HIV AND FERTILITY REVISITED

Sebnem Kalemli-Ozcan
Belgi Turan

HIV and Fertility Revisited
Sebnem Kalemli-Ozcan and Belgi Turan
NBER Working Paper No. 16115
June 2010
JEL No. I12,J11,J13,O11

ABSTRACT

Young (2005) argues that HIV related population declines reinforced by the fertility response to the epidemic will lead to higher capital-labor ratios and to higher per capita incomes in the affected countries of Africa. Using household level data on fertility from South Africa and relying on between cohort variation in country level HIV infection, he estimates a large negative effect of HIV prevalence on fertility. However, the studies that utilize the recent rounds of Demographic Health Surveys, where fertility outcomes are linked to HIV status based on testing, find no effect of the disease on the fertility behavior. This paper tries to bridge this gap by revisiting Young's findings. Young (2005) includes data before 1990, when no data are available on HIV prevalence rates. He assigns all the fertility observations before 1990 with HIV prevalence rates of zero, and this appears to drive the significant negative effect found in his study. When one restricts the sample to the period 1990-1998, where actual HIV data are available, the effect of HIV prevalence on fertility turns out to be positive for South Africa. Simulating Young's model utilizing these new estimates shows that the future generations of South Africa are worse off.

Sebnem Kalemli-Ozcan
University of Houston
Department of Economics
Houston, TX 77204
and NBER
sebnem.kalemli-ozcan@mail.uh.edu

Belgi Turan
University of Houston
belgituran@gmail.com

1 Introduction

In a seminal paper, Young (2005) suggests that population declines brought upon by the HIV epidemic will lead to higher capital-labor ratios and to higher per capita incomes in the affected countries of Africa. He argues that widespread community infection will not only reduce labor, but also lower fertility, both directly through a reduction in the willingness to engage in unprotected sex, and indirectly, by increasing the scarcity of labor and the value of women's time. Using household data on fertility from South Africa and relying on between cohort variation in country level HIV infection, he estimates an eighty percent reduction in fertility due to HIV.

In contrast, the studies that utilize the recent rounds of Demographic Health Surveys (DHS), which link an individual woman's fertility outcomes to her HIV status based on testing, find no effect of HIV on the fertility behavior.[1] This paper tries to bridge this gap by revisiting Young's findings. We argue that given the existing trends in South African data, in particular due to abolition of apartheid and the ongoing demographic transition, Young's key identification strategy might not be appropriate. His strategy rests on constructing a panel by tracing fertility histories of women. This is because for South Africa there is only one Demographic Health Survey (DHS), and this was conducted in 1998. He constructed a panel using each woman's birth history since age 12, covering the period between 1961–1998. He then matches this with country level HIV by age and exploits between cohort variation. Since HIV data are not available until 1990, all women are assigned HIV prevalence rates of zero for the period before 1990. This might not be an appropriate strategy since substituting zeros for HIV before 1990—when the actual HIV rates may or may not truly have been zeros—creates a discrete jump in the HIV data, leading to a mechanical downward trend in the residuals. In fact, when one restricts the sample to post-1990, using the actual HIV data,

---

[1]For example, for thirteen African countries, Juhn, Kalemli-Ozcan, Turan (2008) find no significant effect of the community HIV prevalence on the fertility behavior of HIV negative women. This finding holds both in a cross-section of women and also over time by utilizing the available information on each woman's birth history. Fortson (2009) arrives at the same conclusion both for HIV negative women and also for all women using data from twelve African countries and utilizing different time paths for HIV. See also Fink and Linnemayr (2008) for a similar result.

the effect of HIV prevalence on fertility turns out to be positive and significant for South Africa. Simulating Young's original model using our new estimate shows drastically different results. Although we did not consider the "other" detrimental impacts of the disease, such as a decrease in human capital accumulation, we still find that future generations of South Africa are much worse off.

The rest of the paper is structured as follows. Section 2 describes the data and presents the empirical analysis. Section 3 presents the model in Young (2005) and simulates it using both Young's estimates and our new estimates from section 2. Section 4 concludes.

2   Empirical Analysis

Young (2005) shows that there is a negative effect of country-wide HIV prevalence on individual fertility in South Africa. We start by replicating his results and then show that restricting the analysis to observations where HIV prevalence data are actually available are sufficient to overturn the finding of a significant negative effect of HIV on fertility.

The fertility data are from South Africa 1998 DHS.[2] Following Young (2005), only women who are 25 or older are used in the study, and there are 7276 of them in the data set. The panel is constructed using each woman's birth history since age 12, and it includes the period between 1961–1998. Retrospective fertility is the number of pregnancies of each woman in each year, including pregnancies that were lost before term or resulted in stillbirths.

Young (2005) uses the antenatal clinic sero-prevalence rate for each woman's age group at that time of their life from South Africa Department of Health. These rates are available since 1990. We have attempted to get the same HIV data from the same source, but in spite of many emails and calls these data could not be obtained from South Africa Department of Health. Since South Africa Department of Health is cited as the main source in the U.S. Census Bureau's HIV/AIDS Surveillance Database (2005), we therefore use South Africa HIV prevalence rates by age group from the U.S. Census Bureau, HIV Surveillance Database

---

[2]The survey covers all women, not only the women of color.

(2005). We match Young (2005) number of observations exactly, i.e., 171206 women.

We have to note that the country level HIV estimates for any given country in Africa is based on the HIV-1 incidence among pregnant women, whether it is from country's department of health or from the U.S. Census Bureau, HIV Surveillance Database. The reason is that the country level estimates are coming from the antenatal clinics and not based on testing data from a population based survey. Hence these estimates are in general very high and representativeness of these estimates for the general population is debatable.[3] More recently, DHS started providing results from population based HIV testing. These new estimates are much lower than the U.S. Census Bureau estimates.[4] The new population based DHS estimates are only available for a limited set of countries for their latest survey year though and not available for South Africa.

Young's (2005) identification comes from variation in HIV exposure by age. He controls for the effects of age and cohort using linear (and sometimes polynomial) trends in birth year and age, all of which will control for a smooth trend.[5] To replicate the results of Young (2005), we use a Poisson count model. Table 1 shows the replication.[6] Young (2005) finds the coefficient on HIV as –1.63. Using the data from the U.S. Census Bureau, HIV Surveillance Database, we find a coefficient of –1.36 as shown in column (1) of Table 1. By an imprecise reading of the data off of a graph in the printed version of South Africa Department of Health Report (2004), we find a coefficient of –1.58. All these coefficients are significant at 1 percent level. Following Young's methodology, column (1) assumes a HIV prevalence rate of 0 before 1990. In column (2) only part of the panel that is after 1990 is used in regressions. Columns (3) and (4) repeat the same exercise by using OLS estimation. The coefficient of

---

[3]See Timberg (2006) and McNeil (2007).

[4]See Juhn, Kalemli-Ozcan and Turan (2008) for a comparison of various estimates.

[5]In a country like South Africa one can imagine the existence of more complicated trends due to the abolition of apartheid, which is a discrete change.

[6]Wage index is the estimated wage for each individual depending on own education, where wage regressions are done following Young's methodology. Using OHS (1995) wage is estimated as a function of sex, age, education and their squares. Individual education levels reported in OHS are converted to standardized years of education, as shown in appendix. These educational categories are then used to construct wage index. Wage index is calculated as: $B_E * E_i + B_{E^2} * E_i^2$, where B_E and B_{E^2} are the return to education coefficients coming from the regression of wages on sex, age, education and their squares.

interest changes sign and becomes positive significant when the analysis is restricted to the post-1990 part of the panel.

Table 2 investigates the relationship between HIV and fertility for different sub-periods.[7] This corresponds to decreasing the number of fertility observations that are assigned zero HIV prevalence in each reported regression. The estimation is Poisson. The coefficient increases from –1.36, which is estimated using the years 1961–1998, to 0, which is estimated using the years 1986–1998 and then up to 1.64, when only the 1990–1998 part of the panel used, i.e., the actual data. The coefficient switches from being negative significant to being insignificant and then to being positive significant. Hence, the negative effect found in the 1961–1998 sample is not only due to assuming zero prevalence before 1990, but rather the number of years, where HIV is assumed to be zero. Using the years 1986–1998 also means assigning zero prevalence to the fertility observations before 1990, but now we find no effect of HIV on fertility. This result is consistent with Juhn et al. (2008) and Fortson (2009), who both used a similar identification strategy.

The explanation for Young's large negative finding is straightforward. Figure 1 plots the mean residuals (blue circle line) from a regression of fertility on other control variables except for HIV against the mean residuals (black triangle line) from a regression of HIV on other control variables for every year; an exercise that yields the same OLS coefficient as in column (3) of Table 1 by Frish-Waugh theorem. Assuming zero HIV for every woman before 1990 creates an artificial trend in the residuals—as opposed to the case where we use only the part of the panel after 1990—resulting in a negative association between HIV and fertility. In symbols this can be expressed as follows: Figure 1 plots $residuals_{TFR}$ against $residuals_{HIV}$, where $residuals_{TFR}$ come from a regression of fertility on other control variables and hence $residuals_{TFR} = TFR - \hat{a}X$. $residuals_{HIV}$ come from a regression of HIV on other control variables and hence $residuals_{HIV} = HIV - \hat{b}X$. Given $HIV = 0$ before 1990 and X is increasing in years $residuals_{HIV}$ has a mechanical downward trend. Because HIV is zero until year 1990 and then increasing, the predicted HIV is negative, which is reflected in the

---

[7]We are grateful to Christina Paxson for suggesting this exercise.

Figure 1: Partial Correlation Plot by Years for HIV and Fertility
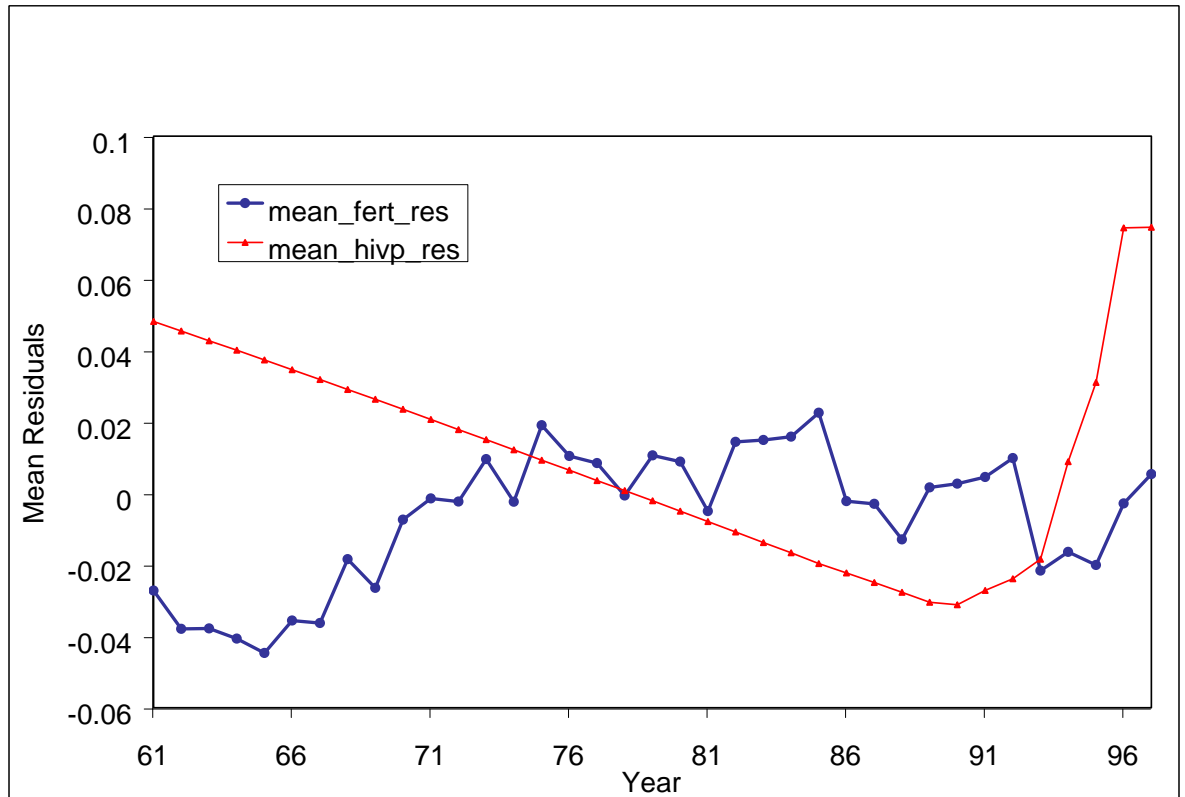
residuals. And since fertility is increasing over time, there is a spurious negative correlation between the negative mechanic trend in HIV pre-1990 and fertility and a positive correlation between the positive trend in HIV post-1990 and fertility. The latter could be also spurious since fertility was increasing even before 1990.

Another way to depict the same relation is shown in Figure 2 that plots mean fertility residual at each year against mean HIV residual at that corresponding year. This figure helps to see the flipping relation due to employing different periods more easily.

It is clear that the significant negative effect of HIV on fertility in South Africa is not a

Figure 2: Partial Correlation Plot for HIV and Fertility
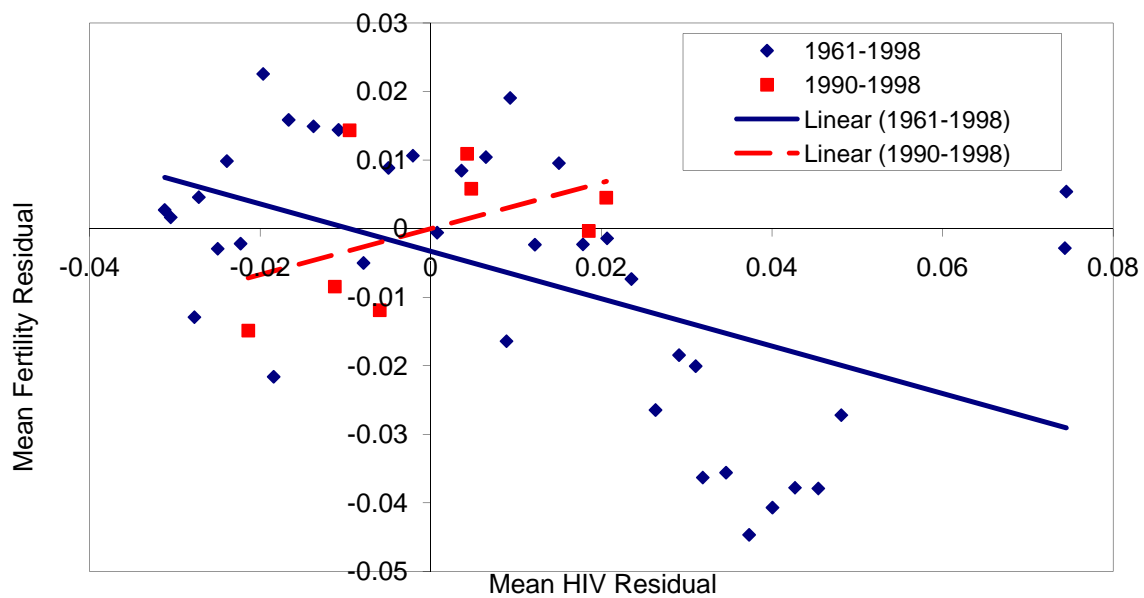
robust finding since it has been driven by creating an artificial discrete trend in the data. In fact out of 171206 observations, 112998 observations are pre-1990. Hence sixty-six percent of the fertility observations are associated with zero HIV prevalence rates since for sixty-six percent of the sample HIV data did not exist. If we focus on the age group of 25-29 that has similar number of observations before and after 1990 we see a striking change in the mean prevalence rate. Before 1990 there were 17537 women aged 25-29. After 1990 there are 13368. The mean HIV prevalence for this group jumps from 0 to 7.2 percent, a jump that is most likely not true in the actual data.

3 Simulation

Although we showed that the estimate for the impact of the disease on fertility behavior changes dramatically from negative to positive when one uses data in the post-1990 period, we still do not know how important this particular input to the results in the Young's paper.[8] Young (2005) brings together different inputs, such as fertility response and human capital accumulation, in the context of a model to simulate the evolution of the South African economy, where a decreasing fraction of population is dying from the disease until the disease ends at some point in the future. However, among all these inputs everything but a negative fertility response contributes to the detrimental impact of the disease on economic development. The negative response of fertility to the disease is most critical since in the absence of any fertility response the other inputs have a negative effect on development. Conditional on the fraction of population that is dying from the disease this will lead to an unambigious decline in income per capita in the absence of a negative fertility response. If fertility responds negatively to the disease then population growth will be reduced drastically not only because of the dying but also because of the lower fertility and hence the income per capita might increase, making the future generations better off.

In order to show the importance of the sign and magnitude of the fertility effect for this

---

[8]We thank our referee to make us see this point and force us to undertake the following exercise.

result, we, first replicate Young's simulation using his exact model and estimates. We focus on his basic scenario and abstract from different scenarios (such as detrimental effect on human capital investment and orphaned children) so that we can show clearly that a change in the estimate for the fertility response turns over the conclusions of his original simulation. Following Young (2005), we assume a Cobb-Douglas production function:

$$Y = AK^{\alpha}(EL)^{1-\alpha} \tag{1}$$

where EL, effective labor, denotes a weighted average of labor inputs.[9] K is the initial level of capital stock in 1995 from Federal Reserve Bank of South Africa's estimate of total fixed capital stock. The path followed by GDP per capita is as follows:

$$k_{t+1} = \frac{(1-\delta)}{(1+n_t)} k_t + \frac{s}{(1+n_t)} A k_t^{\alpha} \tag{2}$$

The depreciation rate, δ, saving rate, s, and labor share, α, are taken as 0.06, 0.175 and 0.62, respectively. Total factor productivity growth is assumed to be zero. We use these values exactly as in Young (2005). We also assume as Young that HIV prevalence rates sinusoidally declines to zero by 2050. We use the same national HIV prevalence rates as before.[10] 1995 birth and death rates are from World Development Indicators Database. We also follow Young on estimating non-AIDS and AIDS mortality. Population growth rate, n_t, is basically determined by AIDS mortality, non-AIDS mortality and fertility behavior. Fertility will be estimated as a function of HIV prevalence and wages as in Young (2005).

We simulate the model under three scenarios: (1) "No HIV," the economy without HIV/AIDS epidemic; (2) "Young's Estimate," the economy with the epidemic and Young's

---

[9]Associated weights are from estimating a wage function using 1995 October Household Survey. Specifically, before-tax wages per hour is estimated as a function of sex, age, education and their squares. Results are shown in Appendix Table A-1. Coefficients from this estimation are used as weights to calculate the effective labor.

[10]Specifically, national HIV prevalence data from US Census Bureau, HIV Surveillance Database is used for pre-2000 period. Post-2000 HIV prevalence evolves according the function $HIV_t = HIV_{2000} * sin[(\pi/2) * (2050 - t)/50]$.

negative fertility estimates, (3) "Our Estimate," the economy with the epidemic and our positive fertility estimate. Scenario (1) and (2) replicates Young (2005). Specifically, in scenario 2, at a 100 percent HIV prevalence rate, fertility would be about 20 percent less. Whereas, in scenario 3, according to our estimate, at 100 percent infection rate fertility would be 500 percent higher.[11] Recall that "our estimate" is nothing but redoing Young's estimation using the actual HIV data between 1990–1998 instead of using 1961–1998 where HIV is assumed to be zero between 1961–1989 as in Young (2005).

Figure 3 shows the evolution of GDP per capita under these two different scenarios relative to the path taken in the absence of epidemic (scenario 1). In scenario 2, "Young's Estimates," GDP per capita remains above "No HIV" path as a result of the large negative fertility response to HIV/AIDS estimated by Young (2005). This is the exact replication of the result found by Young (2005), where future generations are better off because of this increase in GDP per capita. Whereas in scenario 3, fertility response to the epidemic is positive as estimated by us, an effect that dominates the deaths from the disease, and hence GDP per capita decreases and remains under "No HIV" path.[12]

4   Conclusion and Discussion

The relationship between fertility and the HIV/AIDS epidemic is one of the most important missing pieces in the puzzle of AIDS and development. Data from the latest rounds of the DHS surveys show a widespread stall in the demographic transition in Africa, which is inconsistent with declining fertility found in Young (2005) as a result of the HIV epidemic in South Africa. Recent studies using newly available population based HIV-testing data find no significant effect of HIV on the fertility behavior. This paper tries to bridge this gap by revisiting Young's original findings.

---

[11]To ease the interpretation, poisson regression coefficients are converted into incidence rate ratios by exponentiating the poisson regression coefficients in Table 1 of Young (2005) and of this paper. Associated coefficients in Young (2005) and this paper are –1.633 and 1.637, respectively.

[12]See also Figure A-1 in appendix for the scenario without any fertility response to HIV/AIDS, which will correspond to our insignificant estimate.

Using identical individual-level fertility and country-level HIV data as Young (2005), we re-examined the relation between HIV and fertility. The analysis using South African data and cohort-time variation in HIV suggests a significant negative effect of HIV on fertility for the period 1961–1998, a significant positive effect of HIV on fertility for the period 1990–1998, and a zero effect for the period 1986–1998. As a result we cannot draw any generalized conclusion. If one restricts oneself to the years in which there are data, one finds a positive association between HIV and fertility in South Africa—not the negative association that is induced by the artificial trend in HIV rates that are an artifact of assuming zero HIV prior to 1990. Simulating the growth model outlined in Young (2005) with these new estimates show that future generations of South Africa are worse off.

References

Federal Reserve Bank of South Africa. http://www.reservebank.co.za/

Fink, G., S. Linnemayr, "HIV and Fertility: Long-term Evidence from sub-Saharan Africa", Harvard University, mimeo, (2008).

Fortson, J., "HIV/AIDS and Fertility," *American Economic Journal: Applied Economics* (2009), 170–194.

Juhn, C., S. Kalemli-Ozcan, and B. Turan, "HIV and Fertility in Africa: First Evidence from Population Based Surveys," NBER WP 14248, (2008).

McNeil, D., "U.N. to Say It Overestimated AIDS Cases by Millions," New York Times, November (2007).

Republic of South Africa, Department of Health, "National HIV and Syphilis Antenatal Sero-Prevalence Survey in South Africa," (2004).

Timberg, C., "How AIDS in Africa Was Overstated," Washington Post, April 6 (2006).

U.S. Census Bureau, *HIV Surveillance DataBase (2005)*.

World Bank, World Development Indicators, CD-ROM. (Washington, D.C.: The World Bank, 2004).

Young, A., "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations," *Quarterly Journal of Economics* (2005), 423–466.

Table 1: HIV and Individual Fertility in South Africa

Dependent variable is Retrospective Fertility

| Estimation Year	Poisson (1961–1998) (1)	Poisson (1990–1998) (2)	OLS (1961–1998) (3)	OLS (1990–1998) (4)
HIV	−1.359*** (0.190)	1.637*** (0.539)	−0.157*** (0.025)	0.230*** (0.066)
Age	0.502*** (0.008)	0.181*** (0.023)	0.043*** (0.002)	0.004** (0.002)
Age^2	−0.010*** (0.001)	−0.005*** (0.001)	−0.001 (0.001)	−0.001 (0.001)
Wage Index	−0.335*** (0.013)	−0.339*** (0.023)	−0.041*** (0.002)	−0.040*** (0.003)
Year of Birth	−0.003*** (0.001)	−0.065*** (0.014)	−0.006*** (0.002)	−0.007*** (0.002)
Observations	171206	58208	171206	58208
Incidence Rate Ratio	0.257	5.14		

Notes: Only women who are 25 or older are used and there are 7276 of them in the dataset. The panel is constructed using each woman's birth history since age 12, and it includes the period between 1961–1998. Retrospective fertility is the number of pregnancies of each woman in each year, including that were lost before term or resulted in stillbirths. HIV prevalence rates for South Africa are available since 1990, therefore, in columns (1) and (3) HIV prevalence is taken as zero before 1990. In columns (2) and (4) only part of the panel that is after 1990 is used in regressions. Wage index is the estimated wage for each individual depending on own education. Each regression has a constant. Incidence rate ratios are given to ease the interpretation of poisson coefficients. Robust clustered standard errors by individual are in parentheses. ∗∗∗ denotes 1% significance.

Table 2: HIV and Individual Fertility in South Africa: By Different Time-Periods

Dependent variable is Retrospective Fertility

Year	61-98 (1)	65-98 (2)	70-98 (3)	75-98 (4)	80-98 (5)	85-98 (6)	86-98 (7)	87-98 (8)	88-98 (9)	89-98 (10)	90-98 (11)
HIV	-1.359***	-1.247***	-0.976***	-0.910***	-0.831***	-0.723**	-0.357	-0.516	-0.222	0.902**	1.637***
	(0.190)	(0.190)	(0.193)	(0.202)	(0.221)	(0.305)	(0.281)	(0.338)	(0.385)	(0.449)	(0.539)
Age	0.502***	0.496***	0.475***	0.442***	0.402***	0.345***	0.333***	0.302***	0.274***	0.222***	0.181***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.011)	(0.013)	(0.014)	(0.016)	(0.019)	(0.023)
Age^2	-0.010***	-0.009***	-0.009***	-0.009***	-0.008***	-0.007***	-0.007***	-0.006***	-0.006***	-0.005***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Wage Index	-0.335***	-0.334***	-0.330***	-0.322***	-0.327***	-0.333***	-0.331***	-0.327***	-0.326***	-0.325***	-0.339***
	(0.013)	(0.013)	(0.014)	(0.014)	(0.016)	(0.018)	(0.019)	(0.020)	(0.020)	(0.021)	(0.023)
Year of Birth	-0.003**	-0.004***	-0.007***	-0.007***	-0.006***	-0.012***	-0.002	-0.005	-0.011	-0.042***	-0.065***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.005)	(0.006)	(0.008)	(0.010)	(0.014)
Obs.	171206	169487	162305	148030	125493	94588	87312	80036	72760	65484	58208

Notes: Only women who are 25 or older are used and there are 7276 of them in the data set. The panel is constructed using each woman's birth history since age 12, and it includes the period between 1961–1998. Retrospective fertility is the number of pregnancies of each woman in each year, including that were lost before term or resulted in stillbirths. HIV prevalence rates for South Africa are available since 1990, therefore, all the columns before column (11) assumes HIV prevalence is zero before 1990. Wage index is the estimated wage for each individual depending on own education. Each regression has a constant. Robust clustered standard errors by individual are in parentheses. $***$ denotes 1% significance.

13

Figure 3: GDP per Capita (relative to No HIV)
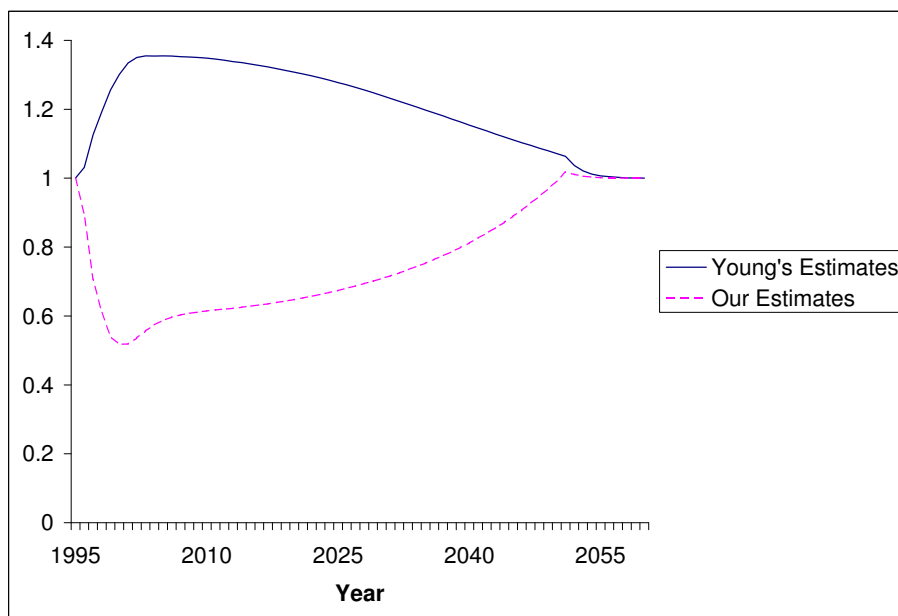
Appendix

Education: Reported education levels are converted into standardized years of education as follows: (i) No schooling or less than one year completed = 0 years; (ii) Sub A/sub B/grade 1/grade 2/Std 1 = 2 years; (iii) Standards 210 = standard year = 2; (iv) Diploma/certificate with Std 9 or lower or further studies incomplete = 13 years; (v) Diploma/certificate with Std 10 or diploma/other postschool complete = 14 years; (vi) Degree or further degree complete = 16 years.

Fertility: We use women's birth histories in DHS to construct the number of births in each year of their lives. We count pregnancies that were lost before term or resulted in still-births as births. We exclude women younger than 25 at the time of the survey from sample. Finally, following Young(2005) we do not include the fertility history since each woman's last birthday.

Before-tax Wages per Hour: 1995 October Household Survey is used to construct before-tax wages per hour. Individuals report their income in terms of intervals and state whether this income is daily, weekly, monthly or annual. Additionally, they report weekly in-kind income (transport, food, and other) they received. Using this information annual income for each individual is calculated. Dividing the annual income by the hours of work in the last seven days the individual worked gives annual before-tax wages per weekly hour of work. Only individuals worked for someone else are used in the calculations. 424 individuals who stated that they worked both for themselves and someone else are excluded from the sample. Individuals who are younger than 25 and at still school-going age are also included in the calculations.
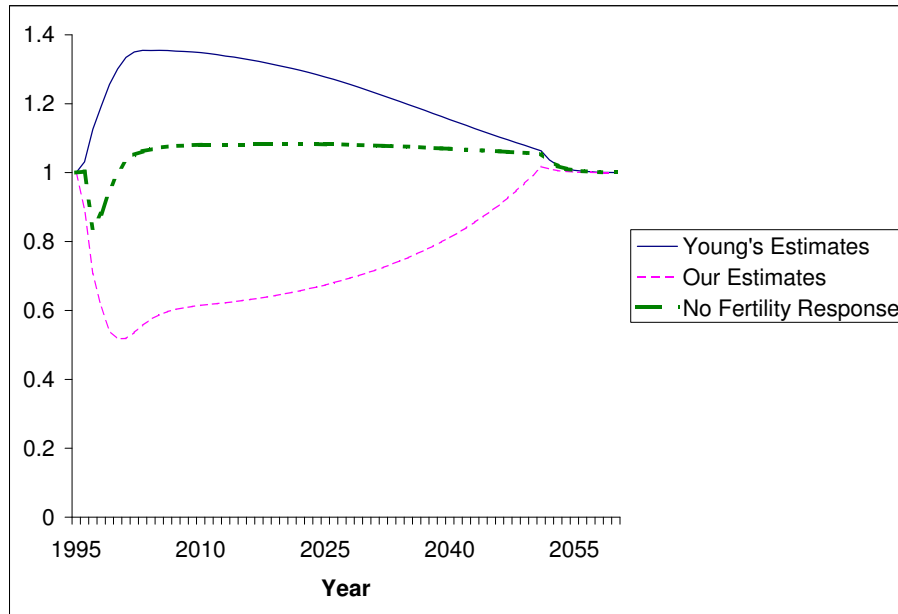
Figure A-1: GDP per Capita (relative to No HIV)

Table A-1: Estimation of Elements of Effective Labor Supply

Dependent Variable: Log Before-Tax Wages per Hour

	Interval Regression
Age	0.082
	(0.003)
Age^2	−0.0008
	(0.000)
Education	0.037
	(0.004)
$Education^2$	0.008
	(0.000)
Sex	−0.218
	(0.100)
Constant	2.932
	(0.051)
N	26290
Pseudo R^2	0.198

Notes: Data is from 1995 October Household Survey. Constructions of log before-tax wages per hour and education variable are described in the appendix. Since wage data are interval coded interval estimation is used. It maximizes the likelihood that the dependent variable falls within the interval brackets. All females and males worked for someone else in the last seven days, including the ones still at school-going age, are used in the regressions. Standard errors are clustered on enumeration area and in parenthesis.