4 August 2009 / Accepted: 12 January 2011 /
Published online: 15 May 2011

Abstract Using country- and region-level data, I investigate the effect of HIV/AIDS on fertility in Africa during 1985–2000. Results differ depending on the variation used and the estimation method. Between estimates that exploit cross-sectional variation suggest a positive significant effect of HIV/AIDS on fertility, whereas within estimates that are identified of off time-series variation show both positive and negative results depending on the HIV/AIDS variable used. These within estimates are insignificant in most of the specifications.

Keywords AIDS · Fertility · Growth

JEL Classification O11 · I12 · J11 · J13

1 Introduction

"Economics is judged ultimately by how well it helps us understand the world, and how well we can help improve it."—Gary Becker

Nobody doubts that AIDS is the plague of the twenty-first century. The impact of the epidemic on economic development is, on the other hand, a

Responsible editor: Junsen Zhang

S. Kalemli-Ozcan (✉)
University of Houston and NBER, Houston, TX, USA
e-mail: skalemli@mail.uh.edu

fiercely debated issue.[1] One of the main underlying reasons for this debate is the ambiguity surrounding the effect of the epidemic on the fertility behavior. So far, the empirical literature that investigates the impact of the disease on the fertility behavior has focused on a single country or on a small set of countries and delivered a widely varying set of estimates. For example, Young (2005) using household data on fertility from South Africa and relying on between cohort variation in country-level HIV infection, estimates a large negative effect of HIV prevalence on fertility. Young (2007) reaches a similar conclusion using similar survey data from 27 countries. Both of these papers use country-level HIV data and individual-level fertility data (births). On the other hand, recent studies using the newly available HIV data based on individual testing from population-based surveys find no significant effect of the disease on individual fertility behavior. Using data from 13 countries, for instance, Juhn et al. (2008) find no significant effect of the community HIV prevalence on the fertility behavior of HIV-negative women. Fortson (2009) and Fink and Linnemayr (2008) arrive at the same conclusion using the same data and employing different empirical specifications.

This paper tries to understand the reasons behind these different results in the literature. The main difference between these papers seems to be the nature of the HIV data used and the sample of countries. Thus, we ask the following question: can we get a general result concerning HIV/AIDS epidemic and fertility by using a comprehensive data set from many African countries over time and using different measures for HIV/AIDS?[2] I investigate the effect of the epidemic on the total fertility rate (TFR), using country- and region-level data both for fertility and HIV from a panel of 44 African countries during 1985–2004. The cross-time nature of the data allows me to exploit both between and within variation with the same data set. This is important since different papers in the literature use different sources of variation; cross-sectional versus time-series. I use four different indicators for HIV/AIDS, two of which are available both at the country and at the regional level.

Results differ depending on the estimation method. Between estimates based on country- and region-level data from Africa suggest a positive effect of HIV/AIDS on fertility. Within estimates that exploit time-series variation

[1] While most of the researchers find negative effects of the epidemic on economic growth, some find no effect and some even find positive effects. Bloom and Mahal (1997) run cross-country regressions of growth of GDP per capita on HIV/AIDS prevalence and find no effect. Papageorgiou and Stoytcheva (2007) find a negative significant effect of AIDS on income per worker but the effect is small. Werker et al. (2006) instrument HIV/AIDS prevalence by national circumcision rates and show that there is no effect of the epidemic on growth of the African countries. Corrigan et al. (2005) show calibration results that imply large negative effects of the epidemic on growth. The results of Lorentzen et al. (2008) imply significant long-run costs of AIDS on various outcome variables.

[2] In a related paper, Kalemli-Ozcan and Turan (2011), we focus only on South Africa and replicate Young (2005) using the exact data and the simulation model he used in Young (2005) for South Africa.

show both positive and negative effects depending on the HIV/AIDS variable used, yielding an insignificant effect in most of the specifications.

The results contrast with those of Young (2007), who find a strong negative effect of the epidemic on fertility using similar data from a subset of African countries and also exploit within-country over-time variation. Specifically, Young (2007) and this paper use the exact same HIV data, though Young (2007) has less countries. The reason for this is the fact that Young (2007) uses individual-level fertility data (births) from surveys whereas this paper uses country-level TFR complied by the World Bank to increase the number of countries. In addition, this paper uses AIDS data. Nevertheless, to reconcile the differences, I replicated Young's (2007) result using his data. I show that Young's (2007) estimates also turns out to be statistically insignificant once the standard errors are clustered at the country level. This is the appropriate clustering since the treatment is at the country level given the country-level HIV variable.

The rest of the paper is structured as follows. Section 2 discusses the literature. Examination of data is provided in section 3. Section 4 presents the econometric framework and the empirical analysis. Concluding remarks are given in section 5.

2 Possible fertility responses

Growth models in the tradition of Becker and Barro (1988) that endogenize fertility show that fertility may decrease as a response to increased life expectancy.[3] Hence, declines in mortality could lead to a quantity–quality trade-off where parents have fewer children but invest more in each child. These models suggest that fertility and mortality are positively related and behavioral response in fertility can undo and even reverse the initial rise in population size due to higher life expectancy. In the context of these models, HIV/AIDS is a shock to adult/child longevity, where fertility should respond positively. However, there are characteristics of HIV/AIDS which suggests that this formulation is overly simplified. First, field evidence suggests that there is a direct biological impact of the disease which lowers the fecundity of infected women, an effect which should be considered separately from the behavioral responses.[4,5]

[3]See Angeles (2011), Cervellati and Sunde (2007), Tamura (2006), Soares (2005), Kalemli-Ozcan (2002), Boldrin and Jones (2002), Lucas (2002), Galor and Weil (1999), and Ehrlich and Lui (1991) among many others.

[4]Many African studies, both clinic and cohort based, indicate lower fertility (around 40%) and childbearing odds among HIV-positive women. See Lewis (2007) for a recent review of these studies.

[5]It is hard to separate the biological effect from the behavioral response without data on individual HIV status. In Juhn et al. (2008), we take a first step in separating these two effects by utilizing recent rounds of the demographic health surveys (DHS) which link an individual woman's fertility outcomes to her *own* HIV status, based on testing.

Second, since it is a sexually transmitted disease, the impact on fertility can come through changes in sexual behavior, assuming individuals have accurate information about the disease. The impact of the disease on sexual behavior in Africa has proven to be a much debated topic, where the estimates so far are inconclusive.[6]

Third, regardless of changes in sexual behavior, it may be the case that infected women who know their own status and have knowledge about mother-to-child transmission would want to reduce fertility rather than give birth to infected children. Again, the field evidence on this channel is mixed.[7]

Last but not least, uninfected people, and people who think they are not at risk, might behave differently. If they know that there is a high level of mortality in their surrounding population, they might reduce their risky sexual activity which will lead to lower fertility as a by-product or they might increase their fertility along the lines of the above-cited models since the epidemic causes a rise in adult and youth mortality. Overall, there might be various responses of fertility to the HIV/AIDS epidemic, which are summarized in Table 1. In principle AIDS variable supposed to be related to the mortality channel since AIDS cases are related to AIDS mortality. HIV, given it is about infection, can work via other channels but it can also represent risk of dying. Hence, it is hard to separate out the respective channels and responses based on the use of HIV or AIDS variables.[8]

A reduction in age-specific fertility rates among HIV-positive woman due to the biological responses may serve to reduce total fertility in a high HIV prevalence country in the absence of any behavioral response from the uninfected woman. Behavioral response from the infected women (if they know their own status or have high-risk perception) might also cause a reduction in fertility. Uninfected women's fertility might also decrease due to a reduction

[6]Mwaluko et al. (2003), Bloom et al. (2000), Stoneburner and Low-Beer (2004), Lagarde et al. (1996), Lindan et al. (1991), Ng'weshemi et al. (1996), Williams et al. (2003) and Caldwell et al. (1999) all find no change or very small change in sexual behavior. Oster (2005), using DHS data on sexual behavior from a subset of African countries finds that sexual behavior changed relatively little since the onset of the epidemic. Other researchers finds some evidence of risky behavior reductions in Zambia and Zimbabwe such as reductions in multiple partners; see Cheluget et al. (2006) and Fylkesnes et al. (2001).

[7]Temmerman et al. (1990) find that in Nairobi a single session of counseling—which is common in most African countries—has no effect on the subsequent reproductive behavior of HIV-positive women. Allen et al. (1993) using cohort data from Kigali, Rwanda, find that in the first 2 years of follow-up after HIV testing, HIV-negative women were more likely to become pregnant than HIV-positive women. However, among HIV-positive women, those with no children were more likely to become pregnant than those with children and married women are more likely to become pregnant than unmarried women. The desire to have children among HIV-positive women altogether was 45%. On the other hand, Noel-Miller (2003) using panel data from Malawi shows that women who have higher subjective HIV risk perceptions for themselves were less likely to have children.

[8]There might also be differences across countries as far as the links between AIDS and mortality and HIV and AIDS are concerned. We use country-specific time trends to partially account for this. A better approach will be country-year dummies which cannot be used given the fact that this is the exact variation we exploit.

Table 1 Possible fertility responses (assuming individuals have a basic level of knowledge about the disease)

	HIV − women	HIV + women
Behavioral response		
Know own status and risk perception (low)	− or +	NA
Know own status and risk perception (high)	−	− (maybe +)
Do not know own status and risk perception (low)	− or +	− or +
Do not know own status and risk perception (high)	− or +	− or +
Biological response	0	−

in risky sexual behavior. Put it differently, for fertility to increase as a result of the epidemic any positive behavioral response of uninfected women have to overcome the negative biological and behavioral responses.

3 Data: sources and issues

3.1 Country level

Fertility I use country-level data on TFR both from the World Bank and World Development Indicators (WDI) and from DHS. TFR is the sum of age-specific fertility rates and hence it is an approximation for the average lifetime fertility of women.

Since the TFR measure differs from the "number of births" measure used by Young (2007), it deserves a little bit more discussion. TFR is the average number of births that women in the sample would have by the time they reach age 50 if they were to give birth at the current age-specific fertility rates. It is the sum of the age-specific fertility rates multiplied by the number of age/age group categories. TFR formula is as follows:

$$\mathrm{TFR} = k * \sum_{a=1}^{k} \frac{\mathrm{Births}_a}{\mathrm{Number\ \ of\ \ Women}_a} \tag{1}$$

where a is age/age group and k is the number of age/age group categories.

Using total fertility rate has several advantages. First, it is insensitive to changes in the age distribution. HIV/AIDS has differential effect on different age groups. Prime age women (15–30) are disproportionately affected by the disease, which changes the age distribution in the sample extensively. Since, TFR is calculated using age-specific fertility rates, it is a better fertility indicator than "number of births" measure used by Young (2007). Second HIV/AIDS may lead to delayed fertility. Therefore, examining the effect of HIV/AIDS on fertility across cohorts using "number of births" may lead to biased estimations as recent cohorts have not completed their fertility. As a stock variable and completed fertility indicator TFR better demonstrates the changes in fertility in the sample.

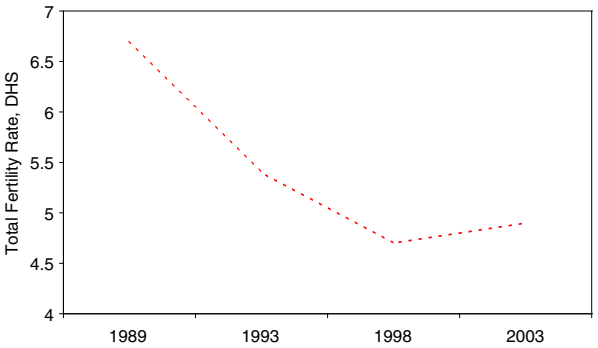


Fig. 1 Fertility in Kenya and demographic health surveys: 1989, 1993, 1998, and 2003

On the other hand, there are some disadvantages of using TFR. First, since by definition women are assumed to give birth at the same age-specific fertility rates, TFR can be relatively slow in detecting changes in fertility across cohorts. Second, the World Bank data includes some projections and interpolations, which may lead biased fertility estimations. The World Bank uses United Nations (UN) World Population Prospects for every 2 years and update the UN data with the latest survey data such as DHS or country census. UN data comes from the countries vital registration system. However, given the weakness of these systems in Africa, WB data has to do some interpolations. Hence, it is also important to use DHS data in spite of the fact that this is survey data and will not be available on continuous time basis. Given the fact that WB data has been used extensively in other work, we use both WB and DHS data to get a comprehensive picture.

WB data on TFR are available for 44 countries and 10 years between 1985 and 2004. DHS data on "number of births per woman ages 15–49" are available for 34 countries, where most countries has only one or two surveys. Only three out of 34 countries have four surveys, ten countries have three surveys, ten countries have two surveys and the remaining 11 countries have one survey. Survey years fall between 1986 and 2004.[9]

Figure 1 plots TFR from DHS for Kenya, a high prevalence country, and shows that after more than a decade of rapid decline, the total fertility rate actually increased starting in the late 1990s. Westoff and Cross (2006) find the increase in fertility in Kenya is most pronounced for the least educated group of women. They also find a significant increase in the percentage of women who report wanting more children for each age group.[10] Similar data from ten other countries show either an uptick for fertility, such as in Nigeria

[9]I also use data on desired fertility rate per woman ages 15–49, available for 34 countries, from DHS. Details of the variables and a full list of countries and survey years are provided in the Appendix.

[10]See also Durevall and Lindskog (2011).

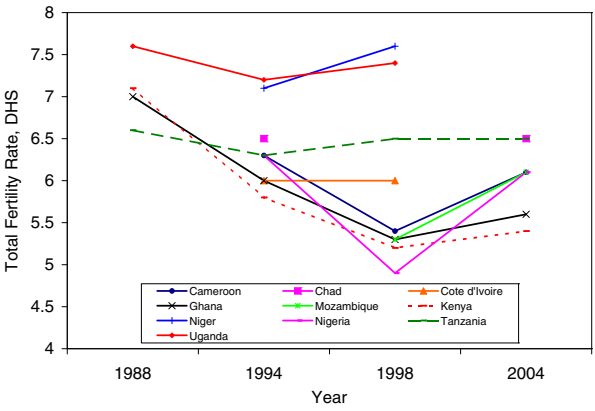


Fig. 2 Fertility in Africa and various demographic health surveys

and Mozambique, or a stall in fertility transition, such as Uganda and Cote D'Ivoire, as shown in Fig. 2.[11]

HIV/AIDS I use four different indicators for HIV/AIDS at the country level, none is perfect and all have different problems. For AIDS, I use data that come from UNAIDS/WHO, Epidemiological Fact Sheets (2003). These are the number of reported AIDS cases available for each country in every year between 1985 and 2004. I multiply the number of reported cases by 100,000 and divide by the country's population in each year, to obtain rate per 100,000 per country per year. According to UNAIDS, AIDS case reports come from surveillance systems of varying quality. Reporting rates vary substantially from country to country and low reporting rates are common in developing countries due to weaknesses in the health care systems. Hence there can be systematic biases such as in countries with worse medical institutions (which is probably correlated with other country characteristics) underreporting will be worse. AIDS case reporting provides information on transmission patterns and levels of infection approximately 5–10 years in the past, limiting its usefulness for monitoring recent HIV infections. Despite these caveats, AIDS case reporting is useful in estimating the burden of HIV-related mortality.

For HIV, I use three different indicators. First, I use data on HIV prevalence rates among pregnant women that are from the US Census Bureau, HIV Surveillance Database (2005). UNAIDS/WHO also provides similar data. This is the indicator that is used by most researchers. Both US Census and UNAIDS databases collect regional estimates of HIV/AIDS prevalence since the early 1980s. The main indicator for the epidemic is the percent HIV-1 incidence among pregnant women for each country and year. However, these estimates are in general very high. Representativeness of these estimates for the general population is also debatable since they are based on pregnant

[11] Each countries survey year is on or around the dates shown on the *x*-axis.

women and high-risk groups, which in turn is the main reason for these inflated estimates.[12] More recently, DHS started providing results from population-based HIV testing. These new estimates are much lower than the UNAIDS and US Census estimates.[13] The new population-based DHS estimates are only available for a limited set of countries for their latest survey though and hence do not provide enough information about variation over time, as the HIV estimates from the Surveillance Database. On the other hand, the time-series variation in these prevalence rates from the Surveillance Database of US Census and UNAIDS is far from perfect. UNAIDS (2006) notes that it is not possible to use previous reports to compare prevalence over time. Using the US Census HIV surveillance database suggests that HIV rates are flat or falling over the 1990s in virtually all countries in Africa, which seems inconsistent with the casual observation. A close inspection of these estimates shows that there is considerable year-to-year variation which calls into question the reliability of the time variation in these data. It has been suggested by many that selection of locations that the estimates are collected are changing over time. Based on these problems, Oster (2011) develops a methodology to estimate HIV prevalence over time from mortality data. To avoid the problem of lack of official mortality statistics for Africa she takes advantage of sibling mortality histories in the DHS. She has HIV estimates only for nine countries since mid-1980s though. I use her estimates as a second indicator for HIV.

As a third indicator for HIV, I use the *projected* HIV from the US Census Bureau, International Programs Center. The International Programs Center uses Estimation and Projection Package (EPP) from WHO/UNAIDS to project adult HIV prevalence among 15–49 years old from US Census Surveillance data between 1985 and 2004. While EPP can be used in all countries with sufficient surveillance data, it is specifically recommended for countries with "generalized epidemics." Generalized epidemics are those that have broken out into the general population or consistent HIV prevalence at over 1% in low-risk individuals. The proxy for low-risk individuals is women attending antenatal clinics. Thus, the inputs to EPP in countries with generalized epidemics are the same surveillance data on HIV prevalence among pregnant women. EPP estimates the trends over time of HIV prevalence by fitting an epidemiological model to data from urban and rural sites.[14] Although EPP model fits a somewhat flexible curve to a not so long time series, the modeling is still an issue of concern given the dynamic nature of the epidemic.

[12] See Timberg (2006) and McNeill (2007).

[13] See Juhn et al. (2008) for a comparison of the various estimates.

[14] It chooses a set minimizing least squares and projects future course based on fitted parameters, such as a parameter for the start year of the epidemic; one for the force of infection (how explosive the epidemic is in its initial stage); one for the fraction of new entrants to the population going into to the at-risk category (a parameter largely determines where the epidemic levels off); and one for the recruitment (a high value means people are brought into the at-risk population as people die of HIV, thus helping to sustain the epidemic at a higher level).

The correlation between different indicators is around 80% on average. The indicators in general suffer from different biases. Classical measurement error is one but there can also be other errors that are not classical. For example, since most of the indicators are based on estimates from antenatal clinics, the measurement error might be correlated with the population attending the clinics, which itself might be correlated with fertility. In case of AIDS the bias is in general downwards since estimated mortality is almost always lower than what is should be.[15]

Other controls All other controls such as female schooling, child mortality, contraception, and GDP per capita are taken from World Bank, WDI, and from DHS. The details of these data are provided in the Appendix.

3.2 Regional level

Fertility I use data on regional total fertility rates from DHS. They are available for 71 regions from 14 countries, whose surveys years fall between 1988 and 2004. A full list of regions is provided in the Appendix.

HIV/AIDS The data for regional HIV rates come from US Census Bureau, HIV Surveillance Database (2005) and available for 40 regions from 13 countries between 1985 and 1990. The overlap between the regional fertility rates and HIV rates give us 32 regions from 12 countries.

4 Econometric framework and empirical analysis

4.1 Framework and identification

Theoretical models of the demand for fertility have the following empirical predictions as summarized in Schultz (1997): (1) increased education of women raises the cost of childbearing and reduces fertility; (2) reduced child mortality, assuming the demand for surviving children is price inelastic, is associated with a decline in fertility; and (3) increased income per capita increases demand for children since they are normal goods. Thus, I control for these determinants, that are shown to be significant in the other empirical studies. Other potential determinants of fertility such as urbanization, average age of population, dependency ratios are all found to be insignificant in the previous studies and

[15] I also use data on *perceptions*, specifically the variable "know someone died of AIDS." The data on the percent female who know someone personally who has the virus that causes AIDS or has died of AIDS are from DHS. This is the ideal measure for the purpose of this paper however since this question has only been asked in the most recent surveys the data are available only for 22 countries whose survey years fall between 1993 and 2004. The results with this measure are available upon request.

hence I do not control for these variables in the main analysis.[16] I estimate Ordinary Least Squares (OLS) regressions of the following form, using both country and regional-level data:

$$\text{TFR}_i = \alpha + \beta \text{HIV/AIDS}_i + \mathbf{X}_i'\gamma + \epsilon_i, \tag{2}$$

where TFR_i is the total fertility rate for country i, $\text{HIV/AID}S_i$ is the indicator for HIV or AIDS for country i, $\mathbf{X}_i$ is a vector of other covariates, and ϵ_i is a random error term.[17] The coefficient of interest is β, the effect of the epidemic on fertility. Recall that four different indicators for HIV/AIDS are used: I use AIDS cases per 100,000 per country per year from UNAIDS. I will call this variable "AIDS." Next, I use the HIV prevalence rates among pregnant women that are from the US Census Bureau. I will call this "HIV." I also use Oster (2005) estimates, which I will call "HIV-Oster." Finally, I use the projections of the US Census Bureau, which I will call "HIV-EPP."

Notice that the regression presented in Eq. 2 only exploits variation *between* countries, using averaged data over time, i.e., it is a "between regression." A "within regression" framework to identify the parameters using only *within* country variation over time is preferable since this framework controls unobserved country heterogeneity. However, as summarized above the information on the time variation is noisy hence I am hesitant to rely solely on within-country time-variation by using first differences or country fixed effects, which further exacerbates the measurement error. I will present results for both frameworks to get a better insight.

To obtain the "within" estimates, I run a panel regression both with country and time fixed effects. I also run the same regressions with a general time trend and country-specific time trends. The "within regressions" are of the form:

$$\text{TFR}_{it} = \mu_i + \lambda_t + \psi \text{HIV/AIDS}_{it} + \mathbf{X}_{it}'\theta + \varepsilon_{it}, \tag{3}$$

where TFR_{it} is the total fertility rate for country i at time t, μ_i is the country fixed effect, λ_t is the time fixed effect, HIV/AIDS_{it} is one of the four indicators for HIV/AIDS, $\mathbf{X}_{it}$ is a vector of other covariates, and ε_i is a random error term.

The econometric framework presented in Eqs. 2 and 3 posits an endogeneity problem since HIV/AIDS is related to sexual behavior and marriage markets, both of which are independently related to fertility. Areas with initially higher levels of sexual behavior will have higher HIV rates and they may also have higher rates of fertility. Also, there are compelling reasons to believe that HIV infection is higher in areas with greater population density and economic activity. Then, country-level HIV rates suffer from an omitted variables bias since countries that are the most economically active may have both higher infection rates and lower fertility, the latter being due to possibly the higher

[16] Controlling for these variables in the robustness analysis yielded same results. See also Schultz (1997).

[17] This regressions is also run at the regional level with country dummies included, i.e., for region r: $\text{TFR}_r = \alpha_i + \beta \text{HIV/AIDS}_r + \mathbf{X}_r'\gamma + \epsilon_r$, where α_i is the country dummy.

cost of women's time. Failing to control any variable that is negatively correlated with the epidemic such as female education will cause a downward bias. There might also be a bias due to simultaneity that are not captured by the fixed effects.

The conditioning variables should take care of the large part of the effect of the differential development levels. The "within regressions" are immune to the unobservable factors that are time-invariant such as religion, climate and culture. However, individuals may start taking less risks as a result of the epidemic over time or across places, which will bias not only "between" but also "within" estimates. In the case of the "within" estimates the bias works against finding a positive effect of the disease on the fertility behavior though. If people start taking less risks (more condoms, fewer partners, or abstaining) because of HIV/AIDS then fertility will decrease as a by-product, and hence a negative relation between fertility and HIV/AIDS will be the result. This would be true assuming that despite changes in sexual activity HIV rates remain high. Ultimately, it is plausible that, societies which lower their level of risky sexual activity are likely to experience declines in HIV rates and in fertility levels. I do not expect this ultimate effect to be dominant for the time period that this paper is concerned with.[18] For the "between" regressions, I will undertake a falsification exercise that investigates the relationship between pre-AIDS fertility and current HIV.[19]

4.2 Descriptive statistics

Table 2 shows the mean, maximum, minimum, and standard deviation of the dependent and independent variables. Fertility rates vary from two to eight children with a mean of six children. For AIDS, the most affected country has prevalence that is 160 times higher than that of the least affected country. The difference in the HIV prevalence between the highest and lowest prevalence country is 250 times in UNAIDS and US Census data but only ten times in the Oster (2005) estimates. GDP per capita moves between $100 and $6,000. The remaining variables also show extensive variation.

4.3 AIDS, HIV, and fertility: between regressions

Table 3 reports the results of the OLS estimation of Eq. 2. Columns (1)–(3) of Table 3 uses the average values of dependent and independent variables over 1985–2000 and show that the first two indicators of the epidemic, namely, AIDS and HIV, are positively significant at 1% and at 5% level, respectively, whereas the other indicator, i.e., the HIV-EPP is not statistically significant.

[18] If sexual behavior declines for some other reason than HIV/AIDS, then this will lead a positive association between fertility and the epidemic since both will decline as a result. One cannot rule this out.

[19] In a previous version of the paper, I also undertook an IV exercise, which yielded similar results.

Table 2 Descriptive statistics

	No. of countries	Mean	SD	Max	Min
Total fertility rate, WB[a]	44	5.71	1.22	8.06	2.08
Total fertility rate, DHS[a]	34	6.07	0.87	7.40	3.90
AIDS (per 100,000)[b]	42	22.38	32.84	162.16	0.02
HIV[c]	44	0.08	0.07	0.25	0.001
HIV-EPP[d]	38	0.06	0.05	0.21	0.004
Secondary school for female[e] (%)	38	24.55	23.62	112.82	4.24
GDP per capita[f] (PPP 1996 $s)	44	798.24	1,217.78	6,168.33	95.92
Infant mortality[g] (per 1,000)	44	100.30	36.06	176.75	14.99
Mortality under 5[h] (per 1,000)	44	159.11	63.80	295.76	17.60

Notes: All variables are averaged over 1985–2000 and 44 countries depending on the availability. See Appendix for more information on the variables

[a]The sum of age-specific fertility rates (number of children that a woman would have if she lived through all of her childbearing years and experienced the current age-specific fertility rates at each age)—from World Bank (WB) and World Development Indicators (WDI) and from demographic health surveys (DHS), www.measuredhs.com, MEASURE DHS, Macro International Inc., respectively. The survey years for the data from DHS fall between 1986 and 2004. Appendix reports the survey years for each country

[b]The number of officially reported AIDS cases per 100,000 per country per year, calculated as multiplying the officially reported AIDS cases by 100,000 and dividing by population—from WHO/UNAIDS, Epidemiological Fact Sheets

[c]Percent HIV-1 sero-prevalence infection rate among pregnant women attending antenatal clinics—from US Census Bureau, HIV Surveillance Database

[d]Estimated national HIV prevalence among 15–49-year olds calculated by fitting an epidemiological model to data (Estimation and Projection Package-EPP) from urban and rural surveillance sites; from US Census Bureau, International Programs Center

[e]The gross enrollment rates from WDI

[f]The Gross Domestic Product (PPP $1996) divided by population—from WDI

[g]The infant mortality rate per 1,000 births—from WDI

[h]The of age 5 and under mortality per 1,000 births—from WDI

Using both indicators AIDS and HIV in a horse race, leads a positive significant coefficient on AIDS (0.23 with a standard error of 0.09) and a negative insignificant one on HIV. Female schooling measured as secondary school enrollment is negative and significant at 1% level, while GDP per capita is insignificant, a result which is probably due to the high correlation between GDP per capita and female schooling.[20] Another important control is infant and child mortality, which is positive and significant at 1% level.

To visually test for outliers, Fig. 3 shows the partial correlation plot for the regression shown in column (1), hence the slope of the solid blue line is 0.14. Note that the reason to use 33 or 35 countries in the regressions as opposed to 44 countries is partly driven by data availability for the controls but also driven by the existence of the outliers. Using other countries lead to partial correlation plots where the relationship is defines by outliers as opposed to the

[20]Using other measures of female schooling yield similar results.

Table 3 AIDS, HIV, and fertility: between regressions

Source for TFR	Dependent variable: total fertility rate (TFR)					
	WB (1)	WB (2)	WB (3)	DHS (4)	DHS (5)	DHS (6)
Log AIDS	0.14[a]	–	–	0.20[a]	–	–
	(0.05)	–	–	(0.08)	–	–
Log HIV	–	0.14[b]	–	–	0.20[a]	–
	–	(0.07)	–	–	(0.09)	–
Log HIV-EPP	–	–	–0.04	–	–	–
	–	–	(0.11)	–	–	–
Know someone died of AIDS	–	–	–	–	–	0.20[a]
	–	–	–	–	–	(0.07)
Female schooling	–0.02[a]	–0.02[a]	–0.02[a]	–0.02[a]	–0.03[a]	–0.02[a]
	(0.006)	(0.006)	(0.006)	(0.009)	(0.01)	(0.009)
Log GDP per capita	–0.11	–0.10	–0.11	–0.01	–0.12	–0.13
	(0.06)	(0.08)	(0.07)	(0.17)	(0.20)	(0.19)
Infant mortality	0.02[a]	0.02[a]	0.01[a]	0.01[a]	0.01[a]	0.01[a]
	(0.003)	(0.003)	(0.003)	(0.004)	(0.01)	(0.007)
R^2	0.84	0.88	0.80	0.63	0.61	0.67
Observations	33	35	30	26	26	22

Notes: Robust standard errors (white correction) are in parentheses. The between regressions report the results using country averages depending on availability, and including a constant

[a]1% significance

[b]5% significance

[c]10% significance

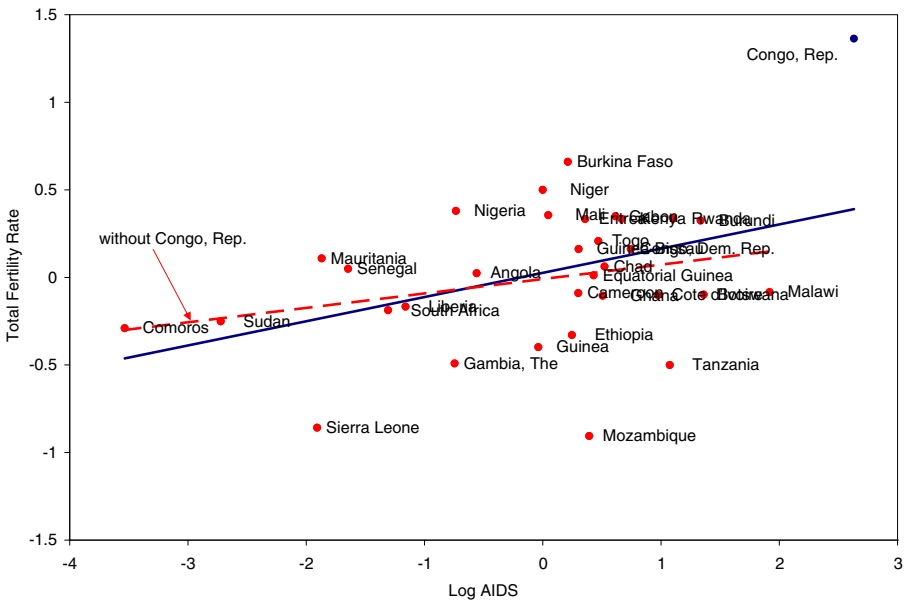


Fig. 3 Partial correlation plot for AIDS and fertility

Table 4 AIDS, HIV, and fertility: falsification exercise

	Dependent variable: total fertility rate in 1980s	
	(1)	(2)
Log AIDS	0.01	–
	(0.06)	–
Log HIV	–	0.08
	–	(0.10)
Controls	Yes	Yes
R^2	0.70	0.72
Observations	32	33

one shown in Fig. 3. If I omit Congo, Rep. the coefficient goes down to 0.10 but stays statistically significant at 5% level as shown by the dashed red line.

Columns (4)–(6) shows similar regressions using data on fertility from DHS. The fertility observations are averaged over the survey years, which change between 1987 and 2004 and from country to country.[21] The point estimates for HIV and AIDS are larger and significant at 1% level for fertility rate. The coefficient estimates for HIV-EPP are insignificant as before and hence not reported.[22]

The indicators of HIV/AIDS are used in logs following Oster (2011). Although using the log of HIV/AIDS makes the quantitative interpretation harder, it has several econometric advantages such as dampening the outliers and making the estimated coefficient immune to the scale effect due to underreporting, assuming underreporting is similar across countries. There might be a concern in using the log form though since log specification in principle compares the countries that have any AIDS to those that do not. I would argue that this is not a serious concern in the case of Africa. First of all due to averaging over time, I do not have any zeros in HIV/AIDS; the only zeros for the initial years of epidemic for few countries are averaged out. Second of all, the sample I am using are composed of countries that are classified as "generalized epidemic countries" with the exception of Comoros, Madagascar, Mauritania, and Sudan. The results are robust to excluding these four countries. The results are also robust to, even stronger, using the non-logged proxies for HIV/AIDS and available from the author upon request.

To deal with the possibility that unobserved country heterogeneity are driving the results, I will undertake a falsification exercise. Table 4 represents the results from a regression, where I regress fertility rates from 1980s, on the current HIV/AIDS, averaged over 1995–2004 and the other controls. Fertility in 1980s is the average of rates in 1980, 1982, and 1987. There is no statistically significant relationship between current HIV/AIDS and fertility in 1980s as shown in columns (1) and (2) and further in Fig. 4. The 95% confidence interval implied by the estimate and the standard error in column (1) does

[21] See Appendix for details on survey years.

[22] I also used desired fertility from DHS obtaining very similar results and hence I do not report them but they are available upon request.

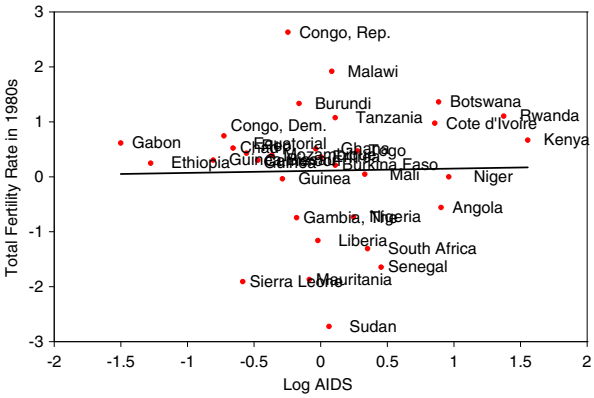


Fig. 4 Falsification plot for AIDS and fertility

not include the estimates from Table 3, however this is not the case for the column (2). Thus, this exercise suggests that time-invariant unobserved country heterogeneity is not driving the results, at least in the case of AIDS indicator.

4.4 Country-level HIV and fertility: within regressions

Table 5 reports results of the OLS estimation of Eq. 3 using country-level data.[23] Standard errors are clustered by country to deal with the possible serial correlation among residuals.[24] For each of the three indicator of HIV/AIDS both pooled regression results with a common time trend (that captures the declining trend of fertility in the absence of HIV/AIDS) and "within" regression results with both country and time fixed effects are shown in columns (1)–(6).[25] AIDS is positively significant both in the pooled and in the fixed effects regressions, HIV is only significant in the pooled regressions, and as before HIV-EPP is not significant. All other control variables yield similar results as before.

I also use country-specific time trends as shown in the last three columns of Table 5. The first two indicators gave insignificant results, however the third indicator, that is HIV-EPP yields a negative significant result. Recall that this variable never turned out to be statistically significant up until this specification. It is hard to interpret this finding but since this variable is composed of built-in trends, accounting for those trends explicitly makes HIV-EPP and country-specific trends highly correlated. To investigate this further,

[23]To deal with zero HIV/AIDS we pursued two different strategies where we dropped the zero observations and we use the HIV as log(1 + HIV). Both of these strategies yielded similar results.
[24]I also perform weighted least squares (WLS) panel regressions; where all observations are weighted in the second step with the inverse of the estimated standard deviations from the first step. Weighting by country's population or log population yields similar results.
[25]I also experimented with a common non-linear quadratic and cubic trend obtaining similar results.

Table 5 AIDS, HIV, and fertility: pooled and within regressions

	Dependent variable: total fertility rate (TFR)								
	Pooled (1)	Pooled (2)	Pooled (3)	Within (4)	Within (5)	Within (6)	Within (8)	Within (9)	Within (10)
Log AIDS	0.09[b]	–	–	0.09[b]	–	–	–0.06	–	–
	(0.04)	–	–	(0.04)	–	–	(0.05)	–	–
Log HIV	–	0.12[b]	–	–	–0.01	–	–	–0.08	–
	–	(0.06)	–	–	(0.05)	–	–	(0.05)	–
Log HIV-EPP	–	–	–0.01	–	–	0.10	–	–	–0.15[a]
	–	–	(0.07)	–	–	(0.10)	–	–	(0.05)
Female schooling	–0.02[a]	–0.03[a]	–0.02[a]	–0.004[a]	–0.001[a]	–0.002[a]	–0.03[a]	–0.003[a]	–0.03[a]
	(0.005)	(0.005)	(0.005)	(0.01)	(0.01)	(0.01)	(0.02)	(0.008)	(0.01)
Log GDP per capita	–0.10	–0.12	–0.12	–0.27	–0.29	–0.43	–0.27	–0.29	–0.13
	(0.06)	(0.08)	(0.07)	(0.31)	(0.27)	(0.29)	(0.27)	(0.27)	(0.24)
Infant mortality	0.01[a]	0.01[a]	0.01[a]	0.003[a]	0.006[a]	0.004[a]	0.007[a]	0.01[a]	0.005[a]
	(0.002)	(0.032)	(0.002)	(0.01)	(0.004)	(0.01)	(0.007)	(0.005)	(0.004)
Common trend	–0.05[a]	–0.04[a]	–0.03[a]	–	–	–	–	–	–
	(0.02)	(0.01)	(0.01)	–	–	–	–	–	–
Country effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	No	No	No	Yes	Yes	Yes	No	No	No
Country trends	No	No	No	No	No	No	Yes	Yes	Yes
R^2	0.76	0.82	0.76	0.87	0.87	0.87	0.95	0.95	0.95
Observations	111	153	139	111	153	139	111	153	139
Countries	33	35	30	33	35	30	33	35	30

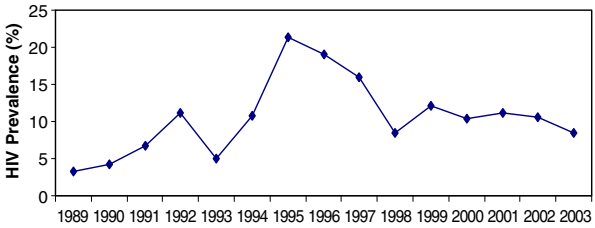
Notes: Robust standard errors (white correction; clustered on countries) are in parentheses. The within regressions report results using country fixed effects. See Table 2 for the detailed explanation of the variables

[a]1% significance

[b]5% significance

[c]10% significance

Fig. 5 Ethopia HIV



I run a regression similar to a diff-in-diff specification such as I regress change in fertility from 1990 to 2004 on the change in HIV/AIDS from 1990 to 2004, obtaining a positive significant result for AIDS and a negative insignificant result both for HIV and HIV-EPP.

Overall, the mixed results can be explained by the use of within-country time-variation as the main identifier.[26] Although this is the preferred strategy given the need to control country-specific time-invariant factors, the available country-level HIV data does not lend itself naturally for this estimation. Figures 5, 6, and 7 show time-series path of HIV data from three countries, Ethopia, Kenya, and Malawi, respectively, where there is a lot of noise. Exploiting noisy variation can lead to different results depending on different specifications. Of course for the HIV-EPP variable the projection package generates a smooth curve but it is not clear why this estimated HIV-EPP variable should be preferred over HIV variable or AIDS variable or vice versa. As argued before all of these measures have errors, each having pros and cons and hence one should investigate the effect of each to get a clearer picture.[27]

4.5 Regional evidence

This section uses regional data on fertility and HIV from 12 African countries. I have data on 32 regions. These are the regions with overlapping data on regional total fertility rates and HIV prevalence rates. Each country's survey year falls between 1998 and 2004. If there is more than one survey year during this period, than the data on the total fertility rate are averaged. I regress the regional total fertility rates on the logarithm of regional HIV prevalence rates among pregnant women averaged over 1990–1995, including country dummies. Unfortunately, the other controls are not available at the regional level.

Table 6 shows the results of the OLS regressions. Both columns show that results are positive and significant at 1% level. To deal with the potential serial

[26] An alternative story that explain the difference between AIDS and HIV and between and within results might the fact that AIDS is a measure of death and HIV is the current infection. We repeat the time-series specifications including lagged variables and the results stay the same. These exercises are available upon request.

[27] The results with HIV-Oster are similar and available upon request. For robustness, we have also tried many other control variables such as contraception, population age and size, and so on; all of the results remain the same and available upon request.

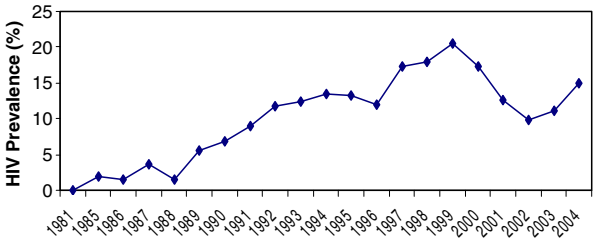


Fig. 6 Kenya HIV

correlation across residuals given the regional data, I cluster at the country level, which raises the standard errors as shown in column (2). Though the results are still significant at 1% level. I also tried a "WLS" specification, where in order to limit the influence of small regions, I weighted by the population and also alternatively by the logarithm of regional population from DHS, averaged over the survey years. Results were similar and hence not reported. The results requires some caution since there might still be heterogeneity in spite of the country effects and clustering the standard errors, due to variation in proximity to road networks and urban-rural differences.[28]

4.6 Reconciling the results with those of Young (2007)

Instead of TFR, Young (2007) uses data on individual-level fertility from DHS and the country-level HIV projections, that is "HIV-EPP" and exploits within country variation over time. For his 27-country samples, there are only two countries which have four surveys, ten countries have three surveys, eight countries have two surveys, and the remaining seven countries have only one survey. Thus the identification rests on limited observations from 20 countries with questionable quality of the time variation as argued before.

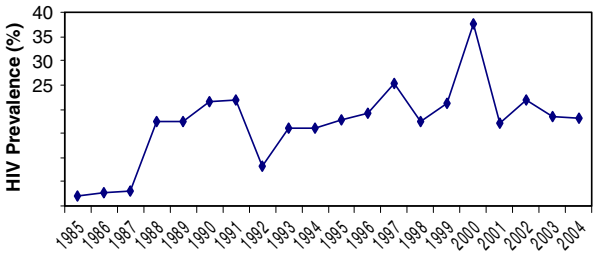
Table 7 shows the replication of Young's (2007) results. Every regression has controls for age and education and also country and year effects. Columns (1)–(4) undertake a Poisson estimation of past-year births following Young (2007) and columns (5)–(8) perform an OLS exercise. The difference between column (1) and (2) and similarly the other column pairs is the inclusion of control variables, which are marital status, urban/rural location, the number and square of born and living children, and the presence of a radio, television, refrigerator, or bicycle (each entered separately) in the household.

Columns (1) and (2) replicates regressions of Young (2007) by clustering at individual level (on case-id). Not clustering at all give very close results. Columns (1) is an exact match to Young (2007) and column (2) is very close.[29]

[28] I also run IV regressions for a smaller sub-sample. In spite of a strong first stage, the second stage regressions gave statistically insignificant results.

[29] Column (1) is an exact match to the working paper version of Young (2007). Differences might be due to different controls. Although I tried to match the controls in Young (2007), some of his specifications do not detail the set of controls used.

Fig. 7 Malawi HIV



However, when I cluster by country in columns (3) and (4), the standard errors for the HIV coefficient get large and the coefficient ceases to be significant. The same story holds for the OLS estimation as shown in columns (4)–(8). Given the fact that the treatment is at the country-time level, it is preferable to cluster either by country-time or by country, to deal with possible serial correlation. Since the autocorrelations can be positive or negative, it is possible for the non-clustered standard error to under or over-estimate the true standard error. In the case of HIV, positively serially correlated residuals lead to underestimated standard errors and hence false significance.

Peterson (2009) and Bertrand et al. (2004) report that, it had been common practice that researchers does not adjust the standard errors for possible dependence in residuals in the panel data sets. Peterson (2009) reports that 42% of the papers published in the last 5 years in finance using panel data by firm and by time does not adjust standard errors at all. He shows that the true standard error is 11 times the estimated and 81% of the time *t*-statistics are falsely significant at 1%. Bertrand et al. (2004) have drawn attention to robust standard error estimation in the context of a special fixed effect model, that is "Differences-in-Differences (DD)," where they show 65% of the time, there

Table 6 HIV and fertility: between regressions at the regional level

	Dependent variable: TFR in 1998–2004	
	(1)	(2)
Log HIV in 1990–1995	0.29[a] (0.05)	0.29[a] (0.08)
Country dummies	Yes	Yes
Cluster	Region	Country
R^2	0.79	0.79
Observations	32	32
Countries	12	12

Notes: Robust standard errors (column 1, clustered on regions; column 2, clustered on countries) are in parentheses. All regressions report results using country fixed effects. Regional TFRs are from DHS, various survey years (mean, 5.07; SD, 1.60; max, 8.7; min, 1.9). Each country's survey year falls between 1987 and 2004. The data are averaged over the survey years. Regional HIV rates (percent HIV-1 sero-prevalence among pregnant women) are from the US Census Bureau, HIV Surveillance Database (2003) (mean, 0.047; SD, 0.079; max, 0.3094; min, 0). HIV prevalence rates are averaged over 1990–1995 or used as a single year depending on the availability

[a]1% significance

Table 7 HIV and individual fertility in a panel of African countries

	Dependent variable is last year's births							
Estimation	Poisson	Poisson	Poisson	Poisson	OLS	OLS	OLS	OLS
Cluster	Individual	Individual	Country	Country	Individual	Individual	Country	Country
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Projected HIV	–1.260[a]	–0.963[a]	–1.260	–0.963	–0.218[a]	–0.234[a]	–0.218	–0.234
	(0.380)	(0.385)	(1.17)	(0.908)	(0.057)	(0.054)	(0.201)	(0.190)
Primary Education	–0.294[a]	0.071[a]	–0.294[a]	0.071[a]	0.060[a]	0.010[a]	0.060[a]	0.010[a]
	(0.054)	(0.057)	(0.005)	(0.001)	(0.009)	(0.010)	(0.001)	(0.000)
Secondary Education	–0.798[a]	0.252[a]	–0.798[a]	0.252[b]	–0.132[a]	0.029[a]	–0.132[a]	0.029[a]
	(0.091)	(0.102)	(0.005)	(0.012)	(0.010)	(0.012)	(0.001)	(0.002)
Tertiary Education	–1.190[a]	0.759[b]	–1.190[a]	0.759[a]	–0.155[a]	0.116[a]	–0.155[a]	0.116[a]
	(0.445)	(0.461)	(0.007)	(0.036)	(0.030)	(0.033)	(0.001)	(0.007)
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age^2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	403,100	350,586	403,100	350,586	403,100	350,586	403,100	350,586

Notes: Countries and survey years are Benin (1996 and 2001), Burkina Faso (1992/1993, 1998/1999, and 2003), Burundi (1987), Cameroon (1991 and 1998), Central Republic of Africa (1994/1995), Chad (1996/1997), Cote D'Ivoire (1998/1999), Ethiopia (2000), Gabon (2000), Ghana (1988, 1993, 1998, and 2003), Guinea (1999), Kenya (1989, 1993, 1998, and 2003), Liberia (1986), Malawi (1992 and 2000), Mali (1987, 1995/1996, and 2001), Mozambique (1997), Namibia (1992 and 2000), Niger (1992 and 1998), Nigeria (1990, 1999, and 2003), Rwanda (1992, 2000), Senegal (1986, 1992/1993, and 1997), South Africa (1998), Tanzania (1992, 1996, and 1999), Togo (1988 and 1998), Uganda (1988, 1995, and 2000/2001), Zambia (1992, 1996, and 2001/2002), and Zimbabwe (1988, 1994, and 1999). Other controls in the regressions are marital status (never, currently or formerly married), urban/rural location, the number and square of born and living children, and the presence of a radio, television, refrigerator, or bicycle (each entered separately) in the household. Robust standard errors (clustered as indicated) are in parentheses

[a]1% significance

[b]5% significance

is false significance because of non-clustering. Out of 92 DD papers only 36 deal with the issue. Peterson (2009) and Bertrand et al. (2004) both show using simulations that clustered standard errors adequately account for the residual dependence created by the state (or firm or country) effect and thus provide unbiased estimates.[30] Peterson (2009) argues that if there are both country and time fixed effects the best practice is to cluster at both levels or if the number of clusters is small in one dimension, like the time dimension, then use a fixed effect for that dimension and cluster on the other dimension, where more clusters are available.[31] As a result the cross-country finding of the negative significant effect of HIV on fertility by Young (2007) is not robust.

5 Conclusions

Using country- and region-level data from a panel of 44 African countries during 1985–2000, I show a positive effect of the epidemic on fertility in between-country and between-region comparisons. I find no robust effect of the disease on fertility in within-country comparisons, however. The within estimates range from positive to negative significance, depending on which country-level HIV/AIDS variable is used, yielding an insignificant effect in most of the specifications. Replicating Young (2007) also suggests an insignificant effect, once the standard errors are clustered by country.

What is the main conclusion then? Although there seems to be a robust positive effect of the disease on fertility in a cross section of countries, this effect is not there once we compare countries over time instead of to each other. The usual explanation for these type of findings is that the cross-sectional results are driven by unobserved country-level heterogeneity which is controlled through the inclusion of country fixed effects in the panel regressions. However, we also get a positive significant result in the regional estimation that also controls for country fixed effects. In addition panel estimation may not be

[30] Peterson (2009) shows this for the standard OLS regression but he reports that his results generalizes to non-linear models too. Bertrand et al. (2004) focuses on a DD model such as; $Y_{ist} = A_s + B_t + cX_{ist} + \beta I_{st} + \epsilon_{ist}$, for individual i, state s, and time t. They also show simple parametric corrections, such as fitting an AR1 process for the error structure, or non parametric corrections, such as block bootstrap, only works with large number of states/cross-sectional units. They show that clustering at state level not just at state-year cell is the best solution.

[31] Kezdi (2004) shows clustered standard errors can be too large in a fixed effects model but he also shows only clustered standard errors are unbiased irrespective of having a country effect, as also shown by Peterson (2009). Peterson (2009) also shows the generalization of the results for the GLS case. Kezdi (2004) shows that the general robust standard error estimator known as the cluster estimator is not only consistent in general but it behaves well in finite samples. His Monte Carlo simulations shows that only cluster estimator gives unbiased results even in small cross-sectional samples. He shows in a fixed effect model with short time series (as here), serial correlation in the error process and the right hand side variables induce severe bias in conventional standard errors. Clustered estimator applied to mean-differenced data is consistent and behaves well in finite sample and it does not get biased with high T or small N.

credible here given the noisy time variation of the country-level HIV/AIDS data, which explains the variety of results found in the panel specifications.

As a result, the main message of this paper is to argue that we should not use HIV/AIDS at the country level, for Africa, since doing so can generate any result. A researcher who just happened to be started with one of the country-level HIV/AIDS measures can easily find a totally different result then another researcher who uses another indicator for country-level HIV/AIDS. The current paper, by showing estimations with all the available country-level measures of the disease, for Africa, tries to make this point. Based on the country-level HIV/AIDS there is not a robust effect of the disease on fertility.

The future research should make use of the newly available HIV data based on blood-testing from DHS surveys. Indeed the results of this paper are consistent with the results from the recent micro studies that started to make use of this new HIV data based on blood-testing from DHS surveys and find no effect of the disease on the fertility behavior.[32]

Acknowledgments The author thanks Emily Oster, Rodrigo Soares, two anonymous referees, and the seminar participants at Bilkent University, Brown University, Harvard University School of Public Health, METU, and participants of the 2005 conference on "Health, Demographics and Economic Development" at Stanford University, of the 2006 AIDS Workshop at Amsterdam Institute for International Development for valuable comments and suggestions. The author also thanks Ms. Laura Heaton of the Health Studies Branch, Population Division, the US Bureau of Census for providing her the HIV estimates.

Appendix

Country-level data

Countries with survey years Benin (1996, 2001), Burkina Faso (1992/1993, 1998/1999, 2003), Burundi (1987), Cameroon (1991, 1998), Central Republic of Africa (1994/1995), Chad (1996/1997), Cote D'Ivoire (1994, 1998), Ethiopia (2000), Gabon (2000), Ghana (1988, 1993, 1998, 2003), Guinea (1999), Kenya (1989, 1993, 1998, 2003), Liberia (1986), Malawi (1992, 2000), Mali (1987, 1995/1996, 2001), Mozambique (1997), Namibia (1992, 2000), Niger (1992, 1998), Nigeria (1990, 1999, 2003), Rwanda (1992, 2000), Senegal (1986, 1992/1993, 1997), South Africa (1998), Tanzania (1992, 1996, 1999), Togo (1988, 1998), Uganda (1988, 1995, 2000/2001), Zambia (1992, 1996, 2001/2002), and Zimbabwe (1988, 1994, 1999).

The following countries have no surveys. Angola, Botswana, Comoros, Congo Democratic Republic, Congo Republic, Equatorial Guinea, Guinea-Bissau, Lesotho, Mauritania, Mauritius, Seychelles, Sierra Leone, Sudan, and Swaziland.

[32] See Juhn et al. (2008), Fink and Linnemayr (2008) and Fortson (2009).

- *AIDS*: The AIDS data come from UNAIDS/WHO, Epidemiological Fact Sheets (2003) and US Census Bureau HIV/AIDS Surveillance Database (2005). These are the number of reported AIDS cases for each country in every year and available for 44 African countries for 1985–2004. I multiply these number of reported incidents by 100,000 and divide by the country's population in each year, converting them to incidence per 100,000 per country per year.
- *Enrollment rates*: Gross school enrollment rates are from World Bank, Word Development Indicators (2006). They are available for 35 countries and years between 1985 and 2004.
- *GDP per capita*: GDP per capita (PPP 2000 $s) is from World Bank, World Development Indicators (2006).
- *HIV*: HIV prevalence rates among pregnant women are from the US Census Bureau, HIV Surveillance Database (2003). UNAIDS/WHO also provides similar data. Both Census and UNAIDS databases collect all studies and estimates of HIV/AIDS prevalence since the early 1980s. They provide information on prevalence, population and other factors and also provide regional estimates. The main indicator for the epidemic is the percent HIV-1 incidence among pregnant women for each country and year.
- *HIV-EPP*: The International Programs Center of the Census Bureau uses Estimation and Projection Package (EPP) from WHO/UNAIDS to estimate and project adult HIV prevalence among 15–49-year olds from surveillance data between 1985 and 2004. While EPP can be used in all countries with sufficient surveillance data, it is specifically recommended for countries with generalized epidemics. Generalized epidemics are those that have broken out into the general population or consistent HIV prevalence at over 1% in low-risk individuals. The proxy for low-risk individuals is women attending antenatal clinics. The input to EPP in countries with generalized epidemics is surveillance data from various sites and years showing HIV prevalence among pregnant women, as well as data from national population-based surveys. EPP estimates the trends over time of HIV prevalence by fitting an epidemiological model to data from urban and rural sites. It tests possible epidemiological parameters, chooses a set minimizing least squares and projects future course based on fitted parameters, such as a parameter for the start year of the epidemic; one for the force of infection (how explosive the epidemic is in its initial stage); one for the fraction of new entrants to the population going into to the at-risk category (a parameter largely determines where the epidemic levels off); and one for the recruitment (a high value means people are brought into the at-risk population as people die of HIV, thus helping to sustain the epidemic at a higher level).
- *Infant mortality*: Infant mortality is the rate per 1,000 live births and from World Bank, World Development Indicators (2006). The data are available for 8 years (1985, 1987, 1990, 1992, 1995, 1997, 2000, and 2004).

- *Total fertility rate*: Data on total fertility rates are from World Bank, World Development Indicators (2006) and available for 10 years (1985, 1987, 1990, 1992, 1995, 1997, 2000, 2002, 2003, and 2004), and 44 countries. DHS data on total fertility rate per woman ages 15–49 are from DHS, www.measuredhs.com, MEASURE DHS, Macro International Inc. The data are available for 34 countries whose survey years fall between 1986 and 2004.

Regional-level data

Regions

Benin	Atacora Province, Atlantique Province, Borgou Province, Mono Province, Oueme Province, and Zou Province.
Ethiopia	Addis Ababa, Dire Dawa, Gambella, and Harari.
Ghana	Accra, the Northern region, the upper East region, and the upper West region.
Lesotho	Maseru, Leribe district, Mafeteng district, Quthing district, and Mokhotlong.
Madagascar	Antananarivo, Antsiranana, Fianarantsoa, Mahajanga, Toamasina, and Toliary.
Malawi	Lilongwe, Blantyre, Mangochi, Mulanje, Mzimba, and Thyolo.
Mali	Bamako, Koulikoro, Mopti, and Sikasso.
Niger	Dosso, Maradi, Niamey, Tahoua, and Zinder.
Nigeria	North East zone, North West zone, South East zone, and South West zone.
Rwanda	Butare, Byumba, Gisenyi, Kigali, and Ruhengeri.
South Africa	Eastern Cape Province, Free State Province, Gauteng Province, Mpumalanga Province, Northern Cape Province, Northern Province, North-West Province, and Western Cape Province.
Tanzania	Dar es Salaam, Rukwa region, Arusha region, and Zanzibar area.
Togo	Kara, Plateaux, and Savanes.
Zimbabwe	Harare, Bulawayo, Manicaland, Masvingo, Mashonaland West Province, and Matabeleland South.

- *Fertility rates*: Regional fertility rates are from DHS, www.measuredhs.com, MEASURE DHS, Macro International Inc., and available for 14 countries, whose surveys years fall between 1988 and 2004.
- *HIV rates–US census*: Regional HIV data come from US Census Bureau, HIV Surveillance Database (2005) and available for 14 African countries. The data are available for 1985–1990 and also for later years for a smaller number of regions.

Individual-level data

Individual-level data are used for 27 countries from 57 Demographic Health Surveys: Benin (1996 and 2001), Burkina Faso (1992/1993, 1998/1999, and 2003), Burundi (1987), Cameroon (1991, 1998), Central Republic of Africa (1994/1995), Chad (1996/1997), Cote D'Ivoire (1994 and 1998), Ethiopia (2000), Gabon (2000), Ghana (1988, 1993, 1998, and 2003), Guinea (1999), Kenya (1989, 1993, 1998, 2003), Liberia (1986), Malawi (1992 and 2000), Mali (1987, 1995/1996, and 2001), Mozambique (1997), Namibia (1992 and 2000), Niger (1992, 1998), Nigeria (1990, 1999, and 2003), Rwanda (1992 and 2000), Senegal (1986, 1992/1993, and 1997), South Africa (1998), Tanzania (1992, 1996, and 1999), Togo (1988 and 1998), Uganda (1988, 1995, and 2000/2001), Zambia (1992, 1996, and 2001/2002), and Zimbabwe (1988, 1994, and 1999).

- *Educational attainment*: This is a categorical variable for woman's educational attainment level. Categories are "No Education", "Primary Education", "Secondary Education", "Tertiary Education" (v106).
- *Fertility:* Measured as number of births or pregnancies in last year for each woman (v209).
- *Controls*: Other control variables from are: Age (v121), year of survey (v007), presence of radio in the household (v120), presence of television in the household (v121), presence of refrigerator in the household (v122), presence of bicycle in the household (v123), urban/rural (v102), number of born children (v201), and number of living children (v201-v206-v207).

References

Allen S, Serufilira A, Gruber V, Kegeles S, Van de Perre P, Carael M, Coates TJ (1993) Pregnancy and contraception use among urban Rwandan women after HIV testing and counseling. Am J Public Health 83(5):705–710

Angeles L (2011) Demographic transitions: analyzing the effects of mortality on fertility. J Popul Econ 23(1):99–120

Becker GS, Barro RJ (1988) A reformulation of the economic theory of fertility. Q J Econ 103(1):1–25

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust difference in differences estimates? Q J Econ 119(1):249–275

Bloom D, Mahal A (1997) Does the AIDS epidemic threaten economic growth? J Econom 77(1):105–124

Bloom S, Banda C, Songolo G, Mulendema S Cunningham A, Boerma JT (2000) Looking for change in response to the AIDS epidemic: trends in AIDS knowledge and sexual behavior in Zambia, 1990 through 1998. AIDS 25:77–85

Boldrin M, Jones LE (2002) Mortality, fertility, and saving in a Malthusian economy. Rev Econ Dyn 5(4):775–814

Caldwell J, Orubuloyeb IO, Caldwell P (1999) Resistances to behavioural change to reduce HIV/AIDS infection in predominantly heterosexual epidemics in third world countries. Health Transition Center

Cervellati M, Sunde U (2007) Human capital, mortality, and fertility: a unified theory of demographic transition. IZA Working Paper 2905

Cheluget B, Baltazar G, Orege P, Ibrahim M, Marum LH, Stover J (2006) Evidence for population level declines in adult HIV prevalence in Kenya. Sex Transm Infect 82(Supplement 1):121–126

Corrigan P, Gloom G, Mendez F (2005) AIDS crisis and growth. J Dev Econ 77(1):107–124
Durevall D, Lindskog A (2011) Uncovering the impact of the HIV epidemic on fertility in Sub-Saharan Africa: the case of Malawi. J Popul Econ 24(2):629-655
Ehrlich I, Lui FT (1991) Intergenerational trade, longevity, intrafamily transfers and economic growth. J Polit Econ 99(5):1029–1059
Fink G, Linnemayr S (2008) HIV and fertility: long-term evidence from sub-Saharan Africa. Harvard University, mimeo
Fortson J (2009) HIV and fertility. Am Econ J Appl Econ 1(3):170–194
Fylkesnes K, Musonda RM, Sichone M, Ndhlovu Z, Monze M, Tembo F, Phiri S, Malamba C (2001) Declining HIV prevalence and risk behaviors in Zambia: evidence from surveillance and population-based surveys. AIDS 15(7):907–916
Galor O, Weil DN (1999) From Malthusian stagnation to modern growth. Am Econ Rev 89(2):150–154
Juhn C, Kalemli-Ozcan S, Turan B (2008) HIV and fertility in Africa: first evidence from population based surveys. NBER WP 14248
Kalemli-Ozcan S (2002) Does mortality decline promote growth? J Econ Growth 7(4):411–439
Kalemli-Ozcan S, Turan B (2011) HIV and fertility revisited. J Dev Econ (in press)
Kezdi G (2004) Robust standard error estimation in fixed-effects panel models. Hung Stat Rev 9:85–116
Lagarde E, Pison G, Enel C (1996) Knowledge, attitudes and perception of AIDS in rural Senegal: relationship to sexual behaviour and behaviour change. AIDS 10(3):327Ű-334
Lewis J (2007) The population impact of HIV on fertility in sub-Saharan Africa. AIDS 18(2):35–43
Lindan C, Allen S, Carael M, Nsengumuremyi F, Van de Perre P, Serufilira A, Tice J, Black D, Coates T, Hulley S (1991) Knowledge, attitudes, and perceived risk of AIDS among urban Rwandan women: relationship to HIV infection and behavior change. AIDS 5(8):993-Ű1002
Lorentzen P, McMillan J, Wacziarg R (2008) Death and development. J Econ Growth 13(2):81–124
Lucas RE (2002) Lectures on economic growth. Harvard University Press, Cambridge
McNeill D (2007) U.N. to say it overestimated AIDS cases by millions. New York Times
Mwaluko G, Urassa M, Isingo R, Zaba B, Boerma JT (2003) Trends in HIV and sexual behavior in a longitudinal study in a rural population in Tanzania, 1994–2000. AIDS 17:2645–2651
Ng'weshemi JZ, Boerma JT, Pool R, Barongo L, Senkoro K, Maswe M, Isingo R, Schapink D, Nnko S, Borgdorff MW (1996) Changes in male sexual behaviour in response to the AIDS epidemic: evidence from a cohort study in urban Tanzania. AIDS 10(12):1415-Ű1420
Noel-Miller CM (2003) Concern regarding the HIV/AIDS epidemic and individual childbearing, evidence from rural Malawi. Demographic Research. http://www.malawi.pop.upenn.edu/Level%203/Papers/PDF-files/noel-2003.pdf
Oster E (2005) Sexually transmitted infections, sexual behavior, and the HIV/AIDS epidemic. Q J Econ 120(2):467–515
Oster E (2011) Routes of infection: exports and HIV incidence in sub-Saharan Africa. J Eur Econ Assoc (in press)
Papageorgiou C, Stoytcheva P (2007) What do we know about the impact of AIDS on cross-country income so far? LSU, mimeo
Peterson M (2009) Estimating standard errors in finance panel data sets: comparing approaches. Rev Financ Stud 22(1):435–480
Schultz PT (1997) Demand for children in low income countries. In: Rosenzweig MR, Stark O (eds) Handbook of population and family economics. Elsevier, Amsterdam
Soares R (2005) Mortality reductions, educational attainment, and fertility choice. Am Econ Rev 95(3):580–601
Stoneburner R, Low-Beer D (2004) Population-level HIV declines and behavioral risk avoidance in Uganda. Science 304:714-Ű718
Tamura R (2006) Human capital and economic development. J Dev Econ 79(1):26–72
Temmerman M, Moses S, Kiragu D, Fusallah S, Wamola IA, Piot P (1990) Impact of single session post-partum counselling of HIV infected women on their subsequent reproductive behavior. AIDS 2:247–252
Timberg C (2006) How AIDS in Africa was overstated? Washington Post
UNAIDS/WHO (2003) Epidemiological fact sheets
UNAIDS/WHO (2006) Global report

US Census Bureau (2005) HIV surveillance database
Werker E, Ahuja A, Wendell B (2006) Male circumcision and the economic impact of AIDS in Africa. Harvard Business School, mimeo
Westoff C, Cross A (2006) The stall in the fertility transition in Kenya. DHS Analytical Studies no. 9
Williams BG, Taljaard D, Campbell CM, Gouws E, Ndhlovu L, Van Dam J, Caral M, Auvert B (2003) Changing patterns of knowledge, reported behavior and sexually transmitted infections in a South African gold mining community. AIDS 17:2099–2107
World Bank (2006) World development indicators
Young A (2005) The gift of the dying: the tragedy of AIDS and the welfare of future African generations. Q J Econ 120(2):423–466
Young A (2007) In sorrow to bring forth children: fertility amidst the plaque of HIV. J Econ Growth 12(4):283–327