HIV, Wages, and the Skill Premium

Ioana Marinescu
University of Chicago

Abstract

The HIV epidemic has dramatically decreased labor supply among prime-age adults in sub-Saharan Africa. Using within-country variation in regional HIV prevalence and a synthetic panel, I find that HIV significantly increases the capital-labor ratio in urban manufacturing firms. The impact of HIV on average wages is positive but imprecisely estimated. In contrast, HIV has a large positive impact on the skill premium. The impact of HIV on the wages of low skilled workers is insignificantly different from 0, and is strongly dampened by competition from rural migrants. The HIV epidemic disproportionately increases the incomes of high-skilled survivors, thus increasing inequality.

I am a grateful to Jane Fortson for abundant comments on this paper and for sharing code with me. I also want to thank David Autor for very useful comments. Finally, I am grateful to participants at the University of Houston/Rice, the University of Chicago, SOLE meetings 2010, and the North American Summer Meetings of the Econometrics Society in St Louis for helpful comments, to Javaeria Qureshi, Ana Sofia Leon and Raffaele Livio De Marco for research assistance, and to Jo Van Biesebroeck for help with the data.

I. INTRODUCTION

The HIV/AIDS epidemic represents one of the greatest public health crises of the past several decades. Since HIV was recognized by the Centers for Disease Control and Prevention in 1981, prevalence has risen worldwide, but particularly in sub-Saharan Africa, which accounts for about two-thirds of current HIV infections (UNAIDS, 2008). Across sub-Saharan Africa, HIV prevalence among adults aged 15-49 is estimated to be 5.0 percent (UNAIDS, 2008). In 2007, 22 million people in sub-Saharan Africa were infected with HIV, and about 1.5 million died from AIDS (UNAIDS, 2008). The burden of disease is particularly large in southern Africa, where, in some areas, more than half of adults are infected.[1]

Because the bulk of HIV infections are sexually transmitted, HIV disproportionately affects prime-age adults. In particular, in most countries, HIV prevalence is highest among women in their 30s and men in their 30s and early 40s (Mishra et al., 2009). In the absence of treatment, HIV-infected patients are expected to live about ten years from the time of infection. For much of that period, they remain asymptomatic; however, patients eventually become very sick with AIDS before death, usually for about one year.

Together, the HIV rates and the age profile of infection imply large increases in both morbidity and mortality among prime-age adults in sub-Saharan Africa, and particularly in southern Africa. AIDS – by changing population composition, risks, and decision-making – has potential spillovers on other sectors, including education and fertility (e.g., Fortson, 2011, Fortson, 2009, Juhn, Kalemli-Ozcan and Turan, 2008). The economic effects of the epidemic are potentially staggering, and economists have devoted considerable effort to estimating and

[1] Prevalence is particularly high in Botswana (23.9%), Lesotho (23.2%), South Africa (18.1%), and Swaziland (26.1%) (UNAIDS, 2007).

forecasting the economic effects of HIV/AIDS. Motivated by evidence from the Black Death which found that the disease raised wellbeing among survivors, Young (2005) develops a model of the HIV/AIDS epidemic showing that future generations of Africans may benefit from AIDS-related reductions in population size, and, in particular, that wages of surviving workers may increase. However, Kalemli-Ozcan (2010) and Santaeulàlia-Llopis (2008) each develop models which support the opposite prediction – that HIV will reduce wellbeing in sub-Saharan Africa. Complementing the theoretical predictions, several studies have assessed the empirical relationship between GDP and HIV in sub-Saharan Africa (see, for example, Bloom and Mahal, 1997, Bonnel, 2000, Ahuja, Werker, and Wendell, 2009). These studies find no significant impact of HIV on economic growth, though they are generally powered to detect only very large effects.

Several studies have also assessed the impact of HIV and antiretroviral (ARV) treatment on labor supply and worker productivity, generally finding moderately sized effects in the short-term (e.g., Thirumurthy, Graff-Zivin and Goldstein, 2008, Fox et al., 2004, and Larson et al., 2008). Looking at the medium-term effects of ARV treatment, Habyarimana, Mbakile, and Pop-Eleches (2010) find that, as much as four years out, firm provision of ARV treatment led to reductions in worker absenteeism. Thirumurthy et al. (2011) show that ARV treatment increases employment and income as far as two years out.

This paper focuses on the impact of HIV on wages. It uses data from the World Bank's Africa Regional Program on Enterprise Development (ARPED) and the World Bank's Enterprise Surveys. These surveys cover urban manufacturing firms and their workers. I calculate regional HIV prevalence from the Demographic and Health Surveys. I then link wage data from manufacturing firms and their employees to regional HIV prevalence estimates.

I develop a simple model to predict the impact of AIDS mortality on wages at the market level. All other things equal, AIDS mortality decreases labor supply. In the short run, as long as the capital stock does not fully adjust downwards, higher mortality leads to an increase in wages and the capital-labor ratio. In the long run, once capital has fully adjusted, there should be no effect on wages. If the production function is Cobb-Douglas, the short-run wage elasticity of AIDS (i.e. the change in wages resulting from the change in labor supply due to AIDS mortality) is equal to the capital share.

However, because the wage data I use covers the urban manufacturing sector only, the estimated short-run wage elasticity of AIDS could be smaller than the capital share. Suppose that these urban manufacturing firms pay wages that are higher than in most other sectors, and above market-clearing levels. If so, they do not need to increase their wages much (if at all) in order to attract new workers and replace those lost to AIDS. Effectively the labor supply to urban manufacturing firms will decline less than the overall labor supply. This means that the short-run wage elasticity of AIDS could be lower than the capital share when estimated on urban manufacturing firms only. Additionally, this raises the possibility for the HIV epidemic to be associated with an increase in the skill premium. Indeed, high paying manufacturing firms can fairly easily replace low skilled workers who die of AIDS by workers from other urban sectors or from rural areas while barely increasing their wages (see discussion of dual labor markets in Fields, 2005). By contrast, high skilled workers who die of AIDS will be harder to replace without increasing wages, since they cannot be easily recruited from outside the manufacturing sector. If the supply of low skilled workers to urban manufacturing firms declines less than the supply of high skilled workers, the impact of HIV on the wages of high skilled workers will be higher than on the wages of low skilled workers. Therefore, I expect a positive association between HIV and the skill premium.

The empirical specifications use within country variation in HIV prevalence to identify the impact of HIV on outcomes both in levels and in growth rates. Consistent with theoretical expectations, I find that higher HIV rates are associated with significantly higher capital-labor ratios. However, HIV has no significant impact on labor productivity. The estimated wage elasticity of AIDS is positive but not always significantly different from zero. In contrast, higher HIV prevalence is consistently significantly associated with a higher skill premium: a 10% increase in HIV prevalence is associated with a 1.5 percentage point greater increase in the wages of workers with 14 years of education or more, as compared to workers with 10 years of education or less. The impact of HIV on workers with 10 years of education or less is not significantly different from 0. Next, following a previous paper in the literature (Ahuja et al., 2009), I instrument regional HIV prevalence with the male circumcision rate in growth rate equations. The IV results are similar to OLS results, and the IV specifications are not underidentified. This provides further evidence for a causal impact of HIV on the capital-labor ratio and the skill premium. Finally, I present additional evidence consistent with dualism in the labor market. First, I show that the positive association between HIV and the skill premium does not hold in small firms, i.e. those firms that are less likely to pay above market-clearing level wages to start with. Second, I show that recent migration from rural areas is an important factor that modulates the impact of HIV on the wages of the least skilled workers. Indeed, the impact of HIV on low-skilled workers' wages is estimated to be positive and significant in the absence of rural-urban migration, but this impact declines significantly as rural-urban migration increases.

This paper is closest in its specific theme to Young (2005), who calibrates a model that formalizes the impact of HIV on wages. However, from a theoretical and econometric perspective, this paper is closest to Borjas (2003) and Acemoglu, Autor and Lyle (2004). Both of these studies analyze the impact of an increase in labor supply on the wages of different groups in the US labor force. This paper makes three key contributions to the literature on the impact of the HIV epidemic on wages and labor market outcomes more broadly. First, the previous literature investigating the impact of HIV on outcomes such as GDP growth and the capital/labor ratio relied on country-level data. Thus, unobserved differences in labor market trends across countries could bias the results. In contrast, my analysis uses within-country variation in HIV prevalence as a source of identification, both in the cross-section and over time. Second, this paper uses a broad sample of Sub-Saharan African countries to investigate the impact of the HIV epidemic on market-level outcomes such as wages and labor productivity. The previous literature has used solid identification strategies to analyze the impact of HIV on workers' productivity and income at the micro level. However, each of these papers is based on variation within a single, small geographic area and over a relatively short period of time. Most importantly, this strand of literature typically concentrates on worker-level outcomes for HIV-positive workers, while I focus on market-level outcomes. This distinction is fundamental. In particular, the impact of HIV on workers' income and productivity is theoretically negative for HIV-infected workers, and the literature on the topic upholds this theoretical expectation. By contrast, the short-run impact of the HIV epidemic on wages at the market-level is theoretically positive if only AIDS mortality is taken into account, and becomes ambiguous when also taking into account the decline in productivity for HIV-infected workers. It is therefore of great interest to analyze the impact of the HIV epidemic on labor markets empirically, and such an analysis complements the studies of the impact of HIV infection on individual labor supply and income. To perform this analysis within a broad sample of countries, the data set I use is unique. Indeed, the Enterprise Surveys and ARPED data I use are, to my knowledge, the only available micro datasets that provide standardized information on wages in a large sample of sub-Saharan African countries. The Demographic and Health Surveys, the only other survey that is comparable in the range of countries and years covered, does not provide any information on wages. Third, my data offers the unique opportunity of examining the relationship between HIV and linked firm- and employee-level outcomes. These two complementary sources of data allow me to explore both firm-level reactions to HIV, such as capital adjustment, and worker-level reactions to HIV, such as changes in hours worked. Investigating both firm-level and worker-level outcomes is crucial in understanding the mechanisms through which HIV affects wages.

The remainder of the paper is organized as follows. Section II presents a simple theoretical framework to quantify the impact of AIDS mortality on wages. Section III presents the empirical estimates of the relationship between HIV and wages, as well as estimates of the impact of HIV on the capital stock, labor productivity, and hours worked. Section IV concludes.

II. THEORETICAL FRAMEWORK

A. THE IMPACT OF HIV ON AVERAGE WAGES

In this section, I assume that the sole way in which HIV affects wages is through a decline in labor supply deriving from AIDS mortality. I adopt a modeling framework adapted from Borjas (2009) to derive the "wage elasticity of AIDS", i.e. the increase in wages due to the reduction in labor supply induced by AIDS mortality. The production function in the economy is a constant elasticity of substitution (CES) production function given by:

$$Q = \left[\alpha K^{\delta} + (1-\alpha)L^{\delta}\right]^{\frac{1}{\delta}}$$

where, K is the capital stock, L is the number of workers, and $\delta \leq 1$. The supply of capital is given by $r = K^{\lambda}$, where r is the price of capital and $\lambda \geq 0$.

Using properties of the CES production function, one can show that:

$$\frac{d \log w}{d \log L} = \frac{-\lambda(1-\delta)s_K}{1+\lambda-\delta-(1-\delta)s_K}$$

where $s_K = rK/Q$ is the share of capital. In the short-run ($\lambda = \infty$), and assuming that the CES production function reduces to a Cobb-Douglas ($\delta = 0$), we have:

$$\frac{d \log w}{d \log L} = -s_K$$

Therefore, under these assumptions, the wage elasticity of AIDS should be positive (remember that AIDS mortality *decreases* labor supply) and equal to the share of capital.

B. THE IMPACT OF HIV ON THE SKILL PREMIUM

I now introduce skill heterogeneity among workers. Maintain the assumptions above and further assume that the labor force L is composed of two groups L_1 and L_2, which we will define as high skill and low skill respectively. The Armington aggregator is given by:

$$L = (\theta_1 L_1{}^{\beta} + \theta_2 L_2{}^{\beta})^{1/\beta}$$

where $\beta \leq 1$ and $\theta_1 + \theta_2 = 1$. The elasticity of substitution between high and low skilled is $\sigma_{12} = 1/(1-\beta)$.

In this model, one can show that the short-run wage elasticity of AIDS does not depend on the detail of the impact of AIDS on the two skill groups, but instead remains at s_K (assuming a Cobb-Douglas production function). Additionally, if AIDS mortality rates are the same for all skill groups, then HIV should not have any impact on the skill premium.

If AIDS mortality differs by skill group, then the skill premium is affected. As I will show later (section III.D.2), there is no evidence that national-level AIDS mortality systematically differs by education in my data. However, there is another reason why the impact of HIV on labor supply may differ by skill group: that is labor market dualism. Urban manufacturing firms in my sample are mostly formal. They likely pay above market-clearing wages to their low skilled workers due to efficiency wages or other institutional features (Fields, 2005). Even if AIDS mortality reduces the overall supply of workers, the firms in my sample may still be able to hire low-skilled workers from the informal sector without having to raise wages. Therefore, AIDS mortality may only create a shortage of high-skilled workers, since these workers cannot be easily hired out of the informal sector (Perry et al., 2007).

The role of dualism can be formalized as follows. Let $m_i = dL_i/L_i$ be the AIDS-induced percent supply shift for group i. Consistent with the idea that, due to dualism, the supply of low-skilled workers to the urban manufacturing sector is not affected by AIDS mortality, assume that $m_2 = 0$. By contrast, assume that the labor supply of the high skilled is diminished by exactly the average AIDS mortality in the overall population. Under these assumptions, one can show that the impact of AIDS mortality on the skill premium is given by:

$$\frac{d\log w_1 - d\log w_2}{m_1} = -\frac{1}{\sigma_{12}}$$

Thus the impact of AIDS mortality on the skill premium only depends on the elasticity of substitution between skill groups. Additionally, under these same assumptions, the overall effect of AIDS mortality on the average wage can be rewritten as:

$$\frac{d\log w}{m_1} = \frac{s_1}{s_L}\frac{d\log w}{d\log L}$$

where s_1/s_L is the share of the wage bill that goes to the high skilled. Thus, in the short run and assuming a Cobb-Douglas production function, the impact of AIDS mortality on average wages in urban manufacturing is $s_K s_1/s_L$. Since the share of the wage bill that goes to the high skilled is strictly less than 1, the assumption of a dual labor market implies that the wage elasticity of AIDS is strictly less than the capital share.

C. DISCUSSION

The above model is a simplified description of the world and does not fully capture the mechanisms at play.

The first set of issues to consider is related to the supply of capital. The above discussion was focusing on short-run effects. If the capital stock adjusts to some degree in response to AIDS mortality, then the wage impact will be smaller (closer to 0). Second, while many macro models assume a Cobb-Douglas production function, recent research based on more disaggregated data (e.g. Antras, 2004, Juselius, 2008) has shown that the elasticity of substitution between capital and labor is less than 1, i.e. $\delta < 0$ in the model above. If that is the case, then the impact of AIDS mortality on wages in the short run should be larger, not smaller than the capital share. I will further discuss the impact of HIV-AIDS on the supply of capital in section III.C.3.

The second mechanism that the simple model presented above neglects is the impact of AIDS mortality on product demand. There would be an increase in product demand (relative to production) if those who survive produce less than they consume, and a decline in product demand if those who survive produce more than they consume (i.e. those who died from AIDS consumed more than they produced). However, since AIDS mortality is concentrated among working-age adults, there should not be much of an effect: the decline in the labor stock likely diminishes production and consumption to similar degrees.

The third issue that could affect the empirical estimate of wage elasticity of AIDS is a fertility response to the HIV epidemic, as hypothesized by Young (2005). However, subsequent empirical evidence on the issue suggests that there has been no fertility response to the HIV epidemic (Fortson, 2009, Kalemli-Ozcan, 2010, Juhn, Kalemli-Ozcan and Turan, 2008, Kalemli-Ozcan and Turan, 2010). We can therefore neglect this channel.

The fourth and final issue relates to measurement: do HIV rates mostly measure AIDS mortality, as we have assumed so far in this theoretical section, or do they also measure something else that affects wages? In fact, HIV rates do not only relate to mortality but also to morbidity. In other terms, because of HIV-related sickness, workers may be less productive. Because of this, wages should decrease with HIV, and hence the empirical estimates of the wage elasticity of AIDS will be biased downwards. In the empirical section III.C.3, I will examine the evidence for a decrease in labor productivity associated with HIV.

Having examined how HIV may impact wages from a theoretical perspective, I now move on to the empirical analysis.

III. DATA AND RESULTS

A. DATA

1. Labor Market Data

Labor market data are drawn from two sources, both from the World Bank: the Africa Regional Program on Enterprise Development[2] (ARPED) and the Enterprise Surveys. Both data sources include the results of firm and employee surveys, with very similar questions across sources. The ARPED data (which cover the early 1990s) include the results of a survey of manufacturing firms from four manufacturing industries: agro, textiles and leather, wood, and metals. A subset of employees from these firms was also interviewed. The survey years for the ARPED data range between 1992 and 1995. The Enterprise Surveys are the updated version of the ARPED data. They were conducted between 2002 and 2007, are similar to the ARPED data, but may include non-manufacturing firms in addition to manufacturing firms. I however restrict the analysis to manufacturing firms only in order to make the data comparable across countries and time. Employees in firms selected for the firm survey could be chosen for the employee survey; up to ten employees per firm were surveyed. The employees do not form a representative sample. Instead, they are designed to be representative of key occupations within the firm. Matched employee surveys are available for only some countries in the Enterprise Surveys.

My cross-sectional firm-level analysis uses data from the Enterprise Surveys from fifteen countries: Burkina Faso, Cameroon, Congo, Cote d'Ivoire, Ethiopia, Ghana, Guinea, Kenya,

[2] This dataset was kindly provided to me by Jo Van Biesebroeck. For more information on this dataset, see Van Biesebroeck (2005).

Malawi, Mali, Niger, Rwanda, Swaziland, Tanzania, Zambia.[3] The survey is designed to be representative of the main industries in each country.

My synthetic panel analysis uses Enterprise Survey data in conjunction with ARPED data from seven countries (Cameroon, Cote d'Ivoire, Ethiopia, Ghana, Kenya, Tanzania and Zambia). Though all countries in the ARPED data have linked employee data (with up to 52 employees per firm), only Cameroon, Ghana, Kenya, Tanzania, and Zambia also have employee data in the Enterprise surveys. Therefore, the synthetic panel analysis using employee data is based on these five countries.

Both the ARPED and Enterprise Survey data have firm-level information about average wages (computed as wage bill divided by the number of workers), sales, labor productivity (as measured by sales per worker), and the capital/labor ratio. I clean the data by removing all firms whose wage bill per worker or sales per worker were in the top or bottom half percent of the distribution[4] (using Purchasing Power Parity adjusted values). Employee data include measures of wages, education, hours worked and job tenure. All data in local currency have been converted to constant prices using the GDP deflator from the IMF (this is the data I use unless otherwise specified), and to 2005 PPP dollars using the United Nations series. Table 1 summarizes these outcomes.

2. HIV Prevalence Data

I link these labor market data to regional HIV prevalence estimates based on calculations from the Demographic and Health Surveys (DHS). The DHS are a series of nationally-

[3] The Enterprise Surveys do not provide employee-level data for three of these countries: Cote d'Ivoire, Ethiopia, and Swaziland. Therefore, employee-level analysis, in the cross-section, includes only 12 countries.

[4] The cleaning is done separately for the 2000s and the 1990s data.

representative cross-sectional household surveys that have been conducted in over 85 countries over the past 25 years. In several countries, a recent survey round included the collection of blood samples for HIV testing. These HIV test results – designed to be representative at the national (and regional) level – can be used to generate estimates of HIV prevalence among adults. The availability of these data represents a significant change in the methods and findings of HIV testing. Whereas previous estimates of HIV prevalence came from testing in antenatal clinics or tests of specific subpopulations (e.g., commercial sex workers), these data cover a broader range of adults, including men. UNAIDS revised its estimates of prevalence based on the results from DHS HIV testing data.

My analysis uses DHS HIV testing data for the 2000s from 15 countries: Burkina Faso, Cameroon, Congo, Cote d'Ivoire, Ethiopia, Ghana, Guinea, Kenya, Malawi, Mali, Niger, Rwanda, Swaziland, Tanzania, Zambia. I calculate HIV prevalence among men and women aged 15-49 in each region (administrative division) within each country. The individual HIV tests are highly accurate[5].

One limitation of these HIV prevalence data is survey non-response. Refusal was the most common reason for non-response, but non-response also occurred because of absence and testing problems. On average, about 20 percent of eligible test respondents were not tested. If non-respondents differ from respondents in their probability of HIV infection, non-response could bias estimates of HIV prevalence. However, Mishra et al. (2006), in an analysis of DHS data from eight countries, find that adjusting HIV prevalence estimates for differences in demographic characteristics between respondents and non-respondents does not have much

[5] DHS testing is done using an initial ELISA test, and then retesting of all positive tests and 5-10 percent of the negative tests with a second ELISA. For those with discordant results on the two ELISA tests, a new ELISA or a Western Blot is performed. The ELISA test by itself has a sensitivity of 99.7% and specificity of 98.5% (Chou et al. 2005). Repeating the ELISA test (or using the Western Blot) as was done in the DHS reduces the ultimate likelihood of a false positive even more.

effect on national prevalence estimates using these data. Furthermore, Fortson (2008), using DHS data from Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania, finds no relationship between non-response and reported sexual behavior.

In addition to showing summary statistics for the labor market data, Table 1 shows estimates of prevalence in the 2000s. Prevalence is measured at the region level; each firm and employee is assigned the regional-level[6] prevalence measured in the most recent DHS data[7].

3. AIDS mortality and HIV

My main empirical specifications will use HIV rates defined at the regional level, within countries. Since theory relied heavily on the assumption that HIV rates reflect AIDS mortality, it is important to investigate the relationship between HIV rates and AIDS mortality. We need to express HIV rates as a function of AIDS mortality in order to be able to derive the wage elasticity of AIDS from a regression of wages on HIV.

AIDS deaths are only available at the national level from UNAIDS. AIDS deaths are recorded annually since 1990; it is believed that there were very few AIDS deaths before that date. To calculate AIDS mortality at the country level, I first take the sum of AIDS deaths between 1990 and the year of the 2000s firm survey I use for each country. I then divide this sum by total adult population in 1989, the year before AIDS deaths started to be recorded: this gives me the national adult AIDS mortality. The reason for dividing by adult population is that most AIDS deaths occur among adults, and, additionally, I am interested in the AIDS-induced reduction in the labor force, not in the total population. For the countries in my sample, total

[6] For the firm and employee data, the region is determined based on the city where the establishment is located.

[7] For a few countries, more than one Enterprise Survey was available in the 2000s. In these cases, I chose to use the survey that was closest in time to the DHS HIV testing sample.

adult AIDS mortality is 7% on average, and ranges from 0.4% to 23.2%. This is a reasonably sized shock to the labor force, if one is to judge by the kind of shocks to the labor force that have been considered in the literature on the wage impact of immigration. For example, in David Card's 1990 paper about the Mariel boat lift, the shock to the labor force was 7%. And, in 2000, the share of immigrants in the US population was 13% (Card, 2005).

The next step is to identify the relationship between national adult HIV prevalence rates and AIDS mortality thus defined. Figure 1 shows a plot of the log of AIDS mortality in the year of the most recent firm survey I use versus the log of the national adult HIV prevalence rate[8] in the same year (also from UNAIDS[9]). As can be seen, the linear fit between these two measures is very good[10]. The coefficient on log AIDS mortality is 0.720. In my wage regressions, I will regress the log wage on the log of HIV prevalence. To recover the wage elasticity of AIDS, we need to multiply the coefficient on the log of HIV prevalence by 0.720 and divide by the average AIDS mortality, i.e. 0.7 (dividing by 0.7 is necessary because we used log AIDS mortality and not AIDS mortality in the first regression). Thus, conveniently, to recover the wage elasticity of AIDS, one just needs to multiply the coefficient on log HIV by 10[11].

[8] This log log specification was used because it provides a much better fit than the same regression in levels.

[9] The national DHS HIV rates and the UNAIDS rates are very highly correlated (0.98) because the UNAIDS data is based on the DHS when available.

[10] Many of the mortality estimates are based on an epidemiological model that has HIV prevalence as its input, so it is not surprising to get a good fit. The epidemiological model was validated using AIDS deaths data when such data was available.

[11] Table 10 in the appendix shows that, when regressing wages on national AIDS mortality and log national HIV respectively, the coefficient on national AIDS mortality is indeed about 10 times the coefficient on log national HIV prevalence.

B. ECONOMETRIC SPECIFICATIONS

1. Cross-Sectional Regressions

I use the same specifications in the firm sample and the employee sample. For each outcome y_i for firm or worker *i*, I run the following regression:

$$y_i = \alpha HIV_r + X_i\beta + \gamma_c + \delta_d + \varepsilon_i$$

where HIV_r is log regional HIV prevalence, X_i is a set of controls, γ_c is a vector of country fixed effects, δ_d is a vector of industry fixed effects, and ε_i is a normally-distributed error term. Outcomes y_i are expressed in logs. Standard errors are clustered by region. The source of identification is thus variation in HIV prevalence across regions, within industries and countries.

In order to estimate the impact of HIV on the skill premium, I use employee data only, since wages by schooling level are not available at the firm level. I use the same specification as above, but I replace the schooling control by dummies for schooling terciles, and I interact these dummies with HIV. Specifically, the schooling terciles are: schooling <= 10, 10< schooling <=13 and schooling > 13.

2. Synthetic Panel Regressions

The relationships we can uncover using cross-sectional identification are descriptively interesting, but they may be driven by unobserved differences in regional outcomes that are correlated with the regressor of interest, i.e., regional HIV prevalence. To address this concern, I use two strategies. First, I use a synthetic panel regression, which allows me to account for differences in regional outcomes that are constant over time. Second, in this synthetic panel regression, I instrument HIV prevalence with male circumcision rate based on self-reports in the DHS.

The most widespread way of addressing the issue of the absence of a true panel over time is to use synthetic cohorts (see Verbeek 2007). Indeed, I do not have panel data, but only a repeated cross section of manufacturing firms and their employees in the early 1990s and the 2000s. To use synthetic cohorts, one must group observations by variables that do not vary over time: region is such a variable for this data.

The data from the 1990s covers only four industries and the sample size is the same for all industries; by contrast, the 2000s data covers more industries and the sample size is unequal across industries. Since the sampling frame in the 2000s reflects the industry composition specific to each country, and I want to make the 1990s cross-section as representative as possible of the 2000s cross-section, I create weights in the 1990s data to make the sample representative of the 2000s data in terms of industries[12]. I then collapse wages by region and time period (1990s and 2000s), taking a weighted mean of these wages. This whole procedure of weighing and collapsing is done twice: once for the firm sample, and once for the employee sample. Note that collapsing the data leaves us with few observations because only seven countries can be matched between the ARPED (1990s) and the Enterprise Surveys data (2000s) for the firm data, and only five countries for the employee data; additionally, within these countries, only a subset of regions can be matched. Because of these small sample sizes, one may be tempted to divide the data in finer cells than by region. However, when using a synthetic panel, it is very important to have a sufficient number of observations per cell (in this case, region by year cell) in order to avoid small-sample bias in estimations (Verbeek, 2007). The asymptotics rely on a fixed number of

[12] Specifically, I regress the indicator for belonging to the 2000s data on industry*region indicators using a logit specification. I then use p/(1-p) as the weight for observations in the 1990s data, where p is the predicted probability of belonging to the 2000s data (see DiNardo, 2002).

cells and the number of observation per cell going to infinity. Hence, what one should pay attention to in this context is the number of observations per cell, and not the number of cells.

Using the collapsed data, I run the following regression:

$$\Delta y_{rt} = \alpha HIV_{rt} + \varepsilon_{rt}$$

where Δy_{rt} is the difference in log wages at the regional level divided by the number of years between the two periods of measurement (1990s and 2000s); dividing by the number of years is important because different countries had different spacings between their two surveys. Thus, Δy_{rt} is approximately the yearly growth in the outcome of interest. HIV_{rt} is the log of regional HIV prevalence in the 2000s and ε_{rt} is a normally distributed error term. Robust standard errors are computed. Additionally, each observation is weighted by $1/(1/\mathrm{n}_{\mathrm{rt}} + 1/\mathrm{n}_{\mathrm{r,t-1}})$ using Stata's analytic weights, where n_{rt} is the number of observations that have been collapsed in region *r* and period *t*. The weight takes into account that a growth rate is calculated over two periods of data, and that means are more precise when there are more observations. For the firm sample, the average number of observations per cell is 46, which is fewer than 80, the minimum number of average observations per cell that has been used in a set of prominent studies using synthetic panel regressions (Verbeek, 2007). Weighting for the number of collapsed observations per cell is thus very important in order to down-weight the growth rates based on cells with few observations.

The regression specification used evaluates whether wages grew more between the early 1990s and the 2000s in regions that have higher HIV prevalence in the 2000s. This specification corresponds closely to the first difference of the cross-sectional specification in the case where HIV prevalence in the early 1990s is 0 and characteristics X_i do not change over time (country fixed effects are dropped when taking the first difference[13]). Is it reasonable to assume that HIV prevalence was 0 in the early 1990s? I cannot use regional HIV prevalence in the 1990s because no reliable data is available at the regional level: indeed DHS HIV testing only happened in the 2000s. National HIV prevalence numbers, themselves not as reliable as the DHS measures, indicate that HIV prevalence in the early 1990s was low but not 0 on average. What is however the case is that cumulative AIDS mortality was essentially 0 in the early 1990s (it was 0.6% on average, with a maximum value of 1.4%). Thus, if HIV is taken to reflect AIDS mortality only, my synthetic panel specification is very close to the first difference of the cross-sectional specification, and hence the coefficient on HIV can be taken to estimate the same underlying relationship in both the cross-sectional and the synthetic panel specifications. Finally, because Δy_{rt} is the difference in log wages at the regional level divided by the number of years between the two periods of measurement (1990s and 2000s), the coefficient on HIV must be multiplied by the average number of years between the two periods of measurement in order to make its magnitude comparable to what was found in the cross-section. Specifically, there were on average 11 years between the 1990s and the 2000s surveys, so, to compare the coefficient from the synthetic panel regressions to the coefficient from the cross-section, the coefficient from the synthetic panel regression must be multiplied by 11, which one can approximate by multiplying by 10.

In order to estimate the impact of HIV on the skill premium, I use the same first-difference specification, but I add schooling terciles dummies, and I interact these dummies with HIV.

[13] The first difference specification was chosen because there are only two periods of data. With only two periods of data, one cannot identify a fixed effect model with a lagged dependent variable. Since I'd like to check for the impact of including the lag dependent variable, the first difference model is useful. Additionally, the first difference specification allows me to control for country specific wage growth rates by including country fixed effects.

Instead of collapsing the employee data by region and year, I collapse it by region, year and schooling tercile. The choice of terciles was made to provide sufficient flexibility while limiting the number of schooling categories. Indeed, having many schooling categories would be problematic when collapsing data for the synthetic panel estimation, as it would lead to too few observations by region*year*schooling cell.

Synthetic cohort regressions allow me to control for *time-invariant* differences in regional outcomes. However, it is possible that unobserved *time-varying* factors drive both regional outcomes and regional HIV prevalence, leading us to wrongly conclude that HIV prevalence causally affects outcomes. Past research has used instrumental variables strategies to assess causality (e.g. Oster, 2010). Given that male circumcision has been shown to reduce the probability of HIV transmission (Auvert et al., 2005; Wise, 2006), Ahuja et al. (2009) and Fortson (2011) have used male circumcision as an instrument for HIV. In this paper, I am investigating the impact of HIV on wages, and it is therefore important to consider whether the exclusion restriction is satisfied in this specific context. While it is difficult to rule out that male circumcision could affect wages through channels other than HIV, one can at least show that male circumcision is not correlated with key determinants of wages and human capital. In particular, Ahuja et al. (2009) show that the male circumcision rate is unrelated to countries' initial income, initial life expectancy or modernity.

Fortson (2011) instruments regional HIV prevalence with regional male circumcision rates based on self-reports in the DHS. However, she finds that the instrument is weak. Indeed, in my sample, circumcision rates do not differ much across regions in countries with very high or very low circumcision rates. Male circumcision is however a good predictor of HIV prevalence overall, i.e. both within and across countries. Since first-difference specifications exploit variation in HIV prevalence both within and across countries, I can instrument HIV prevalence with male circumcision rates with a fairly strong instrument (see first-stage F-statistics and underidentification tests in Table 5). Therefore, this instrumental variables specification allows me to test the robustness of my synthetic panel results to time-varying factors that may be correlated with both HIV and outcomes of interest such as wages.

C. MAIN RESULTS

1. The impact of HIV on wages and the skill premium

How can we compare the estimated coefficient on log HIV in wage regressions to the theoretical predictions? As I have demonstrated in section II.A, the simplest model predicts that the short-run wage elasticity of AIDS is equal to the capital share. For the countries in my sample, the capital share varies between 0.5 (Niger) and 0.83 (Ghana) (based on UNIDO data reported in Rodriguez and Ortega, 2006). Remember that to get the wage elasticity of AIDS, we must multiply the coefficient on log HIV by 10 in cross-sectional regressions, and by 10*10=100 in synthetic panel regressions. Thus, we expect the coefficient on log HIV to lie between 0.05 and 0.083 in cross-sectional regressions and between 0.005 and 0.0083 in synthetic panel regressions.

Table 2 examines the relationship between wages and HIV prevalence using the firm and employee samples. Using the firm cross-section, the relationship between average wages and HIV prevalence is positive and significant (col. 1). This relationship is still positive, but no longer statistically significant in the synthetic panel (col. 2). In column 3 and 4, I use employee level data. The advantage of this sample is that I can use hourly wages instead of yearly wages. Additionally, I can control for a number of individual employee characteristics. These characteristics include occupation dummies, age, age squared, tenure on the job, gender, schooling and union membership. The disadvantage of the sample is that, as mentioned above, it is not representative of employees overall. Similarly to what was found for the firm sample, the coefficient on HIV is significant and positive in column 3. The point estimate is also essentially the same as in the firm sample. In the synthetic panel (col. 4), the coefficient on HIV is larger than in the firm level data, and it is significant at the 10% level. The estimated coefficients in column 1-3 are very similar: they imply that the wage elasticity of AIDS is about 1.5. This point estimate is higher than theoretical predictions. However, the 90% confidence intervals of the estimates in columns 1-4 are fairly wide, and they all include the 0.5 to 0.83 theoretical range. Additionally, as mentioned in section II.C, recent empirical estimates suggest that the elasticity of substitution between capital and labor is less than one. If that is the case, then the impact of HIV on wages should be higher than the baseline Cobb-Douglas case, and therefore the point estimates I find may in fact be accurate. Overall, I conclude that the impact of HIV on average wages is in line with theoretical expectations, but is highly variable and not always significantly different from 0. As suggested in the theory section II.B, the impact of HIV on average wages may be small in the presence of dualism in the labor market for low skilled workers.

In Table 2 columns 5 and 6, I explore the impact of HIV on the skill premium. According to the theory presented above, if there is no dualism in the low skilled market, HIV should have no impact on the skill premium. If, on the other hand, there is dualism in the low skilled market, so that AIDS mortality only decreases high skilled labor supply to urban manufacturing firms, then the coefficient on the interaction between a higher-skill tercile dummy (i.e. second or third tercile of schooling) and the HIV prevalence rate can be interpreted as the inverse of the elasticity of substitution between the skilled group and the lowest skilled group (see section II.B) divided by 10 (this is again to account that HIV proxies for AIDS mortality in this specification). According to Acemoglu (2002), the elasticity of substitution between high and low skilled workers can be reasonably expected to lie between 1.4 and 2, which means that its inverse lies between 0.5 and 0.71. In my empirical application with schooling terciles, I expect the elasticity of substitution between the lowest and the middle schooling tercile to be rather higher than this number (groups with more similar skills are closer substitutes), and the elasticity of substitution between the lowest and the highest tercile to be rather lower than this number. As a result, I expect the interaction between the second tercile dummy and HIV to be around 0.05 or less in the cross-section and 0.005 or less in the synthetic panel, and I expect the interaction between the third tercile dummy and HIV to be around 0.071 or more in the cross-section and 0.0071 or more in the synthetic panel.

The coefficient on the interaction between HIV and the second tercile of education is 0.05, but it is not statistically significant (Table 2, col. 5). Note that the point estimate squares nicely with theoretical predications. The coefficient on the interaction with the third tercile of schooling[14] is a significant 0.15: it is consistent with theory for this coefficient to be larger than 0.071, but then it is indeed quite a bit larger. However, the lower bound of the 90% confidence interval is 0.08, which is much closer to the theoretical prediction. Moving on to the synthetic panel estimates in columns 6, the results follow a similar qualitative pattern, confirming that higher HIV rates are associated with a significantly larger skill premium for workers in the third tercile of schooling[15]. However, the synthetic panel estimates are much larger in magnitude than the cross-sectional estimates. One interesting result in columns 5-6 is that the coefficient on HIV is insignificant and much closer to 0 than in columns 1-4, which suggests that the impact of HIV

[14] If, instead of using terciles, one controls for years of schooling and an interaction of years of schooling with HIV, one finds that the interaction is positive and significant, i.e. that HIV increases the skill premium.

[15] The total impact of HIV on the wages of workers in the third tercile of schooling is 0.076+0.15=0.226 and is significant at 5% in the cross section (col. 5). In the synthetic panel (col. 6), this same impact is -0.01+0.089= 0.079 and also significant at 5%.

for the lowest skilled workers is smaller than for the average worker, and possibly 0. As discussed above, this result is consistent with dualism in the market for low-skilled workers. I will return to this point and provide additional results in section III.C.2. To conclude, I find that the impact of HIV on the skill premium is positive and statistically significant for the third tercile of education, and this result holds both in the cross-section and in the synthetic panel. Additionally, the size of the estimated coefficient is consistent with theory in the cross-section, even though it is larger than predicted in the synthetic panel. Finally, these results strongly suggest that the impact of HIV-AIDS on the labor supply to urban manufacturing differs by skill group; if that were not the case, then HIV should have had no impact on the skill premium (see section II.B).

2. Dual labor markets

One plausible explanation for the positive association between the skill premium and HIV is that there is a dual labor market for low skilled workers and a competitive labor market for high skilled workers, such that the effective labor supply of low skilled workers to the urban manufacturing sector is unaffected by AIDS mortality. This mechanism is difficult to test directly since the Enterprise Surveys and ARPED data do not cover the entire labor market but only the (mostly) formal urban manufacturing sector. Essentially, what we would like to see if this mechanism is at play is that the proportion of the labor force that works in the most informal lowest paid sector decreases in regions with higher HIV, and that the income of workers who stay in this most informal lower paid sector increases. Since I cannot test these implications directly, I will make a series of arguments for why the dual labor market hypothesis seems plausible.

In Sub-Saharan Africa, at least 70% of the labor force works informally (MDG data quoted in Bacchetta et al. 2009). Furthermore, the formal sector is subject to fairly constraining labor regulation (Pierre and Scarpetta, 2004). This factor, together with unionism[16] and the presence of foreign firms in manufacturing makes it more likely that wages in the formal manufacturing sector are set above market-clearing levels. Remarkably, the average wage in the 2000s cross-section of my data corresponds to more than twice the GDP per capita for these countries. By contrast, in the US, the average manufacturing wage is roughly equal to the GDP per capita in 2009 (BLS data). A recent article by Gunther and Launov (2009) shows that, in Cote d'Ivoire, which is one of the countries in my sample, the dual labor market hypothesis is consistent with data on the distribution of wages in the formal and the informal sectors[17]. Overall, these elements make it more likely that there is dualism in the labor market and wages are set above the market-clearing level in the urban manufacturing sector.

The dualism hypothesis is more relevant for low-skilled workers. It is well known that the probability of working in the informal sector strongly decreases with education (Perry et al., 2007). This means that very educated workers have a low probability of being informal, which implies that, if these highly educated workers die, formal sector firms cannot view the informal sector as a reserve of additional highly educated workers. In the working paper version of their article, Gunther and Launov show evidence suggesting that workers who are willing to work in the formal sector but are rationed out tend to have low education.

Finally, in Table 3, I perform some additional analysis to assess to what degree the results can be explained by labor market dualism. For this purpose, I use two strategies. First, a key

[16] The unionization rate in the employee sample varies between 25% in the 2000s and 36% in the 1990s (see Table 1).

[17] While previous literature has found little support for dualism (e.g. Magnac, 1991, and Pratap and Quintin, 2006), it typically used data on Latin American countries, where informality is lower than in Sub-Saharan Africa. Additionally, these other works have relied on a model with two sectors, one formal and one informal. By contrast, Gunther and Launov use a three sector model (formal, informal voluntary and informal involuntary) and find that this model fits the data from Cote d'Ivoire better than the two-sector model.

element in the dualism account of the results is that firms in my sample are mostly formal, and likely pay above market-clearing wages to low skilled workers. While the majority of firms in my sample are likely to be formal, formality is a continuum (Perry et al., 2007) and smaller firms are more likely to evade at least some regulations. Thus, smaller firms are more likely to be informal or partially informal. In practice, I choose 20 workers as a cut-off for small firms because this was the upper limit for the "small firm" strata in the survey design. If the dualism hypothesis holds, then the skill premium should increase less with HIV in smaller less formal firms than in larger more formal firms. Consistent with this hypothesis, HIV does not have an impact on the skill premium in small firms[18] (see interaction between HIV and the third tercile of schooling in Table 3 col. 1 and 2). HIV has a larger impact on the skill premium in large firms (positive interaction between HIV, the large firm dummy, and the third tercile of schooling in col. 1 and 2). However, the difference in the impact of HIV on the skill premium between small and large firms is only significant in the growth specification in column 2. Still, the point estimates for the difference in the impact of HIV on the skill premium between small and large firms are both positive and of similar magnitude in columns 1 and 2. Thus, HIV does not significantly increase the skill premium in less formal firms, and the impact of HIV on the skill premium is generally larger in more formal firms, which is consistent with the dualism hypothesis.

A second strategy I use to assess the plausibility of the dualism hypothesis as an explanation for the results is to examine the impact of rural-urban migration on the wages of workers in different schooling terciles. Rural migrants are more likely to be low skilled and employed in the informal sector. Therefore, I expect that, in urban regions with a higher share of recent rural

[18] I also checked whether HIV has a differential impact in majority foreign-owned firms vs. majority domestic-owned firms and did not find any significant difference.

migrants, there is more competition among workers in the informal sector, and the impact of HIV on the wages of low-skilled workers in manufacturing is closer to 0. I measure the share of recent rural migrants using the DHS: this is the share of adult residents in an urban area that immigrated from a rural area after 1990[19]. In the absence of rural-urban migration, HIV significantly increases the wages of low-skilled workers, as predicted by theory (positive coefficient on HIV in columns 3 and 4 of Table 3). As the share of rural-urban migrants increases, the impact of HIV on the wages of the lowest skill tercile decreases (negative and significant interaction between HIV and the share of rural migrants). For higher skilled workers, rural-urban migration does not dampen as much the positive impact of HIV on wages (positive interactions between HIV and the share of rural-urban migrants for the second and third terciles of schooling in Table 3, cols. 3-4). This exercise suggests that part of the reason why the estimates of the impact of HIV on wages are so noisy is that there are large differences in regional labor markets that affect the impact of HIV on wages. In particular, differences in rural-urban migration play an important role in determining the impact of HIV on the wages of low-skilled workers.

3. The impact of HIV on capital, sales, and hours worked

I now explore some additional outcomes besides wages that can help us understand the link between HIV and wages. In Table 4 columns 1-3, I examine the impact of HIV on the capital-labor ratio and the growth of the capital stock using firm-level data. In the short-run, as AIDS mortality diminishes labor supply, I expect the capital-labor ratio to increase. Theoretically, in the short-run, the impact should be one for one, therefore the coefficient in the cross-section

[19] I chose the 1990 cutoff to be aligned with the approximate beginning of the HIV epidemic. I also experimented with defining recent migrants as those who migrated 10 years prior to the survey date, and the results were essentially unaffected.

should be 0.1 and in the synthetic panel 0.01. I find that indeed higher HIV prevalence rates are associated with a significantly higher capital-labor ratio, and this is so both in the cross-section and in the synthetic panel (columns 1-2). The coefficient is very large: the lower bound of the 90% confidence interval is slightly higher than 0.1 for the cross-section and 0.01 for the synthetic panel. This suggests that HIV may have been associated with capital deepening above and beyond the immediate impact of AIDS mortality on the capital/labor ratio. Firms may thus have actively substituted capital for labor. In column 3, I examine whether the growth in the capital stock has been higher or lower in regions with higher HIV. Since the coefficient on HIV is not significantly different from 0, I cannot reject that the capital stock did not adjust to the HIV-related reduction in labor supply. In fact, the coefficient on HIV is positive, weakly suggesting that HIV was associated with an increase in investment. Overall, these results show that HIV was associated with an increase in the capital-labor ratio, as predicted by theory.

In Table 4 columns 4-6, I examine the impact of HIV on firms' sales. I find that HIV has no significant impact on labor productivity as measured by sales per worker, and this holds both in the cross-section and in the synthetic panel (columns 4-5). HIV does not significantly impact sales growth either, as can be seen in column 6. The absence of an impact of HIV on labor productivity may seem surprising given the morbidity effect of HIV. As mentioned above, previous research has documented a negative impact of HIV on individual workers' productivity. However, one has to remember that labor productivity is measured at the firm level. Given that HIV was associated with a large increase in the capital-labor ratio, measured labor productivity should be positively associated with HIV. Therefore, it is possible that the absence of an impact of HIV on labor productivity results from the countervailing effects of morbidity and the increase in the capital-labor ratio. Controlling for the growth of the capital-labor ratio in column 5 yields a coefficient on HIV that is still positive but more than 100 times smaller (result not shown).

Additionally, the lower bound of the 90% confidence interval for the coefficient on HIV is -0.033 when controlling for the growth of the capital-labor ratio, versus -0.018 in the specification in column 5. This suggests that the impact of HIV on labor productivity is overestimated due to the positive association between HIV and the capital-labor ratio.

In columns 7-8, I investigate the impact of HIV on hours worked using the employee sample. In the cross-section I find no significant impact of HIV on hours worked. In the synthetic panel however, HIV is significantly associated with lower hours worked. The coefficient implies that a 100% increase in HIV prevalence is associated with a 7% decrease in hours (remember that one has to multiply the coefficient by 10 since the growth is calculated over 10 years); at the sample means, this implies that going from 9% HIV prevalence to 18% HIV prevalence decreases weekly hours worked by 3 hours, which is a small effect.

4. Instrumenting HIV prevalence with the male circumcision rate

To test the robustness of my synthetic panel results to endogeneity issues, I instrument HIV prevalence by the male circumcision rate in all first-difference specifications from Table 2 and Table 4. The IV estimates of the impact of HIV on outcomes (Table 5, upper panel) are very similar to the OLS estimates. This similarity does not seem to be driven by a weak instrument problem: indeed, as can be seen in the first stage results in the bottom panel of Table 5, the male circumcision rate is strongly and negatively associated with HIV, and the F-statistics for the first stage are well above 10. While there is, to my knowledge, no formal test of weak instruments in the case of robust standard errors (Baum et al. 2007), the Kleibergen-Paap underidentification test rejects that the equation is underidentified[20] (Table 5). With respect to the substance of the

[20] The male circumcision rate cannot be reliably used as an instrument for within-country variation in HIV. Indeed, if we use the IV specification for the results based on the cross-section in Tables 2 and 4, the male circumcision rate almost always fails the Kleibergen-Paap underidentification test.

results, it is notable that, consistent with OLS results, HIV has a positive and significant impact on the skill premium (Table 5, col. 2), and on the capital-labor ratio growth (Table 5, col. 5). I conclude that my main results are robust to endogeneity issues, and the IV specifications strengthen the case for a causal impact of HIV on the capital-labor ratio and the skill premium.

D. ROBUSTNESS TESTS AND DISCUSSION

1. Additional specifications and robustness tests for the wage results

The impact of HIV on average wages is not very precisely estimated. One reason for this may be that, while the mortality effect of HIV increases wages, the morbidity effect of HIV goes in the opposite direction: workers who are sick with HIV-AIDS are less productive and hence may earn lower wages. We can partially account for this effect by controlling for HIV-related absenteeism at the firm level. Indeed, for a large subset of countries in the 2000s survey, firms were asked whether they suffer from high absenteeism due to HIV-AIDS. However, controlling for this variable in cross-sectional regressions does not affect the HIV coefficient (Table 6, col. 1). Thus, the estimates of the impact of HIV on wages are not affected by HIV-AIDS related absenteeism.

The main analysis clusters standard errors by region. However, one could argue that outcomes are correlated within countries. If I cluster standard errors by country instead of region in Table 2 (results not shown), I find that the results are unchanged in the sense that HIV coefficients that were statistically significant remain so, and those that were statistically insignificant remain so as well.

Another concern that one may raise is that the estimates reported in Table 2 are not representative of the experience of the average worker. This is because large firms have the same number of observations in my firm sample as small firms, which suggests that it may be useful to give larger firms more weight in regressions. However, because the surveys I use oversample large firms, such a weighting scheme results in overweighting large firms compared to the underlying sample of manufacturing firms. Still, I feel it is important to perform this robustness test. I weighted each observation by the number of firm-level employees, using Stata's analytic weights: this is because, in the firm sample, the wage is an average over all employees. For the firm synthetic panel analysis, I collapsed the data to get the sum of all wage bills, and the sum of all employment, and I calculated the average wage after collapsing. The results using weights are reported in Table 9 in the appendix. Both in the cross-section (col. 1-2) and the synthetic panel (col. 3), the impact of HIV on wages is negative and insignificant. Thus, the impact of HIV on wages appears to be more negative when larger firms are given a greater weight. Hence, part of the reason why the impact of HIV on average wages is imprecisely estimated may be that the impact of HIV on wages varies by firm size. This exercise also confirms that the impact of HIV on wages is smaller in larger firms that are more likely to be formal, consistent with the dual labor market hypothesis (see p. 24 and following).

In columns 2 and 3 of Table 6, I add the capital-labor ratio on the right-hand side in order to investigate the degree to which the positive impact of HIV on wages is due to the increase in the capital-labor ratio. As expected, I find that the capital-labor ratio has a significant and positive impact on wages, and the inclusion of the capital-labor ratio makes the coefficient on HIV smaller and insignificantly different from 0. This is consistent with the increase in the capital-labor ratio being the main channel behind the positive impact of HIV on average wages.

The main analysis uses HIV prevalence at the regional level, including both rural and urban areas. However, the ARPED and Enterprise Survey data are drawn exclusively from urban areas, so one might want to restrict the calculation of HIV prevalence (from DHS data) to adults living in urban areas. However, calculating prevalence separately for urban and rural areas within each region is problematic because the small sample size introduces a good deal of sampling variability; in fact, the DHS documentation specifically states that HIV prevalence can be robustly estimated at the regional level and by urban and rural areas nationally, but not for urban and rural populations at the regional level. Therefore, I choose instead to control for the share of urban residents in the HIV testing sample. I have also experimented with regressing wages on HIV rates calculated for urban populations only (results not reported here): the coefficient on urban HIV prevalence was always very similar to the coefficient on regional HIV prevalence when controlling for the share of urban residents in the HIV test sample. In column 4 of Table 6, I add a control for the urban share to the regression of average wages at the firm level on HIV in the cross-section. The inclusion of this control makes the coefficient of HIV insignificantly different from 0, confirming that the estimates of the impact of HIV on average wages are fragile. On the other hand, one may argue that the HIV rate at the regional level better represents the labor market conditions. Indeed, rural-urban migration is high: on average, 24 to 30% of the residents in the cities in our sample have migrated from rural areas in 1990 or more recently (Table 1).

In columns 5-8 of Table 6, I examine the robustness of the impact of HIV on the skill premium. In column 5, I include the urban share in the cross-sectional specification. While the urban share coefficient is positive and highly significant, the coefficient on the interaction between HIV and the third tercile of schooling is unaffected. Including the urban share in the synthetic panel regressions does not affect the results either (col. 6); additionally, the coefficient on urban share itself is not significantly different from 0. In columns 7 and 8, I experiment with including country fixed effects in the synthetic panel regressions; this allows for country-specific wage growth rates. The inclusion of country fixed effects in column 7 does not substantially impact the results. Additionally, in column 8, I include lagged regional sales growth (the sample size shrinks slightly because this variable is not always available). Indeed, one may be concerned that synthetic panel results are driven by trends in wage growth that preceded the advent of the HIV epidemic. In particular, it is possible that regions that were growing more in the late 1980s and early 1990s subsequently experienced both higher growth in the skill premium and higher HIV infection rates. Unfortunately, I do not have information on wages prior to the early 1990s. However, wages are positively correlated with sales, and so sales growth may plausibly proxy for wage growth. Average regional sales growth was thus computed using a question about sales in the previous few years that was asked of firms in the 1990s. Still, the inclusion of this variable does not affect the estimated impact of HIV on the skill premium. Overall, I conclude that the positive impact of HIV on the skill premium is strong and very robust to alternative specifications.

Finally, I perform a falsification exercise where I repeat the specifications from Table 2, columns 1, 3 and 5, but using the ARPED data for firms and workers in the early 1990's. I use on the right-hand side HIV in the 2000. I find that HIV in the 2000 does not have a significant effect on wages and the skill premium in the 1990s (results not reproduced here). This suggests that the HIV epidemic likely had a causal impact on wages and the skill premium.

2. Why does HIV affect the skill premium?

I have argued above that dualism in the low skilled labor market is a plausible explanation for this result. Here I examine and rule out some alternative interpretations.

One interpretation is that this result does not in fact reflect a differential impact of HIV on the labor supply of different skill groups. Indeed, it could be that higher skilled workers are more mobile, and that they dislike living in regions with high HIV. If this is so, then the higher skill premium in regions with higher HIV may reflect a compensating wage differential. This explanation is however unlikely since results based on national HIV rates look very similar (results not reproduced here), and mobility between countries is likely to be much lower than between regions.

A second interpretation is that the skill premium is affected by HIV because high skilled workers died of AIDS at a higher rate[21] than low skilled workers. Specifically, I examine whether AIDS mortality was higher for workers with more than 13 years of schooling. I conduct two supplementary analyses. First, I estimate HIV rates by education tercile in the cross-section. Second, I use multiple rounds of the DHS to assess how the education distribution has changed over time.

First, I use data on individual HIV status – which is only available for the 2000s – to estimate HIV rates separately by education tercile. Using data from Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania, Fortson (2008) found that HIV rates tend to peak among those completing primary school, with generally lower rates among the least and most educated. However, because my employment data come from surveys of manufacturing firms in urban areas from a large group of countries, the patterns of HIV prevalence by educational attainment may be somewhat different in a population which looks more like my sample. Therefore, I use DHS data on HIV status and educational attainment to present a picture of the distribution of HIV by education tercile. In particular, I estimate HIV prevalence by country among adults aged 15-49 living in urban areas, separately for each of three educational categories: completed 10 or fewer years of schooling, completed more than 10 but no more than 13 years of schooling, and completed more than 13 years of schooling. These categories mirror the three education terciles

[21] High skilled workers may also have *lower* mortality rates due to greater access of to ARV therapy. However, if ARV differentially *reduces* the mortality of high-skilled workers, this would dampen the positive impact of HIV on the skill premium. Moreover, ARV access was very limited in Sub-Sahran Africa during the time of my sample. In 2003, ARV coverage was only 2%, and it was still only 17% in 2005 (UNAIDS/WHO, 2006).

used in my synthetic panel analysis. The results are presented in Table 7. Countries with total HIV prevalence rates above five percent are highlighted.

As Table 7 shows, in countries with high HIV rates, HIV prevalence among those with more than 13 years of schooling is generally similar or lower than it is for those with less schooling. Zambia is the only exception to this general pattern: prevalence among urban adults aged 15-49 with more than 13 years of schooling is 20.93%, whereas prevalence among those with between 10 and 13 years of schooling is 15.43% and among those with 10 or fewer years of schooling is 13.75%. In other high prevalence countries, prevalence among adults with more than 13 years of schooling is similar or lower than prevalence among those with 10 or fewer years of schooling. Additionally, when regressing the HIV rate on the schooling tercile dummies within this sample, I find there is no significant difference in HIV prevalence between the first tercile of schooling and the second or third tercile. These results are consistent with the literature on the variation of HIV infection by education, which tends to find a non-monotonic relationship between HIV infection and education (see e.g. Iorio and Santaeulalia-Llopis, 2011).

Based on these estimates, we would expect if anything lower (not higher) mortality among adults in the third education tercile in the coming years. If the distribution of HIV infection by educational attainment has been stable over time, this evidence suggests that differential mortality does not explain the positive correlation between the skill premium and HIV prevalence.

However, current HIV rates may not be a good reflection of past mortality. It is possible that, in earlier cohorts, the distribution of HIV by education was quite different than it is today. By comparing multiple rounds of the DHS, I can estimate how the education distribution has changed over time, and how those changes are related to HIV rates. If adults with more schooling have become less numerous over time in areas with higher levels of HIV, this could provide evidence of differential mortality by education, possibly explaining the estimated positive association between HIV prevalence and the skill premium.

To assess how the distribution of schooling has changed over time across areas with different levels of HIV, I will draw on the household roster from multiple rounds of DHS data. Rather than using data on HIV testing, I will document changes in the distribution of educational attainment for a fixed cohort across rounds of the survey.

The analysis is restricted to countries that are in my synthetic panel sample and have multiple rounds of DHS surveys, and at least one during the early- to mid-1990s. Table 8 shows that the fraction of adults in the highest educational category (more than 13 years of schooling) has increased over time in all but Côte d'Ivoire. The correlation between overall HIV prevalence in the 2000s and the change in the fraction of adults with more than 13 years of schooling is -0.19 and insignificant. This suggests that the positive association between HIV and the skill premium is not driven by changes in the relative supply of educated workers[22] due to mortality.

A fourth interpretation of the impact of the HIV on the skill premium is that, in reaction to the HIV epidemic, firms have started using more capital-intensive technologies. If more skilled workers can better use this technology, HIV may have differentially increased the demand for high-skilled workers, thus explaining the increase in the skill premium. This explanation is consistent with the results from Table 4 showing that HIV is associated with an increase in the capital-labor ratio. However, when I add a control for the capital-labor ratio and an interaction of this control with schooling terciles in columns 5 and 6 from Table 2, I still find that HIV has a

[22] This does not contradict the results from Fortson (2011) showing that, in regions with higher HIV prevalence, cohorts born after the beginning of the HIV epidemic (after 1980) have lower educational attainment than cohorts born before 1980. Indeed, less than 10% of my sample of workers was born after 1980 and so the reduction in human capital for new generations is unlikely to explain the results in my sample. Indeed, if I exclude people born after 1980 from the specification in Table 2, col. 5, I find the impact of HIV on the skill premium to be essentially unchanged.

positive and significant impact on the skill premium, with a magnitude similar to my basic results (appendix Table 11). Furthermore, while the capital-labor ratio has a positive impact on wages, there is no significant interaction between the capital-labor ratio and the third decile of schooling: this means that the wages of high-skilled workers are not differentially higher in firms or regions that use a more capital-intensive technology. Therefore, the data does not support the interpretation that HIV increases the skill premium because the use of more capital-intensive technologies increases the demand for high skilled workers.

I have argued above that dualism in the market for low skilled workers can account for the positive impact of HIV on the skill premium. I have just shown that a number of plausible alternative interpretations of this result are inconsistent with the data. Therefore, I conclude that the positive impact of HIV prevalence on the skill premium is consistent with theory under the assumption of a dual labor market.

IV. CONCLUSION

The HIV/AIDS epidemic has been the cause of a large increase in prime-age adult mortality in Sub-Saharan Africa. Given poverty and low growth rates in the region, economists have been concerned about the impact of the epidemic on the economic outlook of the worst-affected countries. In particular, it has been argued that, despite the terrible death toll, the HIV/AIDS epidemic may offer a chance for growth in the region by decreasing the population, reducing pressure on resources, and increasing the wages of survivors (Young, 2005).

This study uses micro data on urban manufacturing firms and their workers to identify the impact of the HIV/AIDS epidemic on labor market outcomes, including wages. Using within country and across time variation, I find that the impact of HIV prevalence on wages is positive but imprecisely estimated. As expected, I find that HIV significantly increases the capital-labor ratio. Additionally, I uncover a new stylized fact, namely that higher HIV prevalence is significantly associated with a higher skill premium both in the cross-section and in the synthetic panel. When instrumenting HIV prevalence by the male circumcision rate, I continue to find a significant positive impact of HIV on the capital-labor ratio and the skill premium, strengthening the case for a causal impact of HIV. The HIV epidemic thus tends to increase wage inequality in Sub-Saharan Africa, and that this phenomenon is most likely explained by labor market dualism.

The results of this study suggest that the low measured impact of HIV on wages or income per capita in previous macroeconomic studies could be due to the high levels of labor market informality in Sub-Saharan African countries. As AIDS mortality decreases labor supply, some workers move from informal employment, either in agriculture or in the informal urban sector, to the higher-paying formal sector. While this compositional effect should increase the wages of the average worker, wages in the formal sector may not increase much. Because wages and incomes in the formal sector are better measured, and because agricultural income is particularly difficult to measure, measured wages and incomes are likely overweighing more formal economic activities. This may explain why it is difficult to find any positive relationship between measured wages or incomes and HIV prevalence. Further work is needed to confirm that dualism in the labor market is indeed the key reason why it is difficult to estimate a robust positive relationship between AIDS mortality and the average wages of survivors.

References

Acemoglu, Daron 2002. "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, American Economic Association, vol. 40(1), pages 7-72, March.

Acemoglu, Daron, Autor David H., and David Lyle, 2004. "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury," *Journal of Political Economy*, University of Chicago Press, vol. 112(3), pages 497-551, June.

Ahuja, Amrita, Brian Wendell, and Eric Werker. 2009. "Male Circumcision and AIDS: The Macroeconomic Impact of a Health Crisis." Harvard Business School Working Paper 07–025.

Antras, Pol 2004. "Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution," *The B.E. Journal of Macroeconomics*, Berkeley Electronic Press, vol. 0(1).

Auvert, Bertran, Dirk Taljaard, Emmanuel Lagarde, Joëlle Sobngwi-Tambekou, Rémi Sitta, and Adrian Puren (2005), "Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial," *PLoS Med*, 2(11), e298.

Bachetta, Marc and Ernst Ekkehard and Juana P. Bustamante, "Globalization and Informal jobs in Developing Countries - A joint study from the International Labour Organization and the WTO", ILO and WTO 2009.

Bloom, David E., and Ajay S. Mahal (1997), "Does the AIDS Epidemic Threaten Economic Growth?" *Journal of Econometrics*, 77(1), pp. 105-124.

Bonnel, René (2000), "HIV/AIDS and Economic Growth: A Global Perspective," *South African Journal of Economics*, 68(5), pp. 360-379.

Borjas, George J. 2003. "The Labor Demand Curve Is Downward Sloping: Reexamining The Impact Of Immigration On The Labor Market," *The Quarterly Journal of Economics*, MIT Press, vol. 118(4), pages 1335-1374, November.

Borjas, George J. 2009. "The Analytics of the Wage Effect of Immigration," NBER Working Papers 14796, National Bureau of Economic Research, Inc.

Card, David 1990. "The impact of the Mariel boatlift on the Miami labor market," *Industrial and Labor Relations Review*, ILR Review, ILR School, Cornell University, vol. 43(2), pages 245-257, January.

Card, David 2005. "Is the New Immigration Really so Bad?," *Economic Journal*, Royal Economic Society, vol. 115(507), pages F300-F323, November.

Chou R, Huffman LH, Fu R, Smits AK, Korthuis PT (July 2005). "Screening for HIV: a review of the evidence for the U.S. Preventive Services Task Force" . *Ann. Intern. Med.* 143 (1): 55–73.

DiNardo, John 2002. "Propensity Score Reweighting and Changes in Wage Distributions", working paper.

Fields, Gary S., 2005. "A guide to multisector labor market models," Social Protection Discussion Papers 32547, The World Bank.

Fortson, Jane (2008), "The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS," *Demography*, 45(2), pp. 303-322.

Fortson, Jane (2009), "HIV/AIDS and Fertility," *American Economic Journal: Applied Economics,* 1(3), pp. 170-194.

Fortson, Jane (2011), "Mortality Risk and Human Capital Investment," *Review of Economics and Statistics*, 93(1), February 2011, pp. 1-15.

Fox, Matthew P., Sydney Rosen, William B. MacLeod, Monique Wasunna, Margaret Bii, Ginamarie Foglia, and Jonathan L. Simon (2004), "The Impact of HIV/AIDS on Labour Productivity in Kenya," *Tropical Medicine and International Health*, 9(3), pp. 318-324.

Gunther, Isabel and Andrey Launov, (2009). "Informal employment in developing countries: Opportunity or last resort?", working paper.

Habyarimana, James, Bekezela Mbakile, and Cristian Pop-Eleches (2010), "The Impact of HIV/AIDS and ARV Treatment on Worker Absenteeism: Implications for African Firms" *Journal of Human Resources*, 45(4), pages 809-839.

Iorio, Daniela and Raul Santaeulalia-Llopis (2011), "Education, HIV Status and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter?", working paper.

Juhn, Chinhui, Sebnem Kalemli-Ozcan, Belgi Turan (2008), "HIV and Fertility in Africa: First Evidence from Population Based Surveys", NBER Working Paper No. 14248.

Juselius, Mikael, 2008. "Long-run relationships between labor and capital: Indirect evidence on the elasticity of substitution," *Journal of Macroeconomics*, Elsevier, vol. 30(2), pages 739-756, June.

Kalemli-Ozcan, Sebnem (2010), "AIDS, 'Reversal' of the Demographic Transition and Economic Development: Evidence from Africa," *Journal of Population Economics*, forthcoming.

Kalemli-Ozcan, Sebnem and Belgi Turan (2010), "HIV and Fertility Revisited", *Journal of Development Economics*, forthcoming.

Larson, Bruce A., Matthew P. Fox, Sydney Rosen, Margaret Bii, Carolyne Sigei, Douglas Shaffer, Fredrick Sawe, Monique Wasunna, and Jonathon L. Simon (2008), "Early Effects of Antiretroviral Therapy on Work Performance: Preliminary Results From a Cohort Study of Kenyan Agricultural Workers," *AIDS*, 22(3), pp. 421-425.

Magnac, Thierry, 1991. “Segmented or Competitive Labor Markets?”, *Econometrica*, Vol. 59(1), pp.165-187.

Mishra, Vinod, Martin Vaessen, J. Ties Boerma, Fred Arnold, Ann Way, Bernard Barrere, Anne Cross, Rathavuth Hong, and Jasbir Sangha (2006), “HIV Testing in National Population-Based Surveys: Experiences from the Demographic and Health Surveys,” *Bulletin of the World Health Organization*, 84(7), pp. 537-545.

Mishra, Vinod, Amy Medley, Rathavuth Hong, Yuan Gu, and Bryant Robey (2009), “Levels and Spread of HIV Seroprevalence and Associated Factors: Evidence from National Household Surveys,” DHS Comparative Reports 22, http://www.measuredhs.com/pubs/pdf/CR22/CR22.pdf.

Oster, Emily (2010), “Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa”, *Journal of the European Economic Association*, forthcoming.

Perry, Guillermo E. and William F. Maloney, Omar S. Arias, Pablo Fajnzylber, Andrew D. Mason, Jaime Saavedra-Chanduvi, “Informality, exit and exclusion”, World Bank, 2007.

Pierre, Gaelle, Stefano Scarpetta, “Employment Regulations through the Eyes of Employers: Do they Matter and How Do Firms Respond to Them?,” Policy Research Working Paper Series 3463, The World Bank, 2004.

Pratap, S. and Quintin, E., 2006. “Are labor markets segmented in developing countries? A semiparametric approach,” *European Economic Review*, Vol. 50(7), pp.1817-1841.

Rodríguez, Francisco and Daniel Ortega, 2006. "Are capital shares higher in poor countries? Evidence from Industrial Surveys," Wesleyan Economics Working Papers 2006-023, Wesleyan University, Department of Economics.

Santaeulàlia-Llopis, Raul (2008), “Aggregate Effects of AIDS on Development,” working paper, Washington University.

Thirumurthy, Harsha, Joshua Graff Zivin, and Markus Goldstein (2008), “The Economic Impact of AIDS Treatment: Labor Supply in Western Kenya,” *Journal of Human Resources*, 43(3), pp. 511-552.

Thirumurthy H, Jafri A, Srinivas G, Arumugam V, Saravanan M, Angappan SK, Ponnumsamy M, Raghavan S, Merson M, Kallolikar S. (2011), “Two-year impacts on employment and income among adults receiving antiretroviral therapy in Tamil Nadu, India: a cohort study”, *AIDS*, in press.

UNAIDS (2008), “2008 Report on the Global AIDS Epidemic,” Joint United Nations Programme on HIV/AIDS (UNAIDS) 2008.

UNAIDS/WHO (2006), “Progress on global access to HIV antiretroviral therapy a report on "3 by 5", June 2005 and beyond”.

Van Biesebroeck, Johannes (2005), "Firm Size Matters: Growth and Productivity Growth in African Manufacturing", *Economic Development and Cultural Change*, 53(3), April 2005, pp. 545-584.

Verbeek, Marno (2007), "Pseudo Panels and Repeated Cross-Sections", working paper.

Wise, Jacqui, ''Demand for Male Circumcision Rises in a Bid to Prevent HIV,'' Bulletin of the World Health Organization 84:7 (2006), 509–511.

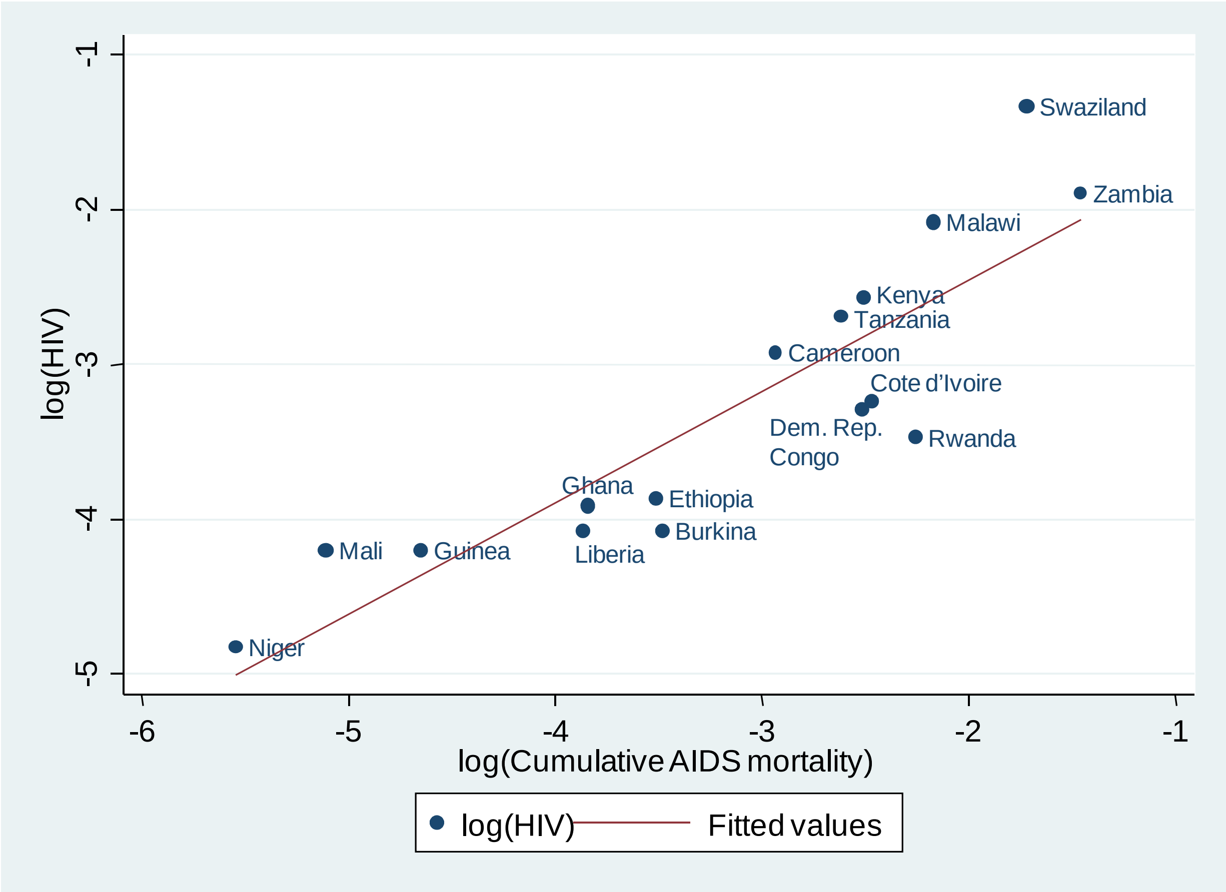
Young, Alwyn (2005), "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations," *Quarterly Journal of Economics*, 120(2), pp. 423-466.

Table 1: Summary statistics

2000's data:	Obs.	Mean	Std. Dev.	Min	Max
Firm cross-section					
HIV prevalence	3868	0.074	0.068	0.001	0.288
Number of HIV testing regions	3868	56	0	56	56
Average yearly wage (ppp)	3298	4846.47	7252.43	64.72	115207.50
Number of workers	3752	8075.03	39798.35	0.00	1507800.00
Sales (ppp)	3368	6049420.00	25736930.00	480.20	454584300.00
Productivity - sales per worker (ppp)	3354	65829.74	160871.40	480.20	2545480.00
Capital-Labor ratio (ppp)	3066	30516.75	206518.00	0.00	7526847.00
Share of rural migrants	3090	0.24	0.14	0.05	0.83
Employee cross-section					
HIV prevalence	10409	0.087	0.070	0.003	0.288
Number of HIV testing regions	10409	39	0	39	39
Hourly wage (ppp)	7256	3.57	42.91	0.00	2239.42
Hours worked	10001	48.34	10.47	0.00	84.00
Schooling	9959	10.62	4.10	0.00	30.00
Age	10384	34.59	9.67	0.00	92.00
Tenure	10240	6.71	6.59	0.00	69.50
Female	10414	0.21	0.41	0.00	1.00
Union status	10480	0.25	0.44	0.00	1.00
Share of rural migrants	8896	0.30	0.12	0.05	0.54
1990's data:					
Firm cross-section					
Number of workers	1588	108.63	397.27	0.00	8345.00
Average yearly wage (ppp)	1458	2884.74	4288.02	0.00	30391.93
Sales (ppp)	1525	4612356.00	22469110.00	35.50	568312900.00
Productivity - sales per worker (ppp)	1488	26401.78	49409.53	0.40	543212.50
Capital-Labor ratio (ppp)	1195	35186.05	142278.90	0.00	4356775.00
Employee cross-section					
Hourly wage (ppp)	5635	1.86	3.14	0.04	62.89
Hours worked	6014	44.75	7.34	4.00	84.00
Schooling	6276	12.21	4.20	0.00	36.00
Age	6683	33.38	9.91	12.00	71.00
Tenure	6660	6.80	6.84	0.00	48.00
Female	6677	0.16	0.37	0.00	1.00
Union status	5575	0.36	0.48	0.00	1.00

Note: PPP values are in 2005 dollars.
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.

Figure 1: The relationship between HIV prevalence and AIDS mortality



Note: Cumulative AIDS mortality is defined as total AIDS deaths between 1990 and the year of the 2000s firm survey I use divided by adult population in 1989. HIV prevalence is as of the year of the 2000s firm survey I use. Source: UNAIDS and United Nations data.

Table 2: HIV, wages and the skill premium

	Firm level Average Wage		Employee level Total Hourly Wage			
	Levels	Growth	Levels	Growth	Levels	Growth
	(1)	(2)	(3)	(4)	(5)	(6)
HIV	0.144**	0.016	0.150*	0.037*	0.076	-0.010
	(0.067)	(0.010)	(0.082)	(0.020)	(0.089)	(0.021)
HIV*2nd tercile school					0.050	0.025
					(0.049)	(0.020)
HIV*3rd tercile school					0.150**	0.089***
					(0.062)	(0.026)
Age			0.047***		0.052***	
			(0.015)		(0.015)	
Age squared/100			-0.039**		-0.045**	
			(0.018)		(0.018)	
Tenure			0.005		0.004	
			(0.003)		(0.003)	
Female			0.017		-0.008	
			(0.043)		(0.045)	
Union membership			0.018		0.047	
			(0.058)		(0.056)	
Schooling			0.072***			
			(0.013)			
2nd tercile school					0.424***	0.084
					(0.138)	(0.061)
3rd tercile school					1.146***	0.319***
					(0.174)	(0.062)
Country fixed effects	X		X		X	
Industry fixed effects	X		X		X	
Occupation fixed effects			X		X	
Observations	2,109	18	7,004	17	7,252	47
R-squared	0.867	0.061	0.828	0.133	0.827	0.239

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: In column 1, the data is from 2000s firm surveys, standard errors are clustered by region. In column 2, the data is from both 2000s and 1990s firm surveys; it is collapsed by region and year (see text for more details). In columns 3 and 5, the data is from 2000s employee surveys, standard errors are clustered by region. In column 4, the data is from both 2000s and 1990s employee surveys; it is collapsed by region and year. In column 6, the data is from both 2000s and 1990s employee surveys; it is collapsed by schooling tercile, region and year; standard errors are clustered by region.
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.

Table 3: HIV, skill premium, and labor market dualism

	By large (L>=20) vs. small firms		By share of rural migrants	
	Total Hourly Wage	Wage Growth	Total Hourly Wage	Wage Growth
	(1)	(2)	(3)	(4)
HIV	0.046	-0.011	0.821**	0.040*
	(0.109)	(0.024)	(0.348)	(0.015)
HIV*2nd tercile school	-0.040	-0.005	-0.115**	-0.019
	(0.050)	(0.032)	(0.052)	(0.009)
HIV*3rd tercile school	-0.056	-0.007	0.018	0.052
	(0.161)	(0.036)	(0.109)	(0.035)
HIV*Large firm	0.044	0.006		
	(0.077)	(0.034)		
HIV*Large firm*2nd tercile school	0.059	0.036		
	(0.070)	(0.050)		
HIV*Large firm*3rd tercile school	0.165	0.104*		
	(0.173)	(0.054)		
HIV*Share of rural migrants			-2.100**	-0.400*
			(0.937)	(0.158)
HIV*Share of rural migrants *2nd tercile school			0.619**	0.308**
			(0.233)	(0.097)
HIV*Share of rural migrants *3rd tercile school			0.491	0.313
			(0.442)	(0.234)
Country fixed effects	X		X	
Industry fixed effects	X		X	
Worker characteristics	X		X	
Occupation fixed effects	X		X	
Observations	7,175	84	5,918	31
R-squared	0.831	0.240	0.840	0.754

Robust standard errors clusterded by region in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: All columns include schooling tercile dummies. Columns 1 and 3 control for the following worker characteristics: age, age^2, tenure, female, union membership. Columns 1 and 2 include interactions of the large firm dummy with schooling tercile dummies as well as a large firm dummy. Columns 3 and 4 include interactions of the share of rural migrants with schooling tercile dummies as well as the share of rural migrants. In columns 1 and 3, the data is from 2000s employee surveys. In columns 2 and 4, the data is from both 2000s and 1990s employee surveys. In column 2, the data is collapsed by schooling tercile, firm size (large and small), region and year. In column 4, the data is collapsed by schooling tercile, region and year. Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.

Table 4: HIV, capital, sales, and hours worked

	Firm-level						Employee-level	
	Capital-labor ratio	Capital-labor ratio growth	Capital growth	Sales per worker	Sales per worker growth	Sales growth	Hours worked	Hours worked growth
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HIV	0.350***	0.050***	0.010	0.070	0.011	-0.019	-0.015	-0.007***
	(0.115)	(0.011)	(0.029)	(0.078)	(0.017)	(0.032)	(0.016)	(0.002)
Age							0.001	
							(0.002)	
Age squared/100							-0.001	
							(0.002)	
Tenure							-0.002***	
							(0.001)	
Female							-0.031*	
							(0.016)	
Union membership							-0.028**	
							(0.012)	
Schooling							-0.001	
							(0.001)	
Country fixed effects	X			X			X	
Industry fixed effects	X			X			X	
Occupation fixed effects							X	
Observations	2,147	18	18	2,167	18	18	8,309	17
R-squared	0.654	0.158	0.003	0.777	0.023	0.016	0.068	0.466

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: In columns 1 and 4, the data is from 2000s firm surveys, standard errors are clustered by region. In columns 2, 3, 5 and 6, the data is from both 2000s and 1990s firm surveys; it is collapsed by region and year (see text for more details). In column 7, the data is from 2000s employee surveys, standard errors are clustered by region. In column 8, the data is from both 2000s and 1990s employee surveys; it is collapsed by region and year.
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.

Table 5: HIV, wages, capital, sales, and hours worked: instrumenting HIV with male circumcision

	Employee level			Firm level				
	Total hourly wage growth	Total hourly wage growth	Hours worked growth	Average wage growth	Capital-labor ratio growth	Capital growth	Sales per worker growth	Sales growth
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HIV	0.052**	0.014	-0.005***	0.013	0.079***	0.041	-0.015	-0.049
	(0.024)	(0.026)	(0.002)	(0.011)	(0.023)	(0.041)	(0.014)	(0.031)
HIV*2nd tercile school		-0.016						
		(0.046)						
HIV*3rd tercile school		0.100***						
		(0.019)						
2nd tercile school		-0.024						
		(0.122)						
3rd tercile school		0.348***						
		(0.042)						
				First stage: dependent variable is HIV				
Male cicrcumcision rate	-1.529***	-1.538***	-1.529***	-1.589***	-1.589***	-1.589***	-1.589***	-1.589***
	(0.309)	(0.339)	(0.309)	(0.209)	(0.209)	(0.209)	(0.209)	(0.209)
Male cicrcumcision rate *2nd tercile school		0.094						
		(0.461)						
Male cicrcumcision rate *3rd tercile school		-0.106						
		(0.432)						
2nd tercile school		-0.023						
		(0.191)						
3rd tercile school		0.021						
		(0.187)						
First-stage R-squared	0.537	0.553	0.537	0.657	0.657	0.657	0.657	0.657
F-statistic of excluded instruments	24.57	12.8; 6.8; 12	24.57	57.64	57.64	57.64	57.64	57.64
Kleibergen-Paap underidentification test P-value	0.095	0.099	0.095	0.051	0.051	0.051	0.051	0.051
Observations	17	47	17	18	18	18	18	18

Robust standard errors in parentheses, clustered by region in col. 2
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: In all columns in the top panel, HIV is instrumented with the male circumcision rate; in column 2, the interactions of HIV with schooling terciles are instrumented by the interactions of the male circumcision rate with schooling terciles. In columns 1-3, the data is from both 2000s and 1990s employee surveys. In columns 4-8, the data is from both 2000s and 1990s firm surveys. In columns 1 and 3-8, the data is collapsed by region and year (see text for more details). In column 2, the data is collapsed by schooling tercile, region and year.
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.

Table 6: HIV and wages: additional specifications

	Firm-level Average wage				Employee-level Total Hourly Wage			
	Levels	Levels	Growth	Levels	Levels	Growth	Growth	Growth
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HIV	0.144**	0.058	0.005	-0.086	-0.072	-0.008	-0.018	-0.085
	(0.066)	(0.071)	(0.011)	(0.052)	(0.082)	(0.019)	(0.056)	(0.064)
HIV*2nd tercile school					0.051	0.025	0.020	0.013
					(0.048)	(0.020)	(0.023)	(0.031)
HIV*3rd tercile school					0.156**	0.084***	0.085***	0.101***
					(0.062)	(0.023)	(0.022)	(0.031)
High HIV absenteeism	0.180*							
	(0.102)							
Capital/Labor		0.179***						
		(0.017)						
Growth of Capital/Labor			0.219*					
			(0.104)					
Urban share				0.491***	0.343***	-0.057	0.015	0.066
				(0.128)	(0.107)	(0.075)	(0.051)	(0.049)
Lagged sales growth								0.124
								(0.221)
Age					0.052***			
					(0.015)			
Age squared/100					-0.000**			
					(0.000)			
Tenure					0.004			
					(0.003)			
Female					-0.011			
					(0.046)			
Union membership					0.053			
					(0.057)			
2nd tercile school					0.420***	0.083	0.073	0.059
					(0.135)	(0.062)	(0.063)	(0.078)
3rd tercile school					1.161***	0.308***	0.331***	0.369***
					(0.174)	(0.054)	(0.064)	(0.082)
Country fixed effects	X	X		X	X		X	X
Industry fixed effects	X	X		X	X			
Occupation fixed effects					X			
Observations	2,109	1,983	18	2,109	7,252	47	47	40
R-squared	0.867	0.888	0.222	0.868	0.828	0.277	0.768	0.783

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: In columns 1, 2 and 4, the data is from 2000s firm surveys, and standard errors are clustered by region. In column 3, the data is from both 2000s and 1990s firm surveys; it is collapsed by region and year (see text for more details). In column 5, the data is from 2000s employee surveys, standard errors are clustered by region. In columns 6-8 , the data is from both 2000s and 1990s employee surveys; it is collapsed by schooling tercile, region and year. In column 6, standard errors are clustered by region. In all columns, urban share is the share of urban residents in the HIV testing sample.
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.

Table 7: HIV prevalence by schooling tercile, 2000s

Country	HIV Prevalence, Adults 15-49			
	All	Education ≤ 10	10 < Education ≤ 13	Education > 13
Burkina Faso	1.84%	1.70%	2.82%	8.43%
Cameroon	5.50%	5.42%	6.39%	4.49%
Congo, Democratic Republic	1.30%	1.35%	1.29%	0.35%
Cote d'Ivoire	5.35%	5.19%	6.24%	6.85%
Ethiopia	1.47%	1.36%	4.91%	0.94%
Ghana	2.17%	2.28%	0.87%	2.43%
Guinea	1.55%	1.51%	2.02%	2.61%
Kenya	6.76%	6.72%	6.80%	7.25%
Liberia	1.60%	1.39%	3.02%	1.72%
Malawi	11.79%	11.37%	16.29%	9.96%
Mali	1.21%	1.15%	1.79%	3.64%
Niger	0.73%	0.69%	1.39%	2.50%
Rwanda	3.07%	2.86%	6.90%	4.75%
Swaziland	25.95%	25.86%	27.09%	23.66%
Tanzania	7.03%	7.01%	7.76%	6.32%
Zambia	14.27%	13.75%	15.43%	20.93%

Notes: Demographic and Health Surveys from Burkina Faso (2003), Cameroon (2004), Côte d'Ivoire (2005), Democratic Republic of the Congo (2007), Ethiopia (2005), Ghana (2003), Guinea (2005), Kenya (2003), Liberia (2007), Malawi (2004), Mali (2006), Niger (2006), Rwanda (2005), Swaziland (2006/7), Tanzania (2003/4), and Zambia (2007). HIV rates are calculated among adults aged 15-49.

Table 8: HIV prevalence and the distribution of education over time

Country	HIV	Fraction of Adults Born 1950-1959 in Each Education Category					
		1990s Round			2000s Round		
		Educ ≤ 10	10 < Educ ≤ 13	Educ > 13	Educ ≤ 10	10 < Educ ≤ 13	Educ > 13
Cameroon	5.5%	91.2%	6.0%	2.7%	89.4%	7.1%	3.4%
Côte d'Ivoire	4.7%	91.3%	4.8%	3.8%	93.9%	3.1%	3.0%
Ghana	2.2%	85.9%	8.5%	5.5%	85.8%	3.1%	11.1%
Kenya	6.8%	80.4%	18.5%	1.1%	79.2%	16.2%	4.6%
Tanzania	7.0%	94.9%	4.6%	0.5%	93.8%	2.8%	3.5%
Zambia	15.6%	86.5%	8.3%	5.2%	88.3%	5.7%	6.0%

Notes: Demographic and Health Surveys from Cameroon (1991, 2004), Côte d'Ivoire (1994, 2005), Ghana (1993, 2003), Kenya (1993, 2003), Tanzania (1991/2, 2003/4), and Zambia (1992, 2001/2). HIV rates are calculated among adults aged 15-49 using data from Cameroon (2004), Côte d'Ivoire (2005), Ghana (2003), Kenya (2003), Tanzania (2003/4), and Zambia (2001/2). Table shows the percentage of adults born 1950-1959 in each educational category.

APPENDIX

Table 9: HIV and average wages, firm surveys, weighted by firm-level employment

	Average Wage		Average Wage Growth
	(1)	(2)	(3)
HIV	-0.050	-0.372	-0.002
	(0.213)	(0.236)	(0.017)
Urban share		0.637*	
		(0.338)	
Country fixed effects	X	X	
Industry fixed effects	X	X	
Observations	2109	2109	18
R-squared	0.813	0.815	0.007

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: In columns 1-2, the data is from 2000s firm surveys, standard errors are clustered by region. In column 3, the data is from both 2000s and 1990s firm surveys; it is collapsed by region and year (see text for more details). In all columns, urban share is the share of urban residents in the HIV testing sample, and observations are weighted by firm-level employment (see details in the text).
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.

Table 10: National HIV prevalence, AIDS mortality and wages, firm surveys

	Average Wage		Wage Growth	
	(1)	(2)	(3)	(4)
AIDS mortality	-2.571***		0.046	
	(0.564)		(0.071)	
HIV		-0.243***		0.005
		(0.076)		(0.008)
Industry fixed effects	X	X		
Observations	2,106	2,106	18	18
R-squared	0.076	0.091	0.006	0.007

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: In columns 1-2, the data is from 2000s firm surveys, standard errors are clustered by region, wages are expressed in 2005 PPP dollars. In columns 3-4, the data is from both 2000s and 1990s firm surveys; it is collapsed by region and year (see text for more details). In all regression, HIV and AIDS mortality are calculated at the country level.
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, UNAIDS.

Table 11: HIV, capital-labor ratio and the skill premium

	Levels	Growth
	(1)	(2)
HIV	0.061	-0.018
	(0.088)	(0.025)
HIV*2nd tercile school	0.020	0.032
	(0.041)	(0.018)
HIV*3rd tercile school	0.124**	0.094***
	(0.055)	(0.027)
Capital-labor ratio	0.043***	0.104
	(0.012)	(0.072)
Capital-labor ratio*	0.005	0.151
2nd tercile school	(0.013)	(0.147)
Capital-labor ratio*	0.010	0.122
3rd tercile school	(0.017)	(0.119)
Age	0.051***	
	(0.015)	
Age squared/100	-0.000**	
	(0.000)	
Tenure	0.004	
	(0.003)	
Female	-0.019	
	(0.038)	
Union membership	0.052	
	(0.054)	
2nd tercile school	0.298**	0.105*
	(0.128)	(0.056)
3rd tercile school	0.991***	0.329***
	(0.197)	(0.066)
Country fixed effects	X	
Industry fixed effects	X	
Occupation fixed effects	X	
Observations	6,874	47
R-squared	0.831	0.302

Standard errors clustered by region in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: In column 1, the data is from 2000s employee surveys. In column 2, the data is from both 2000s and 1990s employee surveys; it is collapsed by schooling tercile, region and year. In column 2, the "capital-labor ratio" variable refers to the growth of the capital-labor ratio in the firms of the region.
Source: World Bank Enterprise Surveys, World Bank Africa Regional Program on Enterprise Development, Demographic and Health Surveys.