HIV Status and Labor Market Participation in South Africa
DRAFT– DO NOT CITE *

James Levinsohn
Ford School of Public Policy
University of Michigan and NBER

Zoë McLaren
Ford School of Public Policy
University of Michigan

Olive Shisana
Human Sciences Research Council
Pretoria, South Africa

Khangelani Zuma
Human Sciences Research Council
Pretoria, South Africa

November 4, 2008

1 Introduction

This paper employs an especially rich South African data set to estimate the causal impact of HIV status on an individual's labor market participation. As such, the paper contributes to a better understanding of the economic impact of HIV/AIDS.

The existence of an impact of HIV on labor market participation is hotly debated in South Africa. Some argue that HIV is a severe constraint on current economic growth because those who are too ill to work represent lost economic output. Others argue that because unemployment is so high (in the 30 to 40 percent range according to the broad definition[1]), HIV has minimal economic impact today. Proponents of this view note that the sex-age cohorts with the highest unemployment are essentially the same as those with the highest HIV prevalence rates.[2] Our estimates speak directly to this issue. This paper provides

---

*Special thanks to John DiNardo, Taryn Dinkelman, Justin McCrary, Nicoli Natrass, Jeff Smith, and Duncan Thomas for especially helpful comments and suggestions. Levinsohn acknowledges support from the NICHD.

[1]The broad definition of unemployment includes individuals who desired employment but had no job search activity within the past month (i.e. discouraged workers).

[2]Another argument that HIV has a minimal impact on employment is that with anti-retroviral (ARV) therapy, HIV+ individuals are not likely to be too sick to work, and hence HIV has only a minimal impact on labor market participation. As discussed below, this argument, while germane to some countries, is probably not highly relevant to South Africa in 2005.

the first nationally representative estimates of the impact of HIV on labor market status, certainly for southern Africa but probably also more broadly for non-rich countries.

There are at least two reasons for the paucity or outright lack of evidence– stringent data requirements and econometric challenges.

Obtaining estimates of the impact of HIV on labor market participation clearly requires data on both labor market participation (which is generally available) and on an individual's HIV status (which is collected fairly rarely).[3] Our econometric methods also require extensive data on correlates of an individual's HIV status.

Even with our rich data set, econometric challenges arise due to econometric endogeneity. In our context, endogeneity arises because an individual's labor market status might affect their HIV status, for example, migrating for employment opportunities may put individuals at greater risk for HIV (reverse causality bias). Alternatively, there may be factors that affect both HIV status and labor market status, such as an individual's age or their willingness to invest in the future (selection bias, omitted variable bias).[4]

Two of the standard approaches to addressing these issues are practically infeasible in this instance. The traditional solution is to apply an instrumental variables approach. For the application at hand, this would require an instrument that is correlated with the individual-level HIV status and is orthogonal to shocks to labor market status. This is a tall order to fill. One class of correlates with HIV status – socioeconomic variables such as income, education, sex, age and even household size – are likely correlated with labor market status, while another class of correlates – health-related variables such as sexual practices and knowledge about HIV – are themselves likely to be correlated with the socioeconomic covariates and, by association, with labor market status.[5] (A closely related approach uses a control functions with an exclusion restriction, and the same concerns that preclude good instruments similarly preclude a convincing exclusion restriction. See Heckman and Navarro-Lozano.)

The other approach to addressing sample selection which has gained currency in development economics is to conduct interventions such that the bias vanishes by design.[6] In the context of a study of HIV status and labor market participation, this would involve

---

[3]There are other data sets available for which the techniques employed in this paper are potentially applicable. These include some of the Demographic Health Surveys (DHS) and the Botswana AIDS Impact Survey (see Levinsohn and McCrary)

[4]These challenges are discussed in detail in Strauss and Thomas.

[5]With country-level data, the fraction of the male population that has been circumcised has been used as an instrument (see Werker, Ahuja and Wendell). At the individual-level, this instrument is problematic. The protective effect may not transfer to women (Wawer, Kigozi, Serwadda, Makumbi, Nalugoda, Watya, Buwembo, Ssempijja, Moulton and Gray) and there is often virtually no variation within large categories of the male population like religion, race and ethnic/tribal affiliation (see McKelvey).

[6]For example, Thomas and et al. use experimental methods to obtain a causal estimate of the effect of improved health (in this case, from increased iron intake) on labor market outcomes.

2

randomly assigning HIV status to individuals and comparing labor market status across the two groups. If the instrumental variables or control function with an exclusion restriction approaches are practically infeasible, the intervention/experiment approach is ethically infeasible. This is one context for which the experimental approach is impossible.

We turn to methods based on the propensity score to address the issues of selection and reverse causality. We investigate the impact of HIV status on labor market participation using two methods– propensity score reweighting and control functions.[7] While both have been used to address the endogeneity issue in the literature, there have been few direct comparisons of multiple estimators for the same problem. We use both methods as a marginal methodological contribution and, more importantly, as a robustness check.

Propensity score reweighting and the control function approach both rely on the assumption that sample selection into being HIV+ is random, conditional on attributes that are observable to the econometrician (selection on observables). Given the richness of the SABSSM II data set, this assumption is weaker than it might be under other circumstances.[8] The data set contains detailed information on sexual practices and knowledge of HIV transmission in addition to household characteristics and other factors that might influence HIV status. We examine the validity of the assumption of selection on observables in section 5 by testing whether a parsimonious specification succeeds in balancing variables not used in the estimation (balancing tests) and experimenting with multiple specifications to confirm that the inclusion of additional controls does not change the results (robustness checks).

In the next section, we briefly discuss the related literature. Section 3 discusses the data that are employed. In section 4, the two approaches to estimation are presented. Section 5 presents results, while section 6 provides both caveats and conclusions.

2   Related Literature

While understanding the relationship between HIV status and labor market participation is important from multiple perspectives, there are few studies that present evidence on the impact of HIV and fewer still that address issues of endogeneity.

A number of studies develop macroeconomic simulation models to examine the effect of HIV/AIDS on economic growth, however these models may be extremely sensitive to

---

[7]These methods are drawn mostly from the program evaluation literature, but they have also been used in other fields of economics. For example, a similar control function approach is used in Olley and Pakes for solving the endogeneity problem in the context of production functions. Levinsohn and McCrary use the propensity score re-weighting technique to address a missing data problem in the context of HIV prevalence.

[8]The survey questionnaire included over 175 questions, many with multiple sub-sections, yielding over four hundred individual covariates.

assumptions about life-expectancy for HIV+ individuals (see Cuddington and Hancock, Kambou, Devarajan and Over, and Arndt and Lewis). Instead of assuming an HIV/AIDS mortality rate, Bloom and Mahal estimate the impact of AIDS on growth by exploiting cross-country variation in HIV prevalence rates. They find that HIV has an insignificant effect on per capita GDP for the sample of 51 countries.

The HIV pandemic not only affects current economic growth, but can have lasting effects on growth rates into the future. Kalemli-Ozcan and Bell, Devarajan and Gersbach model the impact of current HIV prevalence rates on future economic growth. Because HIV+ parents are likely to die before reaching old age, they may invest less in their children's human capital acquisition, lowering the stock of human capital and contributing to lower growth rates in the future. On the other hand, if the AIDS epidemic causes a reduction in fertility that dominates this human capital effect, higher future living standards may result (Young).

Most papers that examine the impact of HIV on economic growth look exclusively at the effect of AIDS mortality, overlooking the effects of the illness on employment and productivity. There are some exceptions. Murray, Sonnenberg, Nelson, Shearer, Bester, Begley and Glynn find that the rate of minor work-related injuries was 30 percent higher for HIV+ miners in South Africa than for HIV- miners. The effect was observable within one year of seroconversion which suggests that the psychological shock has an effect even before the infection affects overall health.[9]

Habyarimana, Mbakile and Pop-Eleches use an instrumental variables approach to examine the effect of health on productivity for mine workers in Botswana. They document an inverse-V-shaped pattern of absenteeism for HIV+ workers in the two years around the inception of anti-retroviral (ARV) therapy. Workers who subsequently enrolled in ARV therapy missed about five times as many days of work as non-enrolled workers in the year prior to ARV therapy inception, but absenteeism rates returned to pre-peak levels after a year of therapy. Non-enrolled workers appear to be a valid control group because the two groups had similar levels of absenteeism in the 1-5 years prior to therapy inception. These findings suggest that health can have large effects on employment outcomes, and that ARV therapy is effective in reducing the disparity in productivity between HIV+ and HIV-workers.[10] Fox, Rosen, MacLeod, Wasunna, Bii, Foglia and Simon examine differences in on-the-job productivity between workers who subsequently died of AIDS and other workers on tea plantations in Kenya. They find that in the last year before death, AIDS victims are less productive (in this case, measured by the quantity of tea leaves picked), are more likely to be reassigned to less strenuous but less lucrative tasks and are more often absent

---

[9]This is likely an underestimate of the impact of HIV because miners who are most affected are more likely to take on easier tasks at work or leave employment altogether.

[10]Thirumurthy, Zivin and Goldstein also found that in western Kenya, labor supply increased within six months of initiating ARV therapy.

from work.

One way to estimate the causal effect of HIV status on employment is to use a plausibly exogenous instrument for HIV status. Werker, Ahuja and Wendell and McKelvey use variation in circumcision to instrument for HIV status because circumcision has been found to be associated with a reduced risk of HIV both in regression analyses (Weiss, Quigley and Hayes) and in randomized controlled trials (Auvert, Taljaard, Lagarde, Sobngwi-Tambekou, Sitta and Puren). They find that HIV/AIDS did not have a measurable effect on economic growth, savings, or fertility behavior in African countries but that there was weak evidence that HIV/AIDS reduced youth literacy, and increased malnutrition. McKelvey finds that across nine African countries and Haiti, HIV+ individuals are significantly less likely to have been employed or to have earned enough to contribute more than half of household expenditures. However, specification checks using two populations who would not be expected to benefit from circumcision (men under 20 years old and those who have never had sex) suggest that unobservable differences that vary with circumcision may be driving the results.

Transitory economic shocks can have short-term effects on an individual's propensity to engage in risky behavior, but potentially long-term effects on health. This is one avenue through which employment outcomes may impact HIV status, and one reason why reverse causality is a salient issue. Women who are economically vulnerable may become active with multiple sexual partners (Dinkelman, Lam and Leibbrandt), or turn to transactional sex to smooth consumption (Dunkle, Jewkes, Brown, Gray, McIntyre and Harlow; LeClerc-Madlala). Migration is another potential response to transitory shocks that can put both men and women at greater risk for HIV (Zuma, Lurie, Williams, Mkaya-Mwamburi, Garnett and Sturm).

3   Data

Our data come from the nationally-representative South African National HIV prevalence, HIV Incidence, Behaviour and Communication Survey (2005) conducted by the HSRC, CADRE and MRC.[11] The survey asked adult respondents questions about demographics, knowledge of HIV, sexual history, knowledge of voluntary counseling and testing (VCT) services, health, mental health, and drug and alcohol use. It also included a household module that asked for basic demographic data for all household members in addition to questions about household infrastructure and participation in government programs. While there were no questions directly addressing income or expenditure, the individual survey

---

[11]A detailed description of the survey methodology is found in Shisana and et al., available online at http://www.hsrcpress.ac.za/product.php?productid=2009

Table 1: Sample sizes

Sample sizes	
Total sample	23275
Aged 15 and up	16398
Agreed to HIV test	12032
Tested HIV+	1351
Infected for more than 6 months	1172
Universe for viral load sample (HIV+ Africans)	1239
Viral load sample	662
Viral load ≥30,000 copies/mL	82

did query labor market participation.[12]

The sample consisted of 23,275 individuals in 10,584 households.[13] Not everyone in a household was sampled. Field workers randomly selected at most one person from each of three age groups (2-14 years, 15-24 years and 25 and up) to be interviewed. While this sampling strategy is an efficient way to obtain a measure of HIV prevalence, it is problematic for any analysis of household dynamics. For example, we are unable to directly measure the impact of having an HIV+ spouse on adult labor supply, since only one household member above the age of 25 was surveyed.

The survey also included an opt-out HIV test for respondents age 2 and older. Dried blood spot specimens were obtained by pricking the respondent's heel, finger or ear. The test results were not made available to survey respondents; instead individuals were referred to a local testing center if they wished to learn their HIV status.[14] The response rate for testing was 65.4 percent overall, and 73.3 percent in the adult sample (age 15 and older) used for analysis (see Table 1). HIV incidence data was collected from all HIV+

---

[12]Respondents were asked to classify their "present employment situation" as one of the following twelve categories: homemaker not looking for work, homemaker looking for work, unemployed not looking for work, unemployed looking for work, informal sector not looking for permanent work, old age pensioner, sick/disabled and unable to work, student/pupil/learner, self-employed full time (40 or more hours per week), self-employed part time (less than 40 hours per week), employed part time (if none of the above), employed full time.

[13]When matching individuals to households based on household identification numbers, approximately 14 percent of the individual observations (3413) could not be matched to household data. Household variables were imputed to zero for these unmatched observations. A specification check suggests that the data were consistent with a pattern of "missing at random"; a dummy variable for being unmatched was not significant in any specification. As with all variables that are imputed, an indicator variable was created that took a value of 1 if the value was imputed and 0 otherwise. These indicator variables are included in the empirical specifications. Individual survey records were matched to HIV incidence and viral load data using the barcode number that identified each specimen.

[14]HIV testing is widely available and free at government health facilities and other organizations in South Africa.

specimens using an enzyme immunoassay that measures the ratio of HIV antibodies to other antibodies to determine the elapsed time since HIV infection.[15] For a random 50 percent sub-sample of the African HIV+ specimens, viral load data (i.e. an estimate of the number of copies of the virus per mL of blood) were also collected.[16] Doctors use viral load as a diagnostic to track the progression of the disease and determine when to begin drug treatment for asymptomatic HIV+ individuals. Sample sizes are given in Table 1.

Table 2 and Figure 1 present pertinent descriptive statistics. Table 2 presents HIV prevalence rates by race. Table 2 indicates that while HIV impacts all races, Africans have, by far, the highest prevalence rates. In our sample, 17.34 percent of Africans are HIV+ while the population group with the next highest prevalence rate, Coloured, has a rate of 2.73 percent.[17] [18]

Figure 1 presents the age profile of HIV prevalence in South Africa– a profile broadly comparable to those of other African countries. HIV prevalence peaks between age 25-30 for women and between 30-40 for men. The age profile for women is about 5 percentage points lower than UNAIDS ante-natal clinic statistics, which may be due to differences in the sampling frame(UNAIDS). The age profile of HIV prevalence beyond approximately age 30 for women and 35 for men reflects the later age at which men tend to become infected and is hence mostly a cohort effect.

Table 3 reports employment status by HIV status separately for men and for women. For our analysis, individuals were assigned to a labor market status using the broad definition of unemployment, which includes discouraged workers. Respondents were classified as unemployed if they report that they are unemployed (either looking for work or not) or if they are a housewife or homemaker who is looking for work. They were classified as not economically active if they reported being a student, an old age pensioner, too sick or disabled to work, or a housewife or homemaker who is *not* looking for work. Otherwise, they were classified as employed.[19] Prevalence rates are higher for the unemployed relative to the employed, and this is especially pronounced for women. Almost 21 percent of

---

[15]Incidence specimen collection and analysis were performed in South Africa using the Calypte HIV-1 BED Incidence Enzyme Immunoassay produced by Calypte Biomedical Corporation (Maryland, USA). BED ratios below 0.80 indicate HIV infection within the previous six months.

[16]Few individuals in the sample were ill with AIDS (i.e. had viral loads above 30,000 copies/mL) because access to ARV therapy can reduce the viral load to below the cutoff and lack of access to ARV therapy hastens death.

[17]These rates apply to our sample which only includes individuals aged 15 and over. Hence they differ from the national prevalence rates reported in Shisana and et al..

[18]The prevalence rates in Table 2 do not correct for sample selection due to non-random opt-out of the HIV test. We have replicated our results on the impact of HIV on labor market participation with weights that do correct for non-random opt-out as computed by HSRC and results are virtually identical. This suggests that selection into testing is independent of the effect of HIV on employment.

[19]Those who responded being in the "other" category are grouped with the employed to generate conservative estimates.

unemployed women are HIV+ (and another 21 percent did not even take the test.) Students and other not economically active individuals (mostly old age pensioners) have lower than average prevalence rates, and these individuals are not included in our analysis. A similar proportion of each labor market group refused the HIV test.

Table 2: Naïve HIV prevalence rates by race

Race	HIV prevalence rate (%)	N
African	17.34	9,664
Coloured	2.73	3,013
Indian	1.27	1,772
White	0.53	1,913
Total	14.14	16,362

[1] Sample includes all individuals aged 15 and over.

[2] Rates are calculated using sample weights only.

Table 3: Employment status by HIV testing status (for men and women)

Employment status	Men					Women				
	HIV+	HIV-	Not tested	Total	N	HIV+	HIV-	Not tested	Total	N
Employed (and other)	8.5	64.1	27.4	100	2,815	10.9	61.5	27.6	100	2,693
Unemployed (broad definition)	12.6	63.0	24.4	100	1366	20.8	58.1	21.2	100	3,037
Student (not economically active)	1.9	68.3	29.8	100	1408	7.7	67.2	25.1	100	1,568
Other not economically active	4.3	72.2	23.5	100	621	6.8	72.1	21.1	100	2,636
Missing data	23.4	58.5	18.2	100	131	5.8	67.3	27.0	100	123
Total	8.1	65.4	26.6	100	6,341	12.8	63.8	23.4	100	10,057

[1] Sample includes all individuals aged 15 and over.

4 Methodologies

4.1 Overview

Generating a plausible counterfactual is the core challenge to identifying a causal effect of HIV. For the situation at hand, this approach entails creating a counterfactual in which one could compare individuals who were virtually identical except for their HIV status and then compare differences in labor market status. The following subsections present two methods, each based on the propensity score, to generate unbiased estimates of the effect of being HIV+ on labor market status. We begin with a brief introduction to our general approach and introduce notation before discussing each of the two approaches to estimation.

Our "treatment" (as the public health and program evaluation literatures refer to it) is HIV status, denoted by D, where $D = 1$ is HIV+ and $D = 0$ is HIV-.[20] If HIV status were independent of our outcome of interest, labor market status (denoted Y), then there is no sample selection problem and we can use a simple "naive" estimator, a simple difference in means, to estimate the treatment effect. However, HIV status is not randomly assigned and is in all likelihood related to our outcomes of interest. We make the conditional independence assumption (CIA) given by:

$$(Y_1, Y_0) \perp\!\!\!\perp D | X. \tag{1}$$

That is, once one conditions on individual characteristics in X, that D contains information only on a respondent's HIV status, not on the value of the potential treated and untreated outcomes (Y_1, Y_0). This assumption is valid insofar as the unobservables are unrelated to the probability of being HIV+ once one has conditioned on the relevant observable individual attributes. That is, selection on unobservables is ignorable.

4.2 The Propensity Score

In this context, the propensity score, denoted $P(HIV+)|X \equiv p(X)$, is the probability that an individual is HIV+ conditional on a vector of observable exogenous covariates (X). These covariates must be exogenous to HIV status in that they cannot be affected by HIV status (i.e. it would be inappropriate to use sexual behavior that may be changed if HIV status is known).[21] Rosenbaum and Rubin show that if the conditional independence

---

[20]We realize that labeling HIV+ as a "treatment" is problematic because of the potential for confusing this with being "treated" for HIV with anti-retroviral drugs. We adopt the standard language of "being treated" used in the program evaluation literature and will refer to drug treatment as anti-retroviral *therapy*.

[21]It is important to note that these are not the same Xs we would choose if we were running a standard instrumental variables (IV) regression; in that case we would want Xs correlated with HIV status but not

assumption is satisfied by conditioning on the vector X, it is satisfied by conditioning on the propensity score (i.e. the propensity score is a sufficient statistic for the X vector in the CIA). Hence, the propensity score makes the problem of finding a comparable control group tractable by reducing the dimensionality of the comparison, while still satisfying the CIA.

4.3 Propensity Score Reweighting

Propensity score reweighting uses the propensity score to create a counterfactual distribution of X in the HIV- (control) population so as to match the distribution of X in the HIV+ population. Essentially, HIV- observations with X characteristics that are most like HIV+ observations (i.e. that have a high $\hat{p}(X)$) receive the most weight, whereas HIV- observations that are very different from the HIV+ population receive less weight.

More formally, Dehejia and Wahba show that

$$\Delta^{ATET} = \frac{1}{N^T}\Sigma_{i=1}^{N^T+N^C}\left(D_iY_i - (1 - D_i)Y_i\frac{p(X_i)}{1 - p(X_i)}\right) \qquad (2)$$

is a consistent estimator for the average treatment effect on the treated (ATET), provided the CIA and the common support condition hold. In the calculation, HIV+ observations receive a weight of 1 (the first term in parentheses) and HIV- observations are weighted with $\omega_i = \frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}$ (the second term in parentheses). This estimator does not impose a functional form on the relationship between HIV status and the outcome of interest (Dehejia and Wahba).

This approach requires overlap in the support of the propensity scores for the HIV+ and HIV- groups (the common support condition). In the Results section, we present density plots to verify the validity of this assumption.[22]

4.4 Control Functions

The control function approach is an alternative econometric strategy for addressing selection bias. The reason one obtains biased estimates from a simple regression of labor market status on HIV status is that the disturbance term in such a regression is correlated with

---

with labor market status. Because actual HIV status is known by the econometrician and can therefore be controlled for in the regression, the covariates in X do not need to be good predictors of HIV status. In fact, selecting the attributes for X depending on predictive power can actually increase bias. See Heckman and Navarro-Lozano.

[22]Frölich demonstrates that an estimated propensity score close to 1 can cause problems. Our values of the estimated propensity scores are not large enough for this to be a concern.

HIV status (the independent variable). The essence of the control function approach is to proxy (or control for) the portion of the disturbance term that is correlated with HIV status. Once the portion of the disturbance term that is responsible for the correlation is expunged, the new error term is uncorrelated with HIV status, and the regression yields unbiased estimates of the impact of HIV on employment status.[23] As with propensity score reweighting, we maintain the assumption of selection on observables.[24]

The data generating process is given by:

$$Y = \beta_0 + \gamma D + f(X) + \epsilon \tag{3}$$

where $f(X)$ is a function of observables X.

Under the assumption of selection on observables, conditioning on $f(X)$ results in a disturbance term, ϵ, that is independent of D and hence, the estimate of the parameter of interest, γ is unbiased. In practice, a polynomial in the estimated propensity score as well as linear terms in X are used to flexibly model $f(X)$. We estimate:

$$Y = \beta_0 + \gamma D + X'\phi + \sum_{i=1}^{k} \beta_i \hat{p}(X)^i + \epsilon. \tag{4}$$

5   Results

Estimating the impact of HIV status on labor market participation is a three-step process. The first step is to estimate the propensity score. The second step is to empirically examine the validity of the CIA and the common support assumption to ensure that the appropriate observable variables are included in the propensity score regression. If the underlying assumptions hold, then one can proceed to the third step, estimating the impact of HIV on labor market participation using propensity score reweighting and a control function approach. This section presents the results of this procedure.

---

[23]The control function approach was developed in Heckman and Robb and has been used to estimate the impact of training on earnings (Heckman and Hotz), the returns to education (Card) and the capitalization of pollution into housing values (Chay and Greenstone).

[24]A similar control function method can be used if there is selection on *un*observables (see Heckman and Navarro-Lozano). However, allowing for selection on unobservables requires an exclusion restriction– a variable that is correlated with HIV status but uncorrelated with labor market status.

5.1 The Propensity Score

There are competing philosophies behind what constitutes a properly specified propensity score regression. One approach is to adopt a relatively parsimonious specification, albeit one still rich enough to satisfy the CIA, while another approach is to include most all plausible regressors. We adopt the former, but experiment with the latter in sensitivity analyses.

The plausibility of the selection on observables assumption clearly rests on having data that can, in our context, account for selection into HIV status. The SABSSM II data set has several *hundred* variables for most respondents. In addition to the usual demographic information, the survey also collected extensive information on sexual practices and knowledge about HIV transmission, which is exactly the type of information needed to account for selection into HIV status. Information on both sexual practices *and* knowledge about HIV transmission is important for predicting HIV status. A respondent who has multiple partners but is well informed on how HIV is transmitted and practices safe sex will have a low probability of being HIV+. Similarly, a respondent who knows very little about how HIV is transmitted but is abstinent will have a low likelihood of being HIV+. Conditional on comprehensive information about an individual's sexual practices and knowledge regarding HIV transmission, unobservables such as attitudes towards risk or moral beliefs may have little explanatory power. Hence, the selection on observables assumption seems especially appropriate given the specifics of our data. The selection on observables assumption is testable, and we investigate the reasonableness (or not) of this assumption by conducting balancing tests on covariates in the propensity score specification as well as plausible covariates that have been excluded in section 5.4.

We estimate the propensity score as the predicted value of a logit regression. Table 4 reports the estimated coefficients and standard errors from our baseline regression.[25][26] We do not discuss the estimated coefficients because the main focus of this paper is on using the propensity score to correct for selection bias rather than on the correlates of seropositivity. Furthermore, it is quite difficult to have much intuition about the marginal impact of a single regressor conditional on the other 44.

---

[25]Elements of the X vector of covariates were imputed to zero for item non-response, and a dummy variable for imputation was included in the specifications. In each specification, we include indicator variables for each province in South Africa as well as indicator variables set to one if the value of a variable was imputed due to missing data. These coefficients are not reported in the table.

[26]We also examine particular cuts of the data (e.g. by sex, by education, by area of residence). For each of these sub-samples, we re-estimate the propensity score. These results are available on request.

Table 4: Propensity Score Logit

COEFFICIENT	
Age	0.145***
	(0.036)
Age-squared	-0.002***
	(0.000)
Female	0.480***
	(0.101)
African	3.185***
	(0.380)
Coloured	1.743***
	(0.413)
Indian	0.315
	(0.546)
Urban resident	0.246**
	(0.108)
Never married	0.409***
	(0.114)
Has had sex	1.960***
	(0.743)
Age at first sex	-0.079
	(0.066)
Age at first sex -squared	0.002
	(0.001)
Completed primary education	0.245
	(0.171)
Completed secondary education	0.266
	(0.165)
Holds a matric qualification	0.190
	(0.189)
Has some post-matric education	-0.579**
	(0.242)
Used a condom at first sex	-0.391***
	(0.146)
Knows HIV transmitted through vaginal sex	-0.422**
	(0.173)
Believes HIV not transmitted through witchcraft	-0.127
	(0.142)
Knows condoms prevent HIV transmission	0.122
	(0.147)
Knows reducing number of partners reduces risk	0.176*
	(0.101)
Male of pension age in household	-0.235
	(0.339)
Female of pension age in household	-0.189
	(0.155)
Constant	-9.522***
	(0.743)
Pseudo-R^2	0.1467
Observations	12032

5.2 Examining the Validity of the Propensity Score Method

We next examine the validity of the CIA and common support assumption under our preferred (base case) propensity score specification. The former is done with balancing tests and the latter by examining empirical distributions.

The intuition behind balancing tests is appealingly clear. The idea of propensity score reweighting is to reweight the distribution of the observables (X's) of the HIV- population so as to match the distribution of the X's for the HIV+ population (see Equation 2). Balancing tests simply examine whether the difference in means of X's between HIV+ and HIV- populations is reduced when the observations are reweighted. If the reweighted means are similar (i.e. not significantly different) for HIV+ and HIV- populations, then the reweighting has achieved its goal, the data are balanced, and the CIA is appropriate.

Table 5 reports the results of balancing for those variables that enter the propensity score regression (i.e. internal balancing).[27] The results in Table 5 apply to the entire sample. Balancing tests were also conducted for each of the four subsamples with comparable results. Row 1 of Table 5 presents the results for age; HIV+ individuals were on average 3.54 years younger than the HIV- population. This difference had a t-statistic of 6.737 so the difference was highly significant, which is denoted by the asterisks in the "before" column. After reweighting, the difference falls to 0.337 years, which is not significantly different from zero. None of the variables in the propensity score regression have a statistically significant difference between the HIV+ and HIV- populations in the reweighted data.

The propensity score weight by design attempts to minimize differences between HIV+ and HIV- groups for the variables included in the propensity score estimation (the X vector). A more stringent balancing test criterion is whether the propensity score weight also succeeds in balancing covariates that were not used in the estimation. Results of external balancing tests for 56 additional variables are given in Table 6. In the unweighted data, 25 of these variables had means that were significantly different between the HIV+ and HIV- populations. After reweighting, only six of those differences were still significant, and in five of the six cases, the reweighting decreased the difference in means.

Based on Tables 5 and 6, we conclude that the propensity score specification is adequately reweighting the data to justify the CIA. While selection on observables is a strong assumption, we find that the richness of our data provides support for the assumption.

---

[27]We also conducted balancing tests for each of the indicator variables for missing data, but these are not reported in the table. All were balanced.

Table 5: Balancing tests for variables used in estimation of propensity score.

Variable	Unweighted				Reweighted		
	Coeff.	Std. E	t-Stat		Coeff.	Std. E	t-Stat
Age	-3.544	0.526	-6.737	***	-0.337	0.488	-0.690
Age-squared	-413.386	45.430	-9.099	***	-28.610	40.733	-0.702
Female	0.117	0.021	5.458	***	-0.002	0.023	-0.066
African	0.211	0.008	27.970	***	0.000	0.003	0.061
Coloured	-0.090	0.005	-19.186	***	0.000	0.003	-0.035
Indian	-0.022	0.002	-11.287	***	0.000	0.001	-0.051
Eastern Cape	-0.011	0.014	-0.788		-0.003	0.015	-0.218
Northern Cape	-0.012	0.002	-4.983	***	0.000	0.002	-0.200
Free State	0.022	0.012	1.799		-0.001	0.014	-0.049
Kwa-Zulu Natal	0.088	0.019	4.671	***	0.002	0.021	0.101
Northwest Province	0.008	0.014	0.555		0.000	0.014	0.000
Gauteng	0.006	0.018	0.344		0.005	0.019	0.270
Mpumalanga	0.038	0.011	3.345	***	-0.001	0.013	-0.083
Limpopo	-0.038	0.011	-3.286	***	-0.002	0.011	-0.175
Urban dweller	-0.011	0.022	-0.501		0.005	0.023	0.205
Never married	0.079	0.021	3.699	***	0.011	0.023	0.469
Has had sex	0.093	0.013	6.978	***	0.001	0.013	0.096
Age at first sex	1.050	0.371	2.834	***	0.047	0.385	0.123
Age at first sex-squared	11.741	8.113	1.447		0.497	8.438	0.059
Completed primary education	0.029	0.019	1.540		-0.005	0.021	-0.251
Completed secondary education	0.034	0.021	1.611		0.007	0.023	0.327
Holds a matric qualification	0.014	0.018	0.781		0.002	0.019	0.108
Completed some post-matric educ	-0.059	0.009	-6.743	***	0.001	0.008	0.136
Used a condom at first sex	-0.028	0.015	-1.883		0.002	0.015	0.158
Knows HIV transmitted through vaginal sex	-0.011	0.011	-0.980		0.000	0.012	0.016
Believes HIV not transmitted through witchcraft	-0.009	0.016	-0.539		0.004	0.017	0.248
Knows condoms prevent HIV transmission	0.042	0.013	3.241	***	0.000	0.014	0.012
Knows reducing number of partners reduces risk	0.015	0.020	0.744		0.000	0.021	0.005
Male of pension age in household	-0.046	0.012	-3.772	***	-0.002	0.012	-0.127
Female of pension age in household	-0.069	0.012	-5.678	***	-0.002	0.013	-0.164
Race missing	-0.001	0.001	-2.260	**	0.000	0.000	-0.057
Has had sex missing	0.011	0.007	1.471		-0.001	0.008	-0.063
Age at first sex missing	0.027	0.017	1.572		-0.003	0.018	-0.146
Used a condom at first sex missing	0.016	0.009	1.822		-0.001	0.010	-0.133

[1] Sample includes all individuals aged 15 and older.
[2] *** significant at 99 percent level, ** significant at 95 percent level.

Table 6: Balancing tests for additional variables *not* used in estimation of propensity score.

Variable	Unweighted				Reweighted			
	Coeff.	Std. E	t-Stat		Coeff.	Std. E	t-Stat	
Used a condom at last sex 1	0.113	0.020	5.664	***	0.037	0.022	1.726	
Used a condom at last sex 2	0.026	0.018	1.424		0.057	0.019	2.987	***
Number of current partners	0.107	0.039	2.770	***	0.032	0.029	1.084	
Number of partners in last year	0.174	0.033	5.217	***	0.066	0.036	1.847	
Had HIV test before	0.070	0.021	3.409	***	0.049	0.022	2.231	**
Got HIV test result	0.058	0.019	3.033	***	0.038	0.021	1.812	
Heard of ARVs	0.012	0.021	0.559		-0.004	0.023	-0.164	
Receives a government pension	0.021	0.018	1.145		0.013	0.020	0.667	
Gone without shelter 1	0.007	0.004	1.688		0.006	0.004	1.449	
Gone without shelter 2	0.002	0.005	0.302		-0.002	0.006	-0.393	
Gone without shelter 3	-0.022	0.010	-2.199	**	-0.006	0.011	-0.561	
Gone without fuel 1	0.029	0.015	2.000	**	0.004	0.016	0.262	
Gone without fuel 2	0.005	0.011	0.408		0.002	0.012	0.152	
Gone without fuel 3	-0.074	0.020	-3.741	***	-0.030	0.021	-1.414	
Gone without water 1	0.009	0.013	0.666		0.000	0.014	0.033	
Gone without water 2	0.015	0.012	1.282		0.008	0.013	0.577	
Gone without water 3	-0.034	0.019	-1.806		-0.009	0.020	-0.465	
Gone without medical treatment 1	0.046	0.018	2.545	**	0.025	0.019	1.318	
Gone without medical treatment 2	0.020	0.012	1.680		0.000	0.013	-0.031	
Gone without medical treatment 3	-0.091	0.021	-4.330	***	-0.027	0.022	-1.187	
Gone without food 1	0.034	0.017	1.940		0.003	0.018	0.184	
Gone without food 2	0.027	0.014	1.921		0.016	0.015	1.052	
Gone without food 3	-0.108	0.022	-4.979	***	-0.052	0.023	-2.246	**
Gone without cash income 1	0.059	0.021	2.809	***	0.017	0.022	0.772	
Gone without cash income 2	0.006	0.010	0.574		-0.011	0.011	-0.977	
Gone without cash income 3	-0.139	0.020	-6.812	***	-0.054	0.022	-2.481	**
Source of water 1	-0.076	0.021	-3.670	***	-0.002	0.022	-0.070	
Source of water 2	0.017	0.016	1.099		-0.006	0.017	-0.382	
Source of water 3	0.052	0.018	2.877	***	0.013	0.019	0.684	
Source of water 4	0.004	0.004	1.083		0.003	0.005	0.603	
Source of water 5	0.005	0.006	0.779		0.003	0.007	0.377	
Source of water 6	-0.003	0.009	-0.300		-0.006	0.010	-0.563	
Source of water 7	0.005	0.003	1.382		0.005	0.004	1.447	
Source of water 8	0.001	0.002	0.362		0.001	0.002	0.420	
Type of toilet 1	-0.095	0.021	-4.518	***	-0.012	0.022	-0.536	
Type of toilet 2	0.075	0.020	3.716	***	0.018	0.022	0.842	
Type of toilet 3	0.016	0.006	2.730	***	0.013	0.006	2.091	**
Type of toilet 4	-0.003	0.002	-1.724		-0.002	0.002	-0.942	
Type of toilet 5	0.012	0.015	0.817		-0.005	0.016	-0.295	
Type of cooking fuel 1	-0.079	0.022	-3.646	***	-0.032	0.023	-1.403	
Type of cooking fuel 2	0.009	0.006	1.354		0.003	0.007	0.483	
Type of cooking fuel 3	-0.005	0.017	-0.324		-0.011	0.018	-0.599	
Type of cooking fuel 4	0.079	0.016	4.991	***	0.048	0.017	2.819	***
Has electricity	-0.048	0.021	-2.298	**	0.009	0.023	0.408	
Owns a refrigerator	-0.116	0.021	-5.398	***	-0.041	0.023	-1.780	
Owns a radio	-0.051	0.021	-2.441	**	-0.016	0.022	-0.712	
Owns a tv	-0.073	0.022	-3.380	***	-0.010	0.023	-0.452	
Has a landline phone	-0.125	0.014	-8.753	***	-0.029	0.015	-1.945	
Migrant household members	0.018	0.014	1.308		0.023	0.014	1.631	
Not matched to household data	-0.001	0.015	-0.053		-0.007	0.016	-0.415	

[1] Full sample includes individuals aged 15 and older.
[2] *** significant at 99 percent level, ** significant at 95 percent level.

17

The balancing tests provide support for the CIA. The other assumption underlying the propensity score reweighting approach is the common support assumption.[28] We examine the appropriateness of the support assumption by comparing the empirical distributions of the propensity scores for the HIV+ and HIV- populations. These distributions are shown in figure 2. In that figure, it is clear that the densities have a common support.

5.3 The Impact of HIV

5.3.1 Estimates

Table 7 presents estimates of the causal impact of HIV on labor market participation. The table is organized such that each column presents estimates resulting from a different estimator and each row presents estimates using different samples of the data.

The first column presents naïve estimates of the impact of HIV status on labor market participation. Estimates in this column are simply the (unconditional) difference in the mean employment status for the HIV+ and HIV- populations. Variables are defined such that the 0.157 figure in the first cell implies that, on average, HIV+ individuals are 15.7 percentage points more likely to be unemployed.

The second column presents the coefficient on HIV status when the variables in the propensity score are included as controls in a simple logit regression of employment status on HIV status. We present these estimates for purposes of comparison. We stress that the estimates in this column should not be interpreted as causal estimates. In this case, the estimates are of similar magnitude to our causal estimates, but that is not necessarily always the case. The logit result implies that for the full sample being HIV+ lowers the probability of employment by 6.8 percentage points.

The third and fourth columns present estimates of the causal impact of HIV on employment using propensity score reweighting and the control function approaches, respectively. Each parameter estimate in columns 3 and 4 is the marginal effect of HIV status from a logit regression. For column 3, it is a simple logit regression of labor market status on HIV status using propensity score reweighting while in column 4, the reported estimate is the coefficient on HIV status (D) in the regression given by (4). Using the entire sample, propensity score reweighting indicates that being HIV+ raises the probability of unemployment by 6.8 percentage points while the control function approach indicates an increase of 7.4 percentage points. Each of these impacts is precisely estimated.[29] Our estimates of the

---

[28]The support assumption is not required for the control function approach.

[29]Both the reweighting and the control function approach use the *estimated* propensity score. Hirano, Imbens and Ridder show that the unadjusted standard errors are (somewhat counter-intuitively) larger than the true standard errors. Hence, inference using unadjusted standard errors is conservative.

causal impact of HIV on employment imply that, all else equal, being HIV+ raises the probability of unemployment by about seven percentage points.

The estimates in the first row of Table 7 apply to the entire sample and, as such, hide substantial underlying heterogeneity in the impact of HIV on labor market status. We investigate this heterogeneity in the remainder of Table 7 by restricting the sample to particular sub-populations.

The second through seventh rows in Table 7 present results obtained from an analysis of sub-samples of the data, using propensity score estimates calculated within the sub-sample alone.

For men, propensity score reweighting and the control function approach give estimates of 0.056 and 0.071, respectively. These are slightly smaller than the point estimates with the full sample, and they are not as precisely estimated. We can no longer reject the null hypothesis that the estimated effect is zero. Compared to the results using the entire sample, the impact of HIV on labor market participation is simply more diffuse among males. For females, the point estimates from propensity score reweighting and the control function approach are 0.072 and 0.067. These are about the same as the estimates obtained with the entire sample, and they remain precisely estimated at conventional levels of statistical significance.

Table 7: Marginal effect of being HIV+ from naïve (unconditional) difference in means, logit with controls, propensity score reweighting (RW) and control function (CF).

Sample	Naïve		Logit		RW		CF		N
Full sample	0.157	***	0.068	**	0.068	***	0.074	**	7467
	(0.024)		(0.029)		(0.026)		(0.029)		
Male	0.074		0.059		0.056		0.071		3087
	(0.040)		(0.045)		(0.045)		(0.047)		
Female	0.166	***	0.065		0.072	**	0.067	**	4380
	(0.026)		(0.033)		(0.029)		(0.034)		
Matric and up	0.183	***	-0.0157		-0.006		-0.008		1798
	(0.054)		(0 .047)		(0.063)		(0.048)		
Less than Matric	0.144	***	0.110	***	0.098	***	0.111	***	2270
	(0.032)		(0.035)		(0.034)		(0.035)		
Rural	0.121	***	0.056		0.048		0.065		3697
	(0.039)		(0.043)		(0.039)		(0.043)		
Urban	0.179	***	0.075	**	0.086	***	0.078	**	5197
	(0.030)		(0.035)		(0.033)		(0.036)		

[1] Full sample and results by sex include individuals aged 15 and older. Results by education level are restricted to individuals aged 25 and older.

[2] Naïve is OLS without controls. Controls used in logit are same as those used in propensity score estimation.

[3] *** significant at 99 percent level, ** significant at 95 percent level.

The next two rows in Table 7 restrict the analysis to respondents aged 25 and older, and divide the sample between individuals whose education level is a Matric or higher and those whose education level is less than a Matric.[30] The message here is clear: the impact of HIV on labor market status is severe for those with lower levels of education and is negligible for those with higher levels. The causal impact of HIV status on labor market status for those with less than a Matric is an increase in the likelihood of unemployment of 9.8 percentage points (with propensity score reweighting) and 11.1 percentage points (with the control function approach.) These are large and precisely estimated impacts.[31]

Tha last two rows highlight urban/rural differences in the impact of HIV on unemployment. Although both HIV and unemployment are more prevalent in rural areas, the causal impact of HIV on unemployment is larger in urban areas where being HIV+ lowers the probability

---

[30]Respondents under 25 may not have completed their schooling. A Matric is about equivalent to a high school education.

[31]We repeated the analysis for subgroups defined by race as well as by groups defined by education and sex, age, and education and age. These results are not reported here, but will be available at Levinsohn's web site, www.umich.edu/ ˜ jamesL. The results in Table 7 capture the gist of these divisions. Women, and especially women with lower levels of education experience the larger causal impacts of HIV on labor market status.

of employment by 8 to 9 percentage points.

5.3.2 Discussion

A concern about the role of selection bias motivated our choice of methodologies. Comparing the naïve estimates in column 1 of Table 7 to the causal estimates in columns 3 and 4 speaks to this issue. The naïve estimates are about twice as large as the causal estimates, and this highlights the importance of selection. Our results are consistent with the hypothesis that HIV+ individuals are more likely to be unemployed irrespective of their HIV status than is the average South African.

Because this study is probably the first to examine the causal impact of HIV on employment outcomes, it is difficult to place the magnitude of the estimated impact in context. There are no other estimates available for comparison. There are at least two economic arguments for why one might have expected no causal impact. First, if unemployment were so pervasive that HIV+ individuals would be unemployed even in the absence of HIV, one would expect no impact. Second, if ARVs were sufficiently widely used, one might expect either no impact or a very tiny impact.[32] Our estimates indicate that these arguments, while perhaps *ex ante* plausible, are on average simply incorrect. Being HIV+ lowers the probability of employment.

A counter-argument to the notion that HIV confers a negligible penalty in the face of extremely high unemployment is that in the presence of high unemployment, even a small disadvantage (e.g. the stigma sometimes associated with HIV) much less a large disadvantage (adverse physical effects of HIV) mean the difference between keeping a job and losing it. The results in Table 7 are consistent with this counter-argument.

One way to gauge the magnitude of our estimates (beyond noting that they are not zero) is to compare the estimated marginal effect of HIV status (D) with the estimated marginal effect of other respondent characteristics (X) in 4. For men, the magnitude of the labor market advantage of being HIV- is approximately equal to the impact of 3 years of age, a Matric qualification (compared to no education), or the absence of a female pensioner in the household. For women it is equal to the impact of 1.5 years of age, some secondary education (compared to no education), or the absence of a male pensioner in the household.

We have estimated the causal impact of HIV status on the probability of unemployment, and in this sense HIV contributes to unemployment in South Africa. It would be wrong, though, to think that HIV was substantially *responsible* for South Africa's dire unemploy-

---

[32]One might expect a tiny impact rather than no impact because even with ARVs, most HIV+ people eventually do become too ill to work.

ment. A simple back-of-the-envelope calculation illustrates why. The overall unemployment rate, UE, can be decomposed as:

$$UE = (\%HIV+)(UE \ if \ HIV+) + (\%HIV-)(UE \ if \ HIV-)$$
$$0.478 = (0.180)(0.6068) + (0.820)(.4502)$$

This equation makes clear that if the unemployment rate of those who are HIV+ were to fall by about 8 percentage points, the impact on the overall unemployment rate would be small.

Our results are conditional on the availability of ARVs as of 2005. We do not have data on which HIV+ respondents were on ARV therapy. Access to ARVs in South Africa is far from universal. One estimate is that in 2005, the year of our sample, only about 18 percent of those who needed ARVs were actually using them. (See Dorrington, Bradshaw, Johnson and Daniel.) As the availability of ARVs changes, the impact of HIV on labor market status, as we estimate it, will change. Because it is likely that ARVs are more widely available today than they were in 2005, it is likely that, *ceteris paribus*, the impact of HIV on unemployment is lower today than it was in 2005. In addition, ARV usage is non-random, and this may in part contribute to the pattern of results in Table 7. For example ARVs are much more likely to be employer-provided in the formal sector than they are in the informal sector. Our findings are consistent with the fact that women and the less-educated tend to be more heavily represented in the informal sector and in domestic (housekeeper) work and hence less likely to receive employer-provided ARVs.

We also found that being HIV+ had virtually no employment impact for better educated workers (and recall this result already accounts for the fact that highly educated individuals are less likely to be HIV+). ARVs may also be contributing to this finding. Employers have greater incentive to invest in ARVs for workers who are more difficult to replace such as highly educated workers.

In sum, HIV appears to reinforce the already existing inequalities in South Africa.

5.4 Sensitivity Analyses

We investigate the sensitivity of our results both to alternative definitions of unemployment and to alternative specifications of the propensity score regression.

As described in section 3, individuals were divided into three groups in our base case specification – unemployed, employed, and not economically active (NEA). The first two groups are included in our analysis while the third is not. In the base case results using the broad definition of unemployment, so-called discouraged workers – workers who were not

actively seeking a job but desired employment – were classified as unemployed. The narrow definition of unemployment classifies discouraged workers as NEA. We repeat the analysis using this narrow definition of unemployment. Results are reported in Appendix Table A1. The results are very similar to those in Table 7. The estimated coefficients are slightly smaller than those reported in Table 7, with only 2 of the 15 coefficients (in the latter 3 columns) differing from their counterpart in Table 7 by more than two percentage points. We also repeated the analysis by combining the NEA with the employed to create a group that might (awkwardly) be called not-unemployed. Again results barely change. Results are reported in Appendix Table A2. Compared to Table 7, coefficients tend to increase, though only 5 of the 15 change by more than two percentage points. Results tend to be more precisely estimated, and this is driven in part by the increase in the sample size when NEA individuals are included in the analysis. From Appendix Tables 1 and 2, we conclude that our findings are robust to alternative definitions of labor market status.

We also experimented with alternative specifications of the propensity score regression. The results obtained from regressions using additional variables in the calculation of the propensity score are virtually identical to results obtained from the preferred specification. These results are not reported here. When including ten variables with information on the household, no point estimate changed by more than two percentage points, and only three of the 15 point estimates changed by more than one percentage point.[33] Adding additional behavioral variables did not change any of the point estimates by as much as one percentage point.[34]

6   Conclusions and Caveats

6.1   Conclusions

Identifying the causal impact of HIV on labor market status requires addressing the issue of selection into being HIV+. In the absence of plausible instruments, we exploit the richness of the data and assume that selection is on observables. Internal and external balancing tests support the validity of this assumption. Employing two estimation strategies, we find that being HIV+ causes, on average, an increase in the likelihood of unemployment of about seven percentage points. It is important to note that this penalty exists despite very high unemployment rates. This average impact hides important heterogeneity. HIV's causal

---

[33]The additional household variables were: type of toilet facility, source of energy for cooking, access to electricity, presence of a land line, and a dummy for whether the household information was missing for the observation.

[34]The additional behavioral variables were condom use at last sex, number of current sexual partners, number of partners in the last year, whether the respondent had been tested for HIV before, whether they had received their test result, and whether they had heard of ARVs.

impact on unemployment is larger (10 to 11 percentage points) for less educated South Africans. The results are robust to multiple alternative econometric specifications.

6.2 Caveats

These results are the first nationally representative estimates of the causal impact of HIV on employment in South Africa. While informative, they are not dispositive. Rather, the results should be interpreted with caution for at least five reasons.

First, our analysis does not account for any general equilibrium effects. In particular, it would be misleading to think that if ARVs were made universally available or a cure for HIV/AIDS were found, that HIV+ individuals would see their likelihood of employment rise on average by about seven percentage points. Rather, the labor market would adjust and these adjustments would depend on supply and demand elasticities. Complicating this analysis, the ability of the labor market to immediately absorb additional healthy workers is questionable.

Second, data limitations preclude an analysis of the indirect labor market impact of having multiple HIV+ adult household members. Recall that the structure of the SABSSM II survey is such that only one adult age 25 or older is sampled from each household. It is unclear in which direction our results may be biased. An HIV- worker could be unemployed because they are caring for an HIV+ spouse, resulting in downward bias (i.e. the true impact is larger than our estimates suggest), or an HIV+ worker might be more motivated to obtain employment to financially support another HIV+ household member, resulting in upward bias.

Third, our results are conditional on the time profile of HIV and prevalence rates as of 2005. Given an approximately nine-year period (on average) of latent HIV infection before AIDS conversion, and HIV prevalence rates that increased from 5 percent in 1996 to 12 percent in 2001,[35] we would expect the "stock" of individuals with AIDS to increase quite sharply between 2003 and 2008. This implies that, *ceteris paribus*, the impact of HIV on unemployment would rise in coming years as the number of HIV+ individuals who are too ill to work increases. However, it is unclear how this effect would interact with any increase in the availability of ARVs.

Fourth, our results are not structural. As such, we are unable to convincingly address the particular avenues through which HIV impacts labor market status. Relatedly, we are unable to conduct detailed policy analysis. For example, increased access to ARVs and successful programs to de-stigmatize HIV might each increase the likelihood of employment,

---

[35]ASSA demographic model predictions cited in Natrass (p. 42). The model on which these estimates are based predicted a 14 percent prevalence rate for 2004 – about the same as that of our 2005 data.

but our approach cannot conduct the counterfactual experiments to estimate the likely impacts of these potential policies.

Lastly, our results are generally not applicable to other countries. South Africa has a stunningly high rate of unemployment, high HIV prevalence rates and a troubled history with the distribution of ARVs – three factors that suggest that it would be misguided to generalize the results of this study to other countries.

References

Arndt, Channing and Jeffrey D. Lewis, "The Macro Implications of HIV/AIDS in South Africa: A Preliminary Assessment," *South African Journal of Economics*, December 2000, *68* (5), 856–87.

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

Bell, Clive, Shantayanan Devarajan, and Hans Gersbach, "The Long-Run Economic Costs of AIDS: Theory and an Application to South Africa," *World Bank Policy Research Working Paper*, October 2003, (3152).

Bloom, David E. and Ajay S. Mahal, "Does the AIDS epidemic threaten economic growth?," *Journal of Econometrics*, 1997, *77* (1), 105–124.

Card, David, "The Causal Effect of Education on Earnings," *Handbook of Labor Economics*, 1999, *3*, 1801–1863. Editor: O. Ashenfelter and D. Card; Publisher:Elsevier Science B.V.

Chay, Kenneth Y. and Michael Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, April 2005, *113* (2), 376–424.

Cuddington, J. and J. Hancock, "Assessing the impact of AIDS on the growth path of the Malawian economy," *Journal of Development Economics*, 1994, *4*, 363–8.

Dehejia, Rajeev H. and Sadek Wahba, "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," in Rajeev Dehejia, ed., *Econometric Methods for Program Evaluation*, Ph.D. Dissertation, Harvard University, 1997, chapter 1.

___ **and** ___ , "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 1999, *94* (448), 1053–1062.

Dinkelman, T., D. Lam, and M. Leibbrandt, "Household and community income, economic shocks and risky sexual behavior of young adults: evidence from the Cape Area Panel Study 2002 and 2005," *AIDS*, 2007, *21* (Suppl 7), S49–S56.

Dorrington, R., D. Bradshaw, L. Johnson, and T. Daniel, "The demographic impact of HIV/AIDS in South Africa: National and Provincial Indicators for 2006," Technical Report, Cape Town 2006. Joint publication of the Centre for Actuarial

Research, the Burden of Disease Research Unit, and the Actuarial Society of South Africa (ASSA).

Dunkle, K., Rachel K. Jewkes, Heather C. Brown, Glenda E. Gray, James A. McIntyre, and Sioban Harlow, "Transactional sex among women in Soweto, South Africa: Prevalence, risk factors and association with HIV infection.," *Social Science and Medicine*, 2004, *59*, 1581–92.

Fox, Matthew P., Sydney Rosen, William B. MacLeod, Monique Wasunna, Margaret Bii, Ginamarie Foglia, and Jonathon L. Simon, "The impact of HIV/AIDS on labour productivity in Kenya," *Tropical Medicine and International Health*, March 2004, *9* (3), 318–24.

Frölich, Markus, "Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators," *Review of Economics and Statistics*, 2004, *86* (1), 77–90.

Habyarimana, James, Bekezela Mbakile, and Cristian Pop-Eleches, "HIV/AIDS, ARV Treatment and Worker Absenteeism: Evidence from a Large African Firm," April 2007.

Heckman, James and Salvador Navarro-Lozano, "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models," *The Review of Economics and Statistics*, February 2004, *86* (1), 30–57.

Heckman, James J. and Richard Robb, "Alternative Methods for Estimating the Impact of Interventions," *Longitudinal Analysis of Labor Market Data*, 1985. Ed: James J. Heckman and Burton Singer; Pub: Cambridge University Press.

___ **and V. Joseph Hotz**, "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 1989, *84* (408), 862–874.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder, "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, July 2003, *71* (4), 1161–1189.

Kalemli-Ozcan, Sebnem, "AIDS, Reversal of the Demographic Transition and Economic Development: Evidence from Africa," *NBER Working Paper*, April 2006, (12181).

Kambou, G., S. Devarajan, and M. Over, "The economic impact of AIDS in an African country: simulations with a computable general equilibrium model of Cameroon," *Journal of African Economies*, 1992, *1*, 109–130.

LeClerc-Madlala, S., "Youth, HIV/AIDS and the importance of sexual culture and context," *Social Dynamics - A Journal of the Centre for African Studies, University of Capetown*, 2002, *28*, 20–41.

Levinsohn, James and Justin McCrary, "Estimating Correct HIV Prevalence Rates: The Case of Botswana," 2008. Working Paper.

McKelvey, Christopher, "Circumcision and the Labor Market Consequences of HIV in Developing Countries," October 2007.

Murray, Jill, Pam Sonnenberg, Gill Nelson, Stuart Shearer, Andre Bester, Arthur Begley, and Judith R. Glynn, "Effect of HIV on work-related injury rates in South African gold miners," *AIDS*, 2005, *19*, 2019–2024.

Natrass, Nicoli, *The Moral Economy of AIDS in South Africa*, Cambridge University Press, 2004.

Olley, S. and A. Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1298.

Rosenbaum, Paul R. and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 1983, *70* (1), 41–55.

Shisana, Olive and et al., *Nelson Mandela/HSRC Study of HIV/AIDS: South African National HIV Prevalence, Behavioural Risks and Mass Media*, Human Sciences Research Council, 2002. Published in Cape Town.

Strauss, John and Duncan Thomas, "Health, Nutrition, and Economic Development," *Journal of Economic Literature*, 1998, *36* (2), 766–817.

Thirumurthy, Harsha, Joshua Graff Zivin, and Markus Goldstein, "The Economic Impact of AIDS Treatment: Labor Supply in Western Kenya," *NBER Working Paper*, 2005, (11871). Forthcoming in JHR (XXX).

Thomas, Duncan and et al., "Causal Effect of Health on Labor Market Outcomes: Experimental Evidence," *California Center for Population Research On-Line Working Paper Series*, September 2006, (070).

UNAIDS, "UNAIDS/WHO AIDS Epidemic Update," Technical Report, UNAIDS/WHO December 2005.

Wawer, Maria, G Kigozi, D Serwadda, F Makumbi, Nalugoda, S Watya, D Buwembo, V Ssempijja, L Moulton, and R Gray, "Trial of Male Circumcision in HIV+ Men, Rakai, Uganda: Effects in HIV+ Men and in Women Partners," February 2008. Paper presented at Conference on Retroviruses and Opportunistic Infections.

Weiss, Helen A., Maria A. Quigley, and Richard J. Hayes, "Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis," *AIDS*, 2000, *14*, 2361–70.

Werker, Eric, Amrita Ahuja, and Brian Wendell, "Male Circumcision and AIDS: The Macroeconomic Impact of a Health Crisis," *Harvard Business School Working Paper*, October 2006, (07-025).

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

Zuma, K., M.N. Lurie, B.G. Williams, D. Mkaya-Mwamburi, G.P. Garnett, and A. W. Sturm, "Risk factors of sexually transmitted infections among migrant and non-migrant sexual partnerships from rural South Africa," *Epidemiol.Infect.*, 2005, *133*, 421–428.

Appendix Table A1

Sensitivity check using narrow definition of unemployment. Marginal effect of being HIV+
from naïve (unconditional) difference in means, logit with
controls, propensity score reweighting (RW) and control function (CF).

Sample	Naïve		Logit		RW		CF		N
Full sample	0.160	***	0.053		0.060	**	0.059	**	6907
	(0.025)		(0.029)		(0.027)		(0.029)		
Male only	0.064		0.038		0.038		0.048		2907
	(0.041)		(0.043)		(0.046)		(0.044)		
Female only	0.177	***	0.061		0.072	**	0.062		3994
	(0.028)		(0.036)		(0.031)		(0.037)		
Matric and up	0.169	***	-0.029		-0.003		-0.019		1748
	(0.054)		(0.040)		(0.062)		(0.042)		
Less than Matric	0.147	***	0.092	**	0.080	**	0.095	**	3390
	(0.033)		(0.037)		(0.036)		(0.037)		

[1] Full sample and results by sex include individuals aged 15 and older. Results by education level
are restricted to individuals aged 25 and older.

[2] Naïve is OLS without controls. Controls used in logit are same as those used in propensity score
estimation.

[3] *** significant at 99 percent level, ** significant at 95 percent level.

Appendix Table A2

Sensitivity check including not economically active (NEA) individuals in the sample.
Marginal effect of being HIV+ from naïve (unconditional) difference in means,
logit with controls, propensity score reweighting (RW) and control function (CF).

Sample	Naïve		Logit		RW		CF		N
Full sample	0.222	***	0.082	***	0.095	***	0.082	***	12032
	(0.021)		(0.021)		(0.023)		(0.021)		
Male only	0.146	***	0.046		0.064		0.065		4529
	(0.037)		(0.034)		(0.041)		(0.035)		
Female only	0.248	***	0.092	***	0.101	***	0.086	***	7488
	(0.025)		(0.026)		(0.027)		(0.026)		
Matric and up	0.199	***	-0.002		0.012		0.001		2173
	(0.051)		(0.035)		(0.060)		(0.037)		
Less than Matric	0.206	***	0.088	***	0.098	***	0.087	***	5739
	(0.028)		(0.029)		(0.031)		(0.029)		

[1] Full sample and results by sex include individuals aged 15 and older. Results by education level are restricted to individuals aged 25 and older.

[2] Naïve is OLS without controls. Controls used in logit are same as those used in propensity score estimation.

[3] *** significant at 99 percent level, ** significant at 95 percent level.
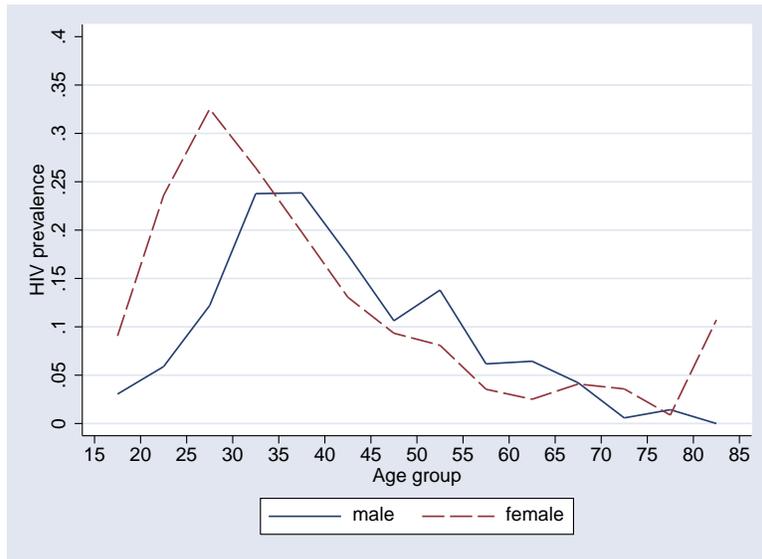
Figure 1: Age profile of naïve HIV prevalence
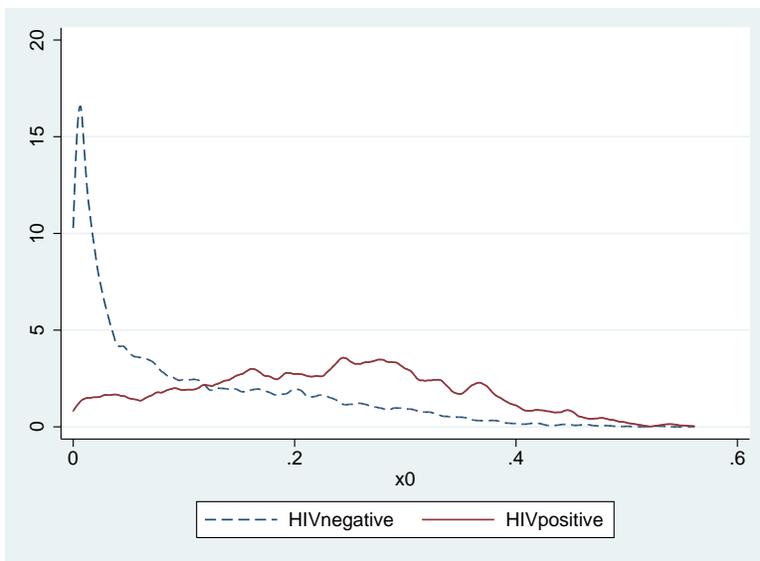
Figure 2: Empirical distributions of the propensity scores