Education, HIV Status, and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter?*

Daniela Iorio

Universitat Autònoma de Barcelona
Barcelona GSE

Raül Santaeulàlia-Llopis

Washington University in St. Louis
Federal Reserve Bank of St. Louis

December 11th, 2011

Abstract

We study the relationship between education and individual HIV status using nationally representative data (Demographic and Health Surveys, DHS) for 18 countries in Sub-Saharan Africa (SSA). Because the diffusion of knowledge on HIV prevention—hence, actual change in sexual behavior—may differ across education groups, we explicitly explore the possibility of a dynamic relationship between education and the probability of being infected with HIV over aggregate stages of the HIV epidemic. Our contribution is twofold. First, we construct an innovative algorithm that positions, for any set of countries, the country-specific evolution of the HIV epidemic in a unified framework—a normalized epidemiological space—to define stages of the HIV epidemic in a comparable manner across SSA countries. Second, using this framework, we exploit epidemiological stage variation across DHS country observations and find that the relationship between education and individual HIV status is dynamic and significantly evolves with the course of the epidemic. Specifically, we show that the education gradient of HIV displays a large U-shaped (positive-zero-positive) pattern over the aggregate stages of the HIV epidemic.

*We thank Andrew Foster, Jane Fortson, Duncan Thomas, Belgi Turan, and Susan Watkins for helpful comments. We thank participants at the XXVI IUSSP International Population Conference (Marrakech) for very helpful comments and discussions. Daniela Iorio acknowledges financial support from ECO2009-11857, ECO2009-07616ECON, the Barcelona GSE, and the government of Catalonia. Raül Santaeulàlia-Llopis thanks the Weidenbaum Center at the Washington University in St. Louis for financial support. We also thank Alannah Glickman, Minho Kim, and Borek Vasicek for excellent research assistance. There is a companion site for this paper at: http://rsantaeulalia.wustl.edu/Stages-of-HIV-Epidemic.html.

1 Introduction

In Sub-Saharan Africa (SSA), where most HIV infections are due to heterosexual intercourse (see UNAIDS (2010)), HIV policy design that specifically distinguishes between highly educated and poorly educated individuals can be particularly relevant if risky sexual behavior—one of the most important margins that policy can affect—differs across education groups. Currently, however, the major international donors in the fight against HIV do not provide any guidance on specific targeting strategies across education groups.[1,2] This lack of policy advice may be motivated by the fact that the current understanding of the sign and size of the relationship between education and HIV status notoriously lacks consensus. That is, the knowledge of which education groups are at major risk of being infected with HIV remains unclear to scholarship, with a large body of mixed evidence that we review below.

In this context, we note one premise that previous work almost invariably shares: The relationship between education and HIV status is assumed to be stationary across stages of the HIV epidemic. Here, we challenge this premise and examine a potentially dynamic relationship between education and the probability of HIV infection by explicitly exploring a nonstationary education gradient of HIV that we hypothesize evolves with the epidemic. That is, could it be that the relationship between education and the probability of HIV infection changes over stages of the HIV epidemic? If yes, by how much?

To the best of our knowledge, we are the first to explore the importance of the stages of the HIV epidemic in a comparable manner across SSA countries. That the HIV epidemic in SSA evolves differently across countries and that these countries are at different stages of the HIV epidemic at any point in time is practically self-evident. In particular, we observe that the peak of HIV prevalence, the year of the HIV peak, the time it takes each country to reach its own peak, and the pace at which each country moves away from its peak differs greatly across

[1] According to Kates et al. (2011), $6.9 billion was given by donor governments to international AIDS assistance in 2010. Of these funds 74% were provided bilaterally (from one country to another), while the remainder went towards multilateral organizations such as UNAIDS and the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM). The United States is by far the largest resource provider for the global fight against AIDS, and it channels its aid through the President's Emergency Plan for AIDS Relief (PEPFAR). Initiated by President George W. Bush for 2003-2008, PEPFAR has continued its activity under the mandate of President Barack Obama, who renewed the efforts for 5 years with few changes in policy implementation.

[2] PEPFAR is, however, precise on the allocation of funds in terms of means. PEPFAR money must be spent as follows: 55% on treatment (mostly the distribution of antiretrovirals); 15% for palliative care; 10% for AIDS orphans; and 20% for prevention programs. Few changes on how means meet ends have been introduced in the second five-year round of PEPFAR funding. One such change is the increase—to about 50% in countries with a generalized epidemic—of money spent on preventing sexual transmission via abstinence, delay of age of first sexual intercourse, monogamy, fidelity, and reduction in the number of sex partners. Welfare assessments of the policy trade-off between HIV treatment versus HIV prevention is, however, a question entirely separate from ours.

SSA countries; Section 4 discusses this cross-country heterogeneity in detail. In our exercise, we take this source of heterogeneity seriously and condition on it to document a set of dynamic relationships among education, individual HIV status, knowledge, and sexual behavior. To do so, we need a definition of the stages of the HIV epidemic that accounts for the large degree of heterogeneity with which the epidemic evolves across countries. Here, we propose an innovative two-dimensional normalization of the HIV epidemic that adjusts for country-specific sizes and associated time paths of the epidemic to define aggregate stages of the HIV epidemic for each individual country in the same normalized space.

Then, using a simple econometric specification that introduces the possibility of variation across epidemiological stages, we find that the education gradient of HIV evolves with the HIV epidemic. Specifically, we find a dynamic relationship between education and individual HIV status that follows a large and significant U-shaped (positive-zero-positive) pattern. During early stages of the epidemic an additional five years of education are associated with 5.5% additional points in the probability of being infected, but as the epidemic progresses past these early stages, this association declines at first to 2.5% and then to 2%. At even later stages of the epidemic, additional years of schooling are associated with no changes in the probability of being infected. Interestingly, this zero HIV-Education gradient reverts to positive in the more advanced stages of the epidemic: about 5% per five more years of education.

To gain a better understanding of this dynamic relationship between education and the probability of infection, we also explore educational disparities in the diffusion of HIV knowledge and actual sexual behavior change across epidemiological stages. First, we find that while more-educated individuals acquire information about the sexual nature of HIV transmission at earlier stages of the epidemic than less-educated individuals, these educational differences in knowledge vanish as the epidemic evolves. Second, we find that as the epidemic builds more-educated individuals significantly change the number of extramarital partners in the past 12 months with respect to less-educated individuals but not other margins of sexual behavior (e.g., condom use, age at first marriage, formation rate of first marriages, and age at sexual debut). Specifically, while more education is associated with more extramarital partners (0.29 per five more years of schooling) at early stages of the epidemic this relationship rapidly declines with the epidemic. Indeed, in mid stages we find that more education is associated with fewer extramarital partners (-.12 per five more years of schooling). Finally, in the most advanced stages of the epidemic, the relationship between education and the number of extramarital partners is nonsignificantly different from zero. We also show the initial decline in the education gradient of HIV is strongly related to the earlier reductions in the number of extramarital partners by the more-educated individuals. However, in the most advanced stages of the epidemic, we find that sexual behavior is quite uniform across age groups. The fact that the last increase in the education gradient of HIV is not associated to sexual behavior change suggests that the effects of earlier use of ARV treatments by the more-educated individuals are starting to be apparent.

The rest of our paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the data we use in our analysis. In Section 4 we document the hetereogeneity of the epidemic across SSA countries and propose a definition for the stages of the HIV epidemic that accommodates that heterogeneity. Our econometric specification to compute the HIV-Education gradient is in Subsection 5.1 and we describe our results in Subsection 5.2. Section 6 reports the educational disparites in the diffusion of knowledge on HIV prevention technologies and in actual risky sexual behavior over stages of the epidemic. In Section 7 we discuss the potential mechanisms underlying the evolution of the HIV-Education gradient using a simple accounting framework. In Section 8, we provide further insights on the relationship between education and HIV status for women, and explore the implications of marital status and aggregate antiretroviral (ARV) coverage on our main results. Section 9 concludes.

2 Related literature

Using nationally representative data from the Demographic and Health Surveys (DHS) for five SSA countries, Fortson (2008) finds education has a positive association with HIV status: adults with primary school are one half more likely to be infected than adults with no schooling once this correlation is conditioned on age, sex, and sector of residence (urban/rural). de Walque (2009) uses the same five DHS countries and finds that incorporating additional controls, such as individual characteristics on marital history and a wealth index, makes education and HIV status uncorrelated. However, when Fortson (2008) further incorporates a quadratic term on education, she finds the education gradient continues to be positive and significant in most of her sampled countries.[3,4] Our analysis differs greatly from this literature in that we incorporate a larger number of DHS countries and exploit the variation across stages of the HIV epidemic for each of these countries when DHS data were collected to study how much the relationship between education

[3]Further, de Walque (2009) also finds that wealth displays a positive association with HIV status. Mishra et al. (2007) find similar results for eight DHS countries in SSA.

[4]It is important to note that if the evaluation of aggregate HIV policies such as those implemented by PEPFAR or GFATM is to be guided using the properties of the education correlates that we intend to document, these properties must unambiguously be extracted from nationally representative datasets. The previous literature that considers nationally representative data to study the relationship among HIV status, education, and risky sexual behavior in SSA is, however, rather small. This can be easily explained by the lack (until very recently) of surveys with this degree of representativeness. The datasets that are filling this gap, and around which a new and exciting literature is rapidly flourishing, are the DHS. Our exercise is one of the first to make extensive use of these data across a large set of SSA countries to study education correlates in the context of the HIV epidemic. A detailed description of the DHS and the countries for which individual HIV testing is available is provided in Section 3.

and HIV status changes over epidemic stages.

While we acknowledge that eliminating the national representativeness property undermines the purposes of aggregate description that we pursue here, we find some aspects of the literature that work with nonrepresentative data relevant for our analysis. A large body of studies emerges if we eliminite the representativeness property.[5] For example, in the context of Manicaland Province, Zimbabwe, Lopman et al. (2007) find HIV prevalence is higher in poorer groups. In KwaZulu-Natal Province, South Africa, Bärnighausen et al. (2007) suggest the highest HIV prevalence is in the middle wealth class. In Limpopo Province, South Africa, Hargreaves et al. (2007) finds no relation between HIV prevalence and wealth.[6] Within this set of nonrepresentative studies, we find particularly interesting the work of de Walque (2007) that incorporates a time-varying dimension to his analysis of several sites in rural Uganda by studying the correlates of education for a span of 12 years (between 1989 and 2001). He finds education and HIV status move from being unrelated to negatively related over time.[7] Whether this is the case in nationally representative samples remains unanswered. Further, here we argue that, given the heterogeneity in both size and the time path of the HIV epidemic across SSA countries, calendar time is not a sufficient statistic to define the stages of the HIV epidemic. A simple example illustrates this by noting that Uganda reached its HIV peak prevalence at 14% in 1989 and after 12 years (de Walque's span of time) its HIV prevalence has declined to 7% (in 2001). However, another country that reaches a similar HIV peak of 14% in 1998, Malawi, after 12 years still has an HIV prevalence of 12% in 2010—a much smaller decline than that of Uganda after the same amount of time since their respective peak years. This exemplifies one of the dimensions in which the HIV epidemic differs across SSA countries—the country-specific speed at which HIV prevalence declines from its peak. We comprehensively describe this and other dimensions of heterogeneity of the HIV epidemic across SSA countries in Section 4. Overall, a more precise definition for the stages of the epidemic that can circumvent the high degree of heterogeneity of the epidemic within SSA is required. We provide such a definition in Section 4.

[5] Surprisingly, most of these studies are conducted outside the economic literature; see the reviews in Hargreaves and Glynn (2002) and Wojcicki (2005). We believe there are two major reasons for the small amount of economic work on the relationship between education and HIV. One reason is the lack of reliable measured data on education and sexual behavior unambiguously linked to individual data on HIV infection that is consistent over time and across countries. Another reason is the above-mentioned unavailability of nationally representative datasets until very recent. We believe the DHS datasets that we use here overcome these problems and will become a major source for economic research on HIV.

[6] In an interesting comparative study across four cities, Glynn et al. (2002) finds that in two cities in Kenya and Zambia educational status and HIV were unrelated. However, in another city in Cameroon highly educated women were less likely to be infected with HIV while education and HIV status was unrelated for men. Finally, in another city in Benin highly educated men were less likely to be infected with HIV while the education and HIV status was unrelated for women.

[7] de Walque (2007) attributes this result to educational disparities in the responsiveness to the Ugandan HIV prevention campaigns (the ABC program: Abstinence, Be faithful, and use Condoms).

Overall, the message is clear: Our current understanding of the sign and size of the relationship between education and HIV provides neither conclusive nor generalizable answers. Similar conclusions are also reached in the reviews of Strauss and Thomas (2007) and Beegle and de Walque (2009). While the current mixed evidence is likely to reflect differences in methodology, sampling strategy, and measures of socioeconomic indicators and HIV status, this may not entirely explain the differing conclusions reached by previous studies. Here, we propose an entirely new look at this problem by providing a definition of the stages of the HIV epidemic that accommodates for the high degree of heterogeneity of the HIV epidemic across SSA countries. This definition of the stages of the HIV epidemic builds a unified framework that makes the country-specific evolutions of the epidemic comparable across countries. Then, in the context of this unified framework, we exploit the cross-country variation of DHS nationally representative datasets across stages of the epidemic to document the dynamic relationship between education and the probability of infection over stages.

Our work is also related to the large set of studies that more broadly describe the relationship between education and health outcomes in diseases other than HIV.[8] Within this group of studies our work relates more closely to those that allow the relationship between education and health to be nonstationary. This is the case of the "fundamental cause" literature described in Cutler et al. (2006)[9] in which the diffusion of information on technological improvements is an argument used to explain the changes in the education gradient in health.[10] Here, in order to investigate the factors underlying the changes in the HIV-Education gradient over stages of the HIV epidemic, we document how the diffusion of knowledge on the sexual nature of the HIV transmission and the use of effective prevention technologies differs across education groups and epidemiological stages. Then, we explore whether our documented diffusion patterns of information on HIV are related to actual changes in risky sexual behavior. The interaction between knowledge regarding HIV and risky sexual behavior has been previously studied in Botwsana by Dinkelman et al. (2006), who find little evidence of sexual behavior change associated with better knowledge. Whether that is finding applies more generally for the entire SSA region—when we control for the stages

[8]Numerous studies document that education is positively associated with a wide range of health outcomes and investments. See, e.g., Elo and Preston (1996), Case et al. (2002), Lleras-Muney (2004), and Kohler et al. (2008).

[9]See also Link and Phelan (1995) and Link et al. (1998).

[10]For example, the more rapid adoption by highly educated individuals of medical innovations and surgical treatmens for heart disease may help to explain the widening of the mortality differentials by education groups in developed countries; see Feldman et al. (1989), Preston and Elo (1995), Goldman and Smith (2005), and Elo (2009). Aizer and Stroud (2010) also find that heterogeneous responses in the smoking behavior across education groups—highly educated individuals respond faster and stronger—to the first publicized report of the negative effects of smoking on health—the 1964 Surgeon General's report "Smoking and Health". de Walque (2004) shows similar evidence for slightly earlier periods (after 1950) when information about the implications of smoking on health started to diffuse.

of the HIV epidemic is an open question that we adress here.

3 The Data

The core of our exercise consists of examining the relationship between HIV and education over the stages of the HIV epidemic. In answering our questions it would be ideal to use nationally representative long panel data for several SSA countries starting in the pre-HIV era. Unfortunately, available nationally representative data are neither long nor panel. However, we show it is possible to construct a normalized path of the patterns of HIV infection by education groups over the stages of the epidemic for several SSA countries. To do so, we combine two sources of data: (i) cross-sectional data from the DHS, and (ii) aggregate data from the World Population Prospects (WPP, 2009) provided by the United Nations.

Demographic and Health Surveys. The DHS are based on nationally representative samples and are available for a large set of SSA countries.[11] While these surveys are primarily health interviews, they also contain cross-sectional information on individual socioeconomic characteristics, knowledge on HIV, several measures of risky sexual behavior (e.g., number of extramarital relationships, condom use and age at sexual debut) and most importantly, a large proportion of adult respondents have been tested for HIV.[12] We use this cross-referenced individual information harmonically collected across SSA countries to conduct our exercise.

We consider the full sample of SSA DHS countries for which individual HIV testing has been conducted (and available as of March 2010): Burkina Faso (2003), Cameroon (2004), Congo (2007), Côte d'Ivoire (2005), Ethiopia (2005), Ghana (2003), Guinea (2005), Kenya (2003), Lesotho (2004), Liberia (2007), Malawi (2004), Mali (2006), Rwanda (2005), Swaziland (2006), Tanzania (2003 and 2007), Zambia (2007), and Zimbabwe (2005-2006). Only one country appears twice in the DHS sample: Tanzania in 2003 and 2007. As we will discuss below, this set of SSA countries provides sufficient observational heterogeneity on the stages of the HIV epidemic to allow us to obtain reliable estimates that evolve over these stages. [13]

Several aspects make DHS datasets appealing for our exercise. An important advantage of these data is that they provide unambiguous individual measures of individual HIV status, education, knowledge on HIV, and risky sexual behavior in a comparable manner across SSA

[11]DHS comprehend countries outside the SSA region. In this paper, we focus on the region that hosts most of the global HIV infections, the SSA region.

[12]Precisely, 87% of the male respondents aged 15 to 50 has been tested for HIV.

[13]Further, our evidence suggests the DHS non-response bias for HIV testing is minimal. We find that statistics for age, schooling, and residence computed for the sample of HIV-tested adults resemble the analogous ones in the overall male sample. These results are available upon request. See also the discussions in Fortson (2008).

countries. First, regarding individual HIV status, the DHS provides a direct measure as individuals have blood testing for HIV, so we need not rely on indirect proxy for HIV obtained from other health outcomes or biomarkers. Second, regarding education variables, DHS collect data on education (i.e., number of years of schooling and maximum degree attained) and also provide an asset-based wealth index.[14] Our preferred choice for measuring education is years of schooling—perhaps the most commonly used measure for education in the previous literature. The reasoning for our choice of years of schooling—rather than the wealth index— is that while wealth (or income for that matter) is influenced by subsequent negative health conditions (such as HIV) or other shocks that will potentially determine one's health status in adulthood, educational attainment is not because, typically, education is completed before individuals in our sample—adults between 15 and 49 years of age—enter adulthood.

Regarding knowledge on HIV we use two DHS questions that have been systematically reported across all countries in our study and that correspond to the knowledge of the sexually transmitted nature of HIV and the knowledge of the use of condoms as prevention technology.[15] Regarding risky sexual behavior, DHS collect data on a wide range of margins of sexual behavior. Here we focus on (i) the number of sex partners (i.e., the extensive margin) in the past 12 months other than spouses and (ii) condom use in last intercourse (i.e., the intensive margin, quality).

In the appendix we extend our analysis to more margins of risky sexual behavior, specifically, the stability of sex partnerships summarized by the age at first marriage and the formation rate of first marriages, as well as the age at sexual debut (i.e., the participation margin).[16,17]

World Population Prospects (2009). The WPP 2009 is constructed by the United Nations (Department of Economic and Social Affairs, Population Division) and provides estimates of the HIV prevalence rates from 1980 until 2008 (at the country level) and their projections from 2009 onward for a large set of SSA countries. These data represent the official 2008 estimates of UNAIDS. While the previous 2006 UNAIDS estimates relied mostly on data aggregations collected

[14]Unfortunately, the DHS do not collect data on income. This task is not easy in many SSA populations as they are mostly rural and a large proportion of the household income is attributed to the household's own agricultural production; see a discussion in Santaeulàlia-Llopis (2011).

[15]Precisely, the questions are "reduce chances of getting AIDS by having 1 sex partner who has no other partners" and "reduce chances of getting AIDS by always wearing a condom."

[16]A detailed description on the construction and sample properties of these variables is provided in Appendix A.

[17]Further, alternative concurrency measures that explicitly consider overlapping partners at the same point in time is available only in a small set of Sub-Saharan DHS countries (see Mishra and Bignami-Van-Assche (2009)) with, unfortunately, insufficient variation across stages of the HIV epidemic to allow a meaningful nonstationary analysis of the type that we do here. However, since HIV is more infectious during the first 6 weeks after infection, concurrency can be an important factor contributing to the rapid spread of HIV. For this reason, we plan to incorporate in our analysis these alternative measures of concurrency as soon as more data on overlapping partners become available in future work.

from antenatal clinics that overestimated prevalence levels, the 2008 UNAIDS data belong to a downward revision largely originated by the appearance of nationally representative surveys such as the DHS. That is, the 2008 UANIDS data that we use do not suffer from overestimation problems. Indeed, the HIV prevalence levels computed from our DHS samples and the HIV prevalence levels from WPP are very similar.[18] The additional data on country-specific ARV coverage used in our robustness exercises are also from the WPP.[19]

Penn World Tables and World Bank Development Indicators. Finally, for all our SSA countries we also use data on real output per capita from the Penn World Tables and data on agricultural share of output from the World Bank Development Indicators. We use these data in our empirical analysis to control for country-specific stages of aggregate economic development.

4 The Stages of the HIV Epidemic

This section builds a definition of the stages of the HIV epidemic. We first discuss a set of challenges that we argue a useful definition must circumvent (Subsection 4.1). To address these challenges we provide an algorithm that normalizes the HIV epidemic of any given set of countries in the same, hence comparable, space (Subsection 4.2). Our definition is provided in Subsection 4.3.

4.1 Challenges for a Definition of the Stages of the HIV Epidemic

Here we describe several dimensions in which the evolution of the HIV epidemic—regarding its level and its time path—is heterogeneous across countries. This heterogeneity in the evolution of the HIV epidemic poses an important set of challenges that we believe a definition of the stages of epidemic must address. For our description purposes we provide some country-specific statistics of the HIV epidemic for our entire DHS sample of SSA countries (Table 2). In addition, Figure 1 shows the country-specific time path of the epidemic for a selected subsample of DHS countries.[20] Using Table 2 and Figure 1, we extract the following important observations.

First, the peak in HIV prevalence differs substantially across SSA countries (see column 1 in Table 2). Further, the HIV prevalence levels largely differ at the time of DHS collection across

[18] See also a detailed discussion of these data in Bongaarts et al. (2008).

[19] Source: United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects: The 2008 Revision. Unpublished Data - Special Tabulations. We thank Patrick Gerland for sharing these data.

[20] This subsample of DHS countries consists of Burkina Faso, Cameroon, Guinea, Lesotho, Malawi, Rwanda and Zimbabwe. This subsample serves expositional purposes only as it is useful to higlight the heterogeneity of the evolution of the HIV epidemic across countries as we describe next. Many other subsample choices would be equally useful.

SSA countries (see column 3). Finally, a large degree of heterogeneity exists for the relative size of the HIV-prevalence (i.e. HIV prevalence over HIV peak) across SSA countries (see column 5).

Second, the time path of the HIV epidemic differs largely across countries. A large degree of heterogeneity exists for the HIV peak year across SSA countries (see column 2 in Table 2); DHS data were collected at different years for different countries (see column 4); and, a large degree of heterogeneity exists for the time distance between the HIV peak year and year DHS data were collected across SSA countries (see column 6). Further, a large degree of heterogeneity exists for the speed by which SSA countries move to the respective HIV peaks and the speed by which SSA countries move away from their respective HIV peaks (see columns 7-10). Several conclusions can be extracted from these observations on the heterogeneity in the time path of the HIV epidemic:

1. **The absolute level of HIV prevalence is NOT a sufficient statistic to describe the country-specific stage of the HIV epidemic.** For example, we use the case of Zimbabwe and Lesotho. The DHS observation of Zimbabwe in 2005 delivers an HIV prevalence of 19.2%, lower than that of Lesotho in 2004, 23.4%. Looking at this statistic only, we would infer that Lesotho is at later stages of the epidemic than Zimbabwe. However, we actually know that Zimbabwe's HIV peak occurred at a higher level and earlier, 29.1% at 2009, than that of Lesotho, 23.8% at 2007, which suggests an opposite ordering over stages. That is, the ordering of DHS countries by HIV prevalence is a mere artefact of the years in which DHS were collected.[21]

2. **The relative level of HIV prevalence is NOT a sufficient statistic to describe the country-specific stage of the HIV epidemic.** One step to resolve the problematic use of the absolute size of the HIV prevalence as a measure of the stages of the epidemic is to compute relative size of HIV prevalence dividing country-specific observations of HIV prevalence by their corresponding HIV peaks. However, this poses, a new set of drawbacks. For example, while Guinea and Liberia have the same relative HIV prevalence, .94, Guinea attains that relative size of .94 before reaching its peak and Liberia does so after reaching its peak, which suggests these two countries are at different stages of the epidemic despite having the same relative HIV prevalence. Another interesting example is the one posed by the Democratic Republic of Congo (DRC) and Ethiopia. At the time of DHS data collection the DRC and Ethiopia have both surpassed their respective peaks and have the same relative prevalence level of .81 (see column 5 in Table 2); however, it took the DRC 19 years to move from its peak to that relative prevalence (see column 6), while it took Ethiopia less than half of that time, about 9 years, to reach the same relative

[21] A similar reasoning occurs if we compare only the DHS observation for Cameroon, 5.8%, versus that of Rwanda, 3.3%, which again results in the opposite ordering to the one attained at their respective peaks: 7.1% for Rwanda and 6.2% for Cameroon.

prevalence. The fact that the transition away from the peak of the DRC substantially slows with respect to that of Ethiopia is, in itself, a phenomenon to which we would like our definition of the stages of the epidemic to be invariant.

Constructively, the arguments posed here against the sole use of the absolute (or the relative) HIV prevalence to define stages of the epidemic also suggest what we need to add to our definition of the stages to resolve the exposed problems: some properties of the time path of the HIV epidemic. That is, we are looking for a definition of the stages of the epidemic that is invariant to both dimensions: country-specific levels and shapes of the HIV time path. To that end, we next provide a two-dimensional (2D) normalization of the evolution of the epidemic.

4.2 A Two-Dimensional Normalization of the Evolution of the Epidemic

This section builds a 2D algorithm that, for all countries, normalizes the country-specific level and time path of the epidemic, thereby making the evolution of the HIV epidemic comparable across countries. Once the evolution of the epidemic is normalized for all countries, the position of each DHS dataset on its associated epidemiological stage (defined in Subection 4.3) readily follows.

Algorithm 1. [A Two-Dimensional Normalization of the Evolution of the HIV Epidemic] *Given the time series of the level of HIV prevalence of each i, we follow three steps to conduct a 2D normalization of the level and time path of the HIV epidemic:*

1. *Interpolate the country-specific time path of prevalence for each country i, $\{\lambda_{i,t}\}_{t_0}^{t_p}$, for $p+1$ interpolation points (years), where p is a positive integer. Then, interpolate the aggregate (across countries) prevalence path as $\lambda_t = \frac{\sum_i^n \lambda_{i,t}\mu_{i,t}}{\sum_i^n \mu_{i,t}}$, where n is the total of number of countries and $\mu_{i,t}$ is the population level of country i at period t. Denote the country-specific interpoland function as $s_i : t \to [0, \max_t \lambda_{i,t}]$, where $\max_t \lambda_{i,t} \in [0,1]$ and $s_i \in \mathcal{S}$, where $\mathcal{S}$ is the collection of functions that can be written as a linear combination of a set of n-known linearly independent basis functions ψ_j, $j = 1, ..., n$,*

$$s_i(t) = \sum_{j=1}^{n} \theta_j \ \psi_j(t)$$

with n unknown θ_j coefficients. Denote the aggregate interpoland as $s(t)$ where $s(t)$ shares the same properties as the country-specific interpolands $s_i(t)$. Importantly, note that $\max_t s_i(t)$ is not necessarily identical across countries or to the aggregate $\max_t s(t)$.

2. *Level normalization*

   (a) *Compute the country-specific peak prevalence,*

   $$s_i(t_*^i) = \max_t s_i(t), \tag{1}$$

   where $t_*^i = \arg\max_t s_i(t)$ *is the period country* i *reaches its peak,* $s_i(t_*^i)$. *Redo equation (1) to obtain the aggregate peak* $s(t_*)$ *and aggregate peak period,* $t_* = \arg\max_t s(t)$.

   (b) *Normalize the country-specific and aggregate interpolands by their respective peak prevalence,*

   $$\widetilde{s}_i(t) = \frac{1}{s_i(t_*^i)}\, s_i(t) \quad \text{and} \quad \widetilde{s}(t) = \frac{1}{s(t_*)}\, s(t),$$

   where $\widetilde{s}_i, \widetilde{s} : t \to \Lambda = [0,1]$ *and* $\arg\max_t s_i(t) = t_*^i = \arg\max_t \widetilde{s}_i(t)$. *Note now that* $\widetilde{s}_i(t_*^i) = \widetilde{s}(t_*) = 1 \ \forall i$.

3. *Time normalization*

   (a) *For* $t_0^i < t^i \leq t_*^i$, *normalize the time interval between the initial period for which data are available,* $t_0^i = 1980$, *and the country-specific peak period,* t_*^i, *by the time interval between the aggregate initial period,* $t_0 = 1980$, *and the aggregate peak period,* t_*. *To do so, we compute the constant of time proportionality for the pre-peak era,*

   $$\alpha_i^L = \frac{t_* - t_0}{t_*^i - t_0^i}.$$

   For $t^i > t_*^i$, *normalize the time between the peak period,* t_*^i, *and the period* t_γ^i *in which country* i *reaches a given threshold* $\gamma \in [0,1]$, *that is,* $t_\gamma^i = \widetilde{s}_i^{-1}(\gamma)$, *over the analogous aggregate interval with* t_* *and* $t_\gamma = \widetilde{s}^{-1}(\gamma)$,

   $$\alpha_i^R(\gamma) = \frac{t_\gamma - t_*}{t_\gamma^i - t_*^i}.$$

   Here, note that t_γ^i *and* t_γ *may not occur at an interpolation node but elsewhere along their respective interpoland.*

   (b) *Normalize the time input of the country-specific interpolands by* α_i^L *and* α_i^R,

   $$\tau = \alpha_i^L(t - t_*^i) \quad \text{for} \quad t \leq t_*^i \tag{2}$$

   $$\tau = \alpha_i^R(t - t_*^i) \quad \text{for} \quad t > t_*^i \tag{3}$$

where $\tau \in T$ are the normalized units of time. Operations (2) and (3) compress/stretch the interpoland[22] *to ensure that for $\tau \leq \tau_*$ (before the peak) the number of normalized periods τ that it takes each country to move from τ_0 to the peak are the same across countries,*

$$\widetilde{s}_i^{-1}(1) - \widetilde{s}_i^{-1}(0) = \widetilde{s}_j^{-1}(1) - \widetilde{s}_j^{-1}(0) = \widetilde{s}^{-1}(1) - \widetilde{s}^{-1}(0) \quad \forall\, i, j\ ,$$

and for $\tau > \tau_*$ (after the peak) the normalized periods τ that it takes each country to move from the peak to a threshold of prevalence γ is the the same across countries,

$$\widetilde{s}_i^{-1}(\gamma) - \widetilde{s}_i^{-1}(1) = \widetilde{s}_j^{-1}(\gamma) - \widetilde{s}_j^{-1}(1) = \widetilde{s}^{-1}(\gamma) - \widetilde{s}^{-1}(1) \quad \forall\, i, j.$$

This allows us to define the evolution of the epidemic for each country and the aggregate,

$$\widetilde{s}_i : \tau \to \Lambda \quad \textit{and} \quad \widetilde{s} : \tau \to \Lambda, \tag{4}$$

in the same—hence comparable—2D normalized space (T, Λ).

Before implementing the algorithm we need to make two choices: the basis functions, $\psi(\tau)$, and the prevalence threshold γ for the time normalization after the peak. First, we specify $\widetilde{s}(\tau)$ as a B-spline with cubic pieces and solve for the θ_j coefficients accordingly.[23] Splines are the most common finite element method of interpolation. Our choice of splines as interpolands obeys our desired manageability of the interpoland given the size of the Lagrangian interpolation problem poised by 71 (1980-2050) interpolation data points.[24] Second, our choice of γ responds to balance between minimizing the use of projection of U.N. data for our set of DHS countries and maximizing the number of countries that have already surpassed the threshold $\gamma =$ at the time of DHS data collection. Our search for this balance suggests a value of $\gamma =.8$. This choice of γ implies that one fourth of the countries in our dataset have already passed the threshold.[25]

Next, we apply our algorithm to the set of SSA DHS countries for which U.N. estimates and projections of the HIV prevalence time path are available.[26] The results are depicted as

[22]The interpoland s_i horizontally compresses when $\alpha_i^k > 1$ with $k = \{L, R\}$ and expands otherwise.

[23]We find our results are robust to the use of linear or quadratic pieces

[24]For instance, the use of a monomial basis—that is, a polynomial interpoland as a spectral method—would require an order-72 polynomial to ensure the interpoland passes through all interpolation points. Here note that interpolation by regression that allows for a polynomial interpoland of degree lower than the number of interpolation nodes does not ensure the interpoland passes through all interpolation points.

[25] Specifically, the countries where the HIV epidemic has surpassed γ at the time of DHS collection are Burkina Faso (2003), DRC (2007), Rwanda (2005), and Zimbabwe (2005).

[26]The interpolation Lagrangian points, that is, the country-specific prevalence time series, $\lambda_{i,t}$, are retrieved

snapshots of 5 year intervals from 1980 to 2015 in Figures A-1 through A-8 (see Appendix 9) where the normalized level of HIV prevalence is on the vertical axis and the normalized time on the horizontal axis. This way, the level and time path of the epidemic are entirely comparable across countries.[27]

What is most relevant is to find where to position each DHS observation along each of the the country-specific normalized epidemics. The position of each country i on its normalized HIV time path at the period t_{DHS} at which its respective DHS data were collected can be easily computed by solving for τ_i in

$$\widetilde{s}_i^{-1}\left(\frac{\lambda_{i,t_{DHS}}}{s_i(t_*^i)}\right) = \tau_i. \tag{5}$$

The results of this exercise are shown in the scatterplot in Figure 2. Each data point $(\tau_i, \widetilde{s}_i)$ in Figure 2 represents a DHS dataset. The main observation we extract from this plot is that for a relatively small span of years for which our DHS sample was collected—from 2003 to 2007—there is a large degree of heterogeneity across country positions over the normalized HIV epidemic.

4.3 A Definition of the Stages of the HIV Epidemic

Next, we define the stages of the HIV epidemic as intervals of a continuous real variable,

$$\omega(\tau,\zeta) = \frac{\zeta}{\tau - \tau_*} \to \mathcal{R}^1, \tag{6}$$

with the pair (τ,ζ) belonging to the 2D normalized space, $T\times\Lambda$. Geometrically, $\omega(\tau,\zeta)$ represents the slope of the arrays from the origin in the (T,Λ) space with the following limiting properties:

$$\lim_{\tau\to\tau_*^-}\omega(\tau,\zeta) = -\infty,\ \lim_{\tau\to\tau_*^+}\omega(\tau,\zeta) = \infty,\ \text{ and }\ \lim_{\tau\to-\infty}\omega(\tau,\zeta) = \lim_{\tau\to+\infty}\omega(\tau,\zeta) = 0. \tag{7}$$

Definition 1. [Stages of the HIV Epidemic] *Given a set of stage thresholds* $\{\zeta_0, ..., \zeta_j, ..., \zeta_n\}$ *with* $\zeta_j > \zeta_{j+1}$ *for all* j*, the stage* j *of the HIV epidemic consists of all pairs* $(\tau,\zeta)\in T\times\Lambda$ *such that* $\omega(\widetilde{s}^{-1}(\zeta_{j+1}),\zeta_{j+1}) \le \omega(\tau,\zeta) \le \omega(\widetilde{s}^{-1}(\zeta_j),\zeta_j)$*, where* $\widetilde{s}(\tau)$ *is the normalized (population-*

from the U.N. population division estimates (1980-2008) and projections (2009-2050, medium-variant).

[27] Figures A-1 through A-8 show the results of our algorithm in the form of snapshots (i.e., every five years). Additionally, we supply a companion site for this article http://rsantaeulalia.wustl.edu/Stages-of-HIV-Epidemic.html that hosts the outcome of our algorithm in a more user-friendly video form with interactive applications. Options such as tracking one or several countries over aggregate epidemiological stages, 2D normalizations by SSA regions or worldwide regions, and changing the size of the bubbles to represent other variables such as population or HIV prevalence are also available in that interactive video form.

weighted) aggregate of the HIV epidemic defined in (4).

Our choice for the stage thresholds $\{\zeta_0, ..., \zeta_j, ..., \zeta_n\}$ pursues the maximization of both countries per stage of the epidemic and number of stages. To do so, we set $\zeta_0 = 1$ and $\zeta_j = \zeta_1 - .05j\ \forall j$. This implies the following allocation of DHS datasets, $\omega(\tau_i, \widetilde{s}_i(\tau_i))$, over stages of the epidemic as follows.

- Stage ≤0: Guinea (2003), Mali (2006), Swaziland (2006), and Lesotho (2004).[28]
- Stage 1: Cameroon (2004), Malawi (2004), Liberia (2007), Zambia (2006), and Tanzania (2003).
- Stage 2: Ghana (2003), Ethiopia (2005), and Tanzania (2007).
- Stage 3: Kenya (2003), Cote d'Ivoire (2005), and DRC (2007).
- Stage ≥4: Burkina Faso (2003), Rwanda (2005), and Zimbabwe (2005).

Figure 2 depicts these allocations in two dimensions in the normalized space $T \times \Lambda$. Further, Figure 3 zooms in the quadrant $[0, 10] \times [.7, 1]$ in $T \times \Lambda$. It is easily seen that the large heterogeneity across stages of the HIV epidemic is sufficient to provide reliable nonstationary estimates.[29]

5 The HIV-Education Gradient

Our empirical analysis consists of posing a simple econometric specification suitable for documenting the potentially nonstationary properties of the HIV-Education gradient over stages of the HIV epidemic.

5.1 Econometric Specification

We consider two model specifications, a linear probability model (LPM) and a linear model (depending on the dependent variable of interest), where the HIV-Education gradient is allowed

[28]Note that while Guinea (2003) and Mali (2006) had not yet reached their respective HIV peak at the time DHS data were collected, we include them in stage 0 since they are very close to the peak.

[29] We add two remarks regarding two SSA DHS countries for which individual HIV status data are available. First, for the country of Niger we only have U.N. data for 1980-2007; in that interval of time, Niger attained its HIV peak in 2006-07. However, we lack a projected path for Niger. This means that we can normalize the rising side of Niger's HIV epidemic but not its declining side. Second, for Senegal we not only have data for a restricted period, 1980-2007, but, unlike Niger, Senegal poses an additional drawback. Its HIV epidemic had not peaked when the DHS data were collected. Hence, for Senegal we unable to normalize both the rising and the declining sides of its HIV epidemic. For these reasons, we omit Niger and Senegal from our benchmark analysis. We find that the inclusion of Senegal and Niger in Stage 0 of the epidemic does not qualitatively alter our findings.

to change over the stages of the HIV epidemic defined in Section 4. Let s_{ij} denote the educational attainment of an individual i who lives in stage j of the HIV epidemic. Also, let y_{ij} be the individual's HIV status, a dummy variable equal to 1 if the individual HIV testing result is positive and zero otherwise. We estimate a linear projection of the type

$$y_{ij} = \alpha_0 + \sum_{j>0} \alpha_j \mathbf{1}_j + \left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j \right) s_{ij} + \beta\, x_{ij} + \psi\, m_{gj} + \varepsilon_{ij}, \tag{8}$$

where ε_{ij} is a contemporaneous term that is independently and identically distributed (across individuals) $N(0, \sigma_\varepsilon^2)$, and $\mathbf{1}_j$ is an indicator function that is equal to 1 when the stage of the HIV epidemic is j and zero otherwise. That is, if the stage of the epidemic is $j = 0$, then the intercept is α_0 and the slope is γ_0. However, if the stage of the epidemic is $j > 0$, the associated intercept is $(\alpha_0 + \alpha_j)$ and the slope is $(\gamma_0 + \gamma_j)$. Therefore, α_j measures the proportionate difference in the probability of being infected with HIV in the aggregate epidemiological stage j relative to stage $j = 0$, and γ_j is the difference in the HIV-Education gradient between individuals in stage j of the epidemic and individuals in stage $j = 0$. That is, we explicitly exploit the cross-country variation in the epidemiological stages of the HIV epidemic. While we believe it would be interesting to also explore the within-country variation (e.g., across regions), this is not possible because we cannot recover the stage of the HIV epidemic at this level of disaggregation.[30]

The vector x_{ij} corresponds to a set of individual characteristics that are likely to be correlated with both education and HIV status. Hence, controlling for these characteristics reduces the impact of omitted variables bias. Precisely, we find it is important to control for the type of area in which agents live because the HIV prevalence is, on average, higher in urban than in rural areas (respectively, 5.4% and 3.1% for our entire sample); it is within urban areas where adult education levels are also higher (the average number of years of schooling in rural areas is 3.36, while in urban areas it reaches 4.11). Thus, a positive association between education and HIV may be driven by the fact that people living in urban areas are more likely to be both HIV-positive and more educated. Similarly, we also control for age because HIV prevalence is increasing with age (the DHS age sample is 15-49 years), and education is negatively correlated with age since younger cohorts are more educated than older cohorts.[31]

Last, we control for country-specific aggregate economic variables, m_{gj}, that correct for the current stage of economic development for each country.To do so, we use country-specific measures of output per capita and, following the literature on structural transformation that

[30] Note that in order to define stages of the HIV epidemic across regions using the methodology described in Section 4, we need data on the time paths of aggregate HIV prevalence by region that are—to date—not available.

[31] Here, note that for a given year, age effects are cohort effects.

accounts for development processes (e.g., Hansen and Prescott (2002), Gollin et al. (2002), Gollin et al. (2007), and Herrendorf et al. (2009)), we also use country-specific agricultural shares of output. All our specifications are weighted least squares regressions, where the weights are proportional to the relative population size of each country. By doing so, when we pool a number of countries in the same stage of the HIV epidemic, the relative DHS sample size of a given country corresponds to the relative population size of the country. We then combine these weights with the individual weights provided by the DHS surveys. The results are robust when we cluster the standard errors at the country level and use the individual weights.[32]

5.2 Results

Table 3 shows the behavior of the HIV-Education gradient under the stationary and nonstationary econometric specifications of the linear probability model where the dependent variable is the individual's HIV status.

5.2.1 Stationary Specification

The econometric specification commonly used to study the HIV-Education gradient considers only a stationary gradient, that is, $\alpha_j = \gamma_j = 0$ for all $j > 0$ in (8). Hence, without taking into account the stages of the epidemic, the stationary specification is simply a restricted version of the econometric model (8): its estimates are listed in column 1 in Table 3. Pooling all the DHS countries for which data are available data on individual HIV testing, we find the HIV-Education gradient is highly significant and positive, about .4%, holding age and area of residence constant, which means the probability of being HIV-positive increases by about .4% per additional year of schooling. This suggests that completing 5 additional years of schooling increases the probability of being HIV-positive by 2%, which is not small if we consider that the average prevalence is at 4.5%. In other words, five years of schooling (more than the average) increases the likelihood of individuals being HIV infected by between one-third and one-fourth, precisely, by a factor of (5.8-4.5)/4.5=.29 with respect to the average. Further, HIV positivity significantly increases with age and is significantly higher in urban areas. As we expected, these findings are consistent with those obtained by Fortson (2008), who specifies a similar stationary econometric model for a subset of the DHS countries considered in this study.

Result 0. *If we assume the HIV-Education gradient is stationary, the gradient is positive.*

Next we explore how much this gradient changes over aggregate epidemiological stages.

[32] The standard errors obtained with this procedure are in "Sheet: HIV-Education, cluster" in our online appendix.

5.2.2 Nonstationary Specification

Our nonstationary specification follows the econometric model posed in equation (8); the estimates are reported in columns 2, 3, and 4 in Table 3. These columns differ solely in (i) the set of controls we include and (ii) whether we allow for the intercept and the HIV-Education gradient to be stage-specific. Column 2 represents our full econometric specification (henceforth, our benchmark model). Column 3 represents the minimal departure (from column 1) that accommodates for the HIV-Education gradient to be stage specific. Finally, column 4 augments the model presented in column 3 to include stage-specific intercepts.

Our benchmark specification shows how the evolution of the HIV epidemic shapes the relationship between education and HIV status. The main finding is the following:

Result 1. *The HIV-Education gradient is significantly nonstationary and displays a large positive-zero-positive U-shaped pattern over the stages of the HIV epidemic.*

This result raises a warning flag against the standard assumption that the gradient is stationary over stages of the HIV epidemic. The properties of the nonstationary behavior of the HIV-Education gradient are perhaps better understood by looking at Figure 4, which displays an isomorphic representation of the estimated results for our stationary and nonstationary benchmark models with 95% confidence intervals. Specifically, at Stage 0 an additional years of schooling raises the probability of being HIV-positive by 1.1%. That is, when individuals live in a country that is close to its HIV peak, the HIV-Education gradient is significantly positive and remarkably high (about three times larger than the stationary specification gradient). As the HIV epidemic evolves, the HIV-Education gradient starts to sharply decline. The decline of the gradient is such that while an additional year of schooling at Stage 1 raises the probability of being infected by 1.1-.6=.5%, at Stage 2 this probability is similar to its stationary counterpart: 1.1-.7=.4%. In other words, at Stage 2 the rise in the probability of being infected associated with one additional year of schooling is already about one-third of its value at the earliest stage. The gradient continues to decline to the point where its nonstationarity becomes dramatically self-evident at Stage 3, where the gradient is not significantly different from zero.[33] That is, in contrast with previous stages of the epidemic, at Stage 3 an additional year of schooling does not increase the probability of being infected. Finally, Stage 3 also represents an inflexion point. When the epidemic moves farther away to Stage 4, the HIV-Education gradient again turns significantly positive: 1.1-.1=1%.

The changes of the HIV-Education gradient are significant across stages of the HIV epidemic.

[33] In earlier versions of this paper, we showed that at Stage 3 of the epidemic the HIV-Education gradient can be negative. This is the case when we remove western African countries from the analysis (i.e., those countries with the lowest HIV prevalence levels).

That is, the difference between the estimates of γ_j and γ_{j-1} are statistically significant for all j. Specifically, we test whether the HIV-Education gradient at stage j and the HIV-Education gradient at stage $j+1$ (i.e., two consecutive stages) are significantly different from each other and for all j. We find that all these differences are significant. The difference between γ_0 and γ_1 is significant with a p-value of 0.000, the difference between γ_1 and γ_2 has a p-value of 0.030, between γ_2 and γ_3 the p-value is 0.005, and between γ_3 and γ_4 the p-value is 0.000.

Regarding the control variables, the estimated coefficients of age and area of residence remain significantly positive, and are quite similar to those in column 1.[34] The aggregate economic variables, output per capita, and agricultural share of output are also significant. While output per capita is negatively related with the individual probability of being infected, interestingly, a lower agricultural share of output (which is typically associated with a higher degree of economic development) increases the individual probability of being infected. Thus, there are two offsetting effects in terms of aggregate economic development: One generated by the level of output per capita that pushes the probability of infection downward, and one generated by the composition of output that pushes the probability of infection upwards as countries move from agricultural to industrial production. The latter is consistent with agricultural goods being produced in rural areas, and industrial goods in urban areas. Finally, the econometric specifications reported in columns 3 and 4 of Table 3 also show a significant nonstationary U-shaped pattern of the HIV-Education gradient. In particular, Stage 3 also represents an inflexion point for the HIV-Education gradient in these specifications.

6 The Role of Knowledge on HIV and Sexual Behavior Change

In SSA—where the majority of the HIV infections occur through sexual intercourse (see UNAIDS (2010))—a natural candidate to account for the nonstationary behavior of the HIV-Education gradient is the difference in risky sexual behavior across education groups. The hypothesis that the patterns of risky sexual behavior in response to the HIV epidemic differ across education groups builds on the argument that more-educated individuals tend to be better informed about how HIV is transmitted and, hence, are more likely to adopt safer sexual practices and use prevention technologies. If this hypothesis were correct, we should observe disparities on the diffusion of knowledge on HIV across education groups and a decline in the risky sexual behavior of more-educated individuals relative to that of less-educated individuals over stages of the HIV epidemic.

[34] While we introduce age linearly, we do find that the estimated coefficients for the HIV-Education gradient are robust when age enters nonlinearly. For example, when we add a squared term for age, the estimated γ_0 is 0.012, γ_1 is -0.008, γ_2 is -0.009, γ_3 is -0.014, and the γ_4 is -0.008; the estimated coefficient for age is 0.013 and for age squared -0.0002. All these coefficients are significant at the 1% level.

Next, we explore these two observations.[35]

6.1 The Knowledge-Education Gradient

The DHS provides two questions regarding ways to avoid HIV infection that are informative about the knowledge of respondents on the sexual nature of HIV transmision and, in particular, on the effectiveness of two specific prevention strategies. Specifically, respondents answer two questions: (i) *"Can you (the respondent) reduce the chances of getting HIV by having one sex partner who has no other partners?"* and (ii) *"Can you (the respondent) reduce the chances of getting HIV by always wearing a condom?"*.[36,37]

Table 4 shows the education gradients for knowledge regarding HIV under econometric specification (8), where y_{ij} refers to either the knowledge of the reduction of sexual partners (column 1) and the knowledge of the use of condoms (columns 2-4) as prevention strategies. In column 1, our working sample is unrestricted, while we condition the sample to those individuals who have at least one (two) extramarital partner(s) in column 2 (column 3), and to those individuals who report having had last intercourse with a casual partner in column 4 (since we do not want fertility decisions to interfere in the reasons behind condom use). Overall, we find that more-educated individuals are unambiguously better informed and more rapidly informed about both these HIV prevention strategies. Further, our estimation results for the nonstationary Knowledge-Education gradients reveal interesting diffusion patterns of HIV knowledge over aggregate epidemiologi-

[35] Albeit not using nationally representative data, a small set of papers describes changes in sexual behavior across education groups over time. In the context of south west Uganda, de Walque et al. (2005) finds that highly educated individuals are those who are more likely to adopt safer sexual behavior once the mortality and morbidity effects of AIDS become clearer as the epidemic evolves. de Walque (2007) also finds greater responsiveness among more-educated individuals to HIV prevention campaigns in several sites of rural Uganda.UNAIDS (1997), Asiimwe-Okiror et al. (1997), Stoneburner and Low-Beer (2004), and UNAIDS (2005) provide further evidence of sexual behavior changes. Finally, a precise indication of sexual behavior change in response to a specific policy-driven change in the environment is in Dupas (2011). She finds, in a randomized experiment conducted in 328 primary schools in Kenya, that information campaigns on the relative risk of infection by the age of sexual partner moves females to choose younger partners. However, neither of these studies discusses whether there are differential changes in sexual behavior across education groups, which is the main focus of our paper.

[36] DHS pose further general questions regarding the knowledge of sexually transmitted diseases (STDs) and HIV (e.g., *"Have you ever heard of STDs?", "Have you ever heard of HIV?", and "Can a healthy person have HIV?"*), as well as regarding ways to avoid HIV infection that are not related to sexual behavior (e.g., whether a person can get HIV from mosquito bites, from sharing food with someone who has HIV, and from witchcraft). However, in our exercise we are explicitly interested in exploring the interaction between knowledge on HIV and actual sexual behavior change. For this reason we focus on questions regarding knowledge on HIV that are related to sexual behavior as stated above. Further, answers to these two questions have been systematically reported by the DHS across all countries and years in our study, which is not always the case for the other knowledge questions.

[37] Dinkelman et al. (2006) also look at these two specific questions regarding the knowledge on HIV for the case of Botswana. However, they do not look at cross-sectional variation across educations groups and over stages of the HIV epidemic, which is our focus here.

cal stages. These diffusion patterns are shown in Figure 5 for selected Knowledge-Education gradients. The results are clear and can be summarized as follows:

Result 2. *The Knowledge-Education gradients show a significantly nonstationary inverted-U shape over stages of the HIV epidemic: More-educated individuals have more HIV knowledge at earlier stages of the epidemic, but these educational differences vanish as the epidemic evolves.*

Specifically, while there are no significant educational differences on HIV knowledge at Stage 0, at Stage 2 five more years of schooling are associated with 5*1.5=7.5% more chances of knowing that fewer sexual partners reduces the risk of infection (see column 1 in Table 4) and with 5*2.3=11.5% more chances of knowing that condom use reduces the risk of infection (see column 2) conditional on having at least one extramarital partner. This educational gap on HIV knowledge stops growing as the epidemic evolves further and tends to disappear in the most advanced stages of the epidemic. These results are robust to having more than one extramarital partner (see column 3) or having casual sex (see column 4).

Next, we investigate whether these educational disparities in the diffusion of knowledge on HIV prevention strategies are reflected in educational disparities in the actual adoption of safer sexual practices: Are more-educated—hence, better-informed—individuals more likely to adopt safer sexual practices?

6.2 The Risky Sex-Education Gradient

Here, we study the relationship between education and alternative measures of risky sexual behavior. The margins of risky sexual behavior on which we focus are those associated with the HIV knowledge questions in the previous section: (i) the number of sex partners other than spouses during past 12 months (henceforth, the Partners-Education gradient)[38] and (ii) a dummy variable equal to 1 if the respondent used a condom during the last intercourse (henceforth, the Condom-Education gradient).[39] In Appendix A, we explore additional margins of risky sexual behavior, including age at first marriage, the formation rate of first marriages, and age at sexual debut.

First, we use our nonstationary benchmark model to compute the Partners-Education gradient.

[38] For the extensive margin of risky sexual behavior we use the number of sex partners other than spouses (i.e., extramarital partners) in the past 12 months and note that for individuals who are single or do not cohabit, all sex partners are extramarital. Our variable choice is explained by the fact that the number of extramarital partners in the past year does not have such a shortcoming, while the alternative of using the total number of lifetime partners as endogenous variable inevitably inherits behavior of the pre-HIV era.

[39] When y_{ij} is a binary variable, as in the HIV-Education gradient and the Condom-Education gradient, we find the coefficients of our LPM and the partial effects from an analogous probit model are quite similar in terms of size and significance. For ease of exposition, we report and discuss the results for the LPM only. The estimates obtained with a probit model are in "Sheet: HIV-Education PROBIT" in our online appendix.

For this computation, y_{ij} in (8) now is the number of sex partners other than spouses in the past year (the extensive margin of risky sexual behavior). The results are shown in column 1 in Table 5 and summarized by the following:

Result 3. *The Partners-Education gradient is significantly nonstationary over stages of the HIV epidemic: More-educated individuals have more extramarital partners at earlier stages of the epidemic, but after a decline and then a reversal in the sign (from positive to negative) of the relationship between education and partners, the educational differences in the number of extramarital partners disappear in the most advanced stages of the epidemic.*

Specifically, at Stage 0, five more years of schooling are associated with .057*5=.29 more extramarital partners. That is, for countries close to the HIV peak, individuals with 5 more years of schooling have more than half (precisely, .29/.47=.61) as many more extramarital partners than average. The Partners-Education gradient starts to decline moving away from the peak and shows large behavioral changes across educational groups over stages of the epidemic. At Stage 1, five more years of schooling are associated with (.057-.048)*5=.05 more extramarital partners (close to one-sixth of its initial value at Stage 0) with respect to average. Stage 2 shows again large declines for the Partners-Education gradient with respect to its peak value, although with a small reverse with respect to the previous stage. While the Partners-Education gradient in both Stages 1 and 2 is significantly lower than that of Stage 0, the difference between Stages 1 and 2 is not significant. The decline is quantitatively more evident in Stage 3, where we find a significant reverse in the sign of the gradient. Precisely, at Stage 3, 5 more years of schooling are associated with significantly (.057-.080)*5=-.12 fewer extramarital partners in the past 12 months, an extraordinarily large behavioral change across educational groups that implies an absolute drop of (.29+.12)=.41 extramarital partners from Stage 0 to Stage 3 for every 5 years of schooling. Finally, the Partners-Education gradient becomes not significantly different from zero at the most advanced Stage 4.

Second, we compute the Condom-Education gradient. Table 5 reports the Condom-Education gradient using three different samples. Since the choice of condom use as HIV prevention technology is likely to be related to the frequency of respondents' engagement in risky sexual activities,[40] we study specifically the sample of respondents who engage in these type of activities. First, we consider the respondents with at least one sexual partner other than spouses in the past 12 months, $n > 0$ (see column 2). Second, we consider the further restricted sample of at least two extramarital partners, $n > 1$ (see column 3). Finally, we partition our sample by the type of sex partner with whom individuals had last sexual intercourse, and focus on those who had sex with

[40]See our discussion in Section 3 on the frequency of condom use reported in Table 1.

casual partners (i.e., commercial sex workers or casual acquaintances) (see column 4). Our main result is:

Result 4. *The Condom-Education gradient does not evolve over stages of the HIV epidemic: There are no main educational disparities in condom use over stages of the HIV epidemic.*

Specifically, for those with at least one extramarital partner, 5 more years of schooling imply 2.1*5=10.5% more chances of using a condom at Stage 0 of the epidemic (see column 2 in Table 5). While the gradient at Stage 0 is significant, its size is rather small and the gradient does not display significant changes after Stage 0. For individuals with at least two extramarital partners or who had casual sex in the last intercourse, the use of condoms does not show significant differences across education groups in any stage of the HIV epidemic (see, respectively, columns 3 and 4 in Table 5).

In summary, we found that educational disparities in the knowledge that reducing the number of sexual partners is a useful HIV prevention technology (see Result 2) are accompanied by actual reductions in the number of extramarital partners (see Result 3) across education groups. In particular, at the earliest stages of the epidemic, when there are no HIV knowledge differences across education groups, more-educated individuals have more extramarital partners than less-educated individuals. However, as the diffusion of knowledge occurs in favor of the more-educated individuals, the association between number of extramarital partners and education declines and actually becomes negative. Finally, in the most advanced stages of the epidemic we find that the diffusion of knowledge is more complete and symmetric across education groups and, consistently, we do not find education disparities in the number of extramarital partners. This is not the case with condom use because even though we find educational disparities in the diffusion of knowledge that condom use helps prevent infection—an information assymetry that favors the more-educated—(see Result 2) we do not find significant differences in the actual use of condoms across education groups (see Result 4). Similarly to the findings for condom use, when we extend our analysis to more margins of risky sexual behavior (see Appendix A), we find either empirical evidence of change across education groups as the epidemic evolves (e.g., age at first marriage) or small associations with the education gradient of HIV (e.g., the formation rate of first marriages and age at sexual debut).

7 What Is Behind the Evolution of the HIV-Education Gradient? A Discussion

Here we take a pure accounting perspective to explore the factors underlying the evolution of the HIV-Education gradient described in Subsection 5.2. Our goal is to document dynamic relationships between the U-shaped behavior of the education gradient of HIV and a set of potential explanations for it, specifically, risky sexual behavior, ARVs, and human capital accumulation. Note that to determine actual mechanisms of causation requires the endogenization of the factors that we explore, which is beyond the scope of this paper and we leave for immediate future research. Further, these mechanisms (risky sex, ARVs, and human capital choices) have potentially complex interrelationships. In this context, the discussion in this Section on the empirical dynamic relationships among these factors is informative about which set of factors is useful to endogeneize and when.

To structure our discussion, we construct a simple epidemiological framework using a Markovian process that tracks the population between the ages of 15 and 49. This population is heterogenous in two dimensions: (i) completed educational attainment defined by schooling years s and (ii) HIV status, either negative or positive $h = \{-,+\}$. Denote by $\mu^h_{s,\tau}$ the measure of individuals with s years of schooling and HIV status h in stage of the epidemic τ. Then, the evolution of the population of any given education group s over epidemiological stages can be tracked by

$$\begin{bmatrix} \mu^+_{s,\tau} \\ \mu^-_{s,\tau} \end{bmatrix} = \begin{bmatrix} \gamma^+_s & \lambda_s\,\gamma^-_s \\ 0 & (1-\lambda_s)\,\gamma^-_s \end{bmatrix} \begin{bmatrix} \mu^+_{s,\tau-1} \\ \mu^-_{s,\tau-1} \end{bmatrix} + \begin{bmatrix} \phi^+_{s,\tau-1} \\ \phi^-_{s,\tau-1} \end{bmatrix}, \tag{9}$$

where $\gamma_s = \{\gamma^+_s, \gamma^-_s\}$ are the survival rates of individuals with s schooling years and these differ by HIV status, λ_s is the rate at which individuals with s schooling years become infected with HIV, and $\phi_{s,\tau-1} = \phi^-_{s,\tau-1} + \phi^+_{s,\tau-1}$ denotes the new adult population—individuals who are becoming adults at stage τ with schooling years s.

Using the population law of motion defined in (9), the probability of being HIV-positive for individuals in education group s in stage τ (i.e., the HIV prevalence of schooling group s in stage τ) is

$$HIV_{s,\tau} = \frac{\mu^+_{s,\tau}}{\mu^+_{s,\tau} + \mu^-_{s,\tau}} = \frac{\gamma^+_s\,\mu^+_{s,\tau-1} + \lambda_s\,\gamma^-_s\,\mu^-_{s,\tau-1} + \phi^+_{s,\tau-1}}{\gamma^+_s\,\mu^+_{s,\tau-1} + \gamma^-_s\,\mu^-_{s,\tau-1} + \phi^+_{s,\tau-1} + \phi^-_{s,\tau-1}}, \tag{10}$$

and we compute the HIV-Education gradient between education groups s and $s+1$ at τ as[41]

$$G(\mu_{s+1,\tau}, \mu_{s,\tau})(\lambda_{s+1}, \lambda_s, \gamma_{s+1}, \gamma_s, \phi_{s+1,\tau}, \phi_{s,\tau}) = \frac{HIV_{s+1,\tau+1} - HIV_{s,\tau}}{HIV_{s,\tau}}. \tag{11}$$

We are interested in which factors can generate changes in the HIV-Education gradient over epidemiological stages. We can see from equations (10) and (11) that three mechanisms can explain the evolution of the gradient: changes in the rate of new HIV infections across education groups $\{\lambda_{s+1}, \lambda_s\}$, changes in the survival rates across education groups $\{\gamma_{s+1}, \gamma_s\}$, and changes in the education and HIV composition of individuals that reach adulthood $\{\phi_{s+1}, \phi_s\}$. Further, after some algebra, the following model implications arise:

$$\frac{\partial G}{\partial \lambda_{s+1}} > 0,\ \frac{\partial G}{\partial \lambda_s} < 0,\ \frac{\partial G}{\partial \gamma^+_{s+1}} > 0,\ \frac{\partial G}{\partial \gamma^+_s} < 0,\ \frac{\partial G}{\partial \phi^+_{s+1}} > 0,\ \text{ and }\ \frac{\partial G}{\partial \phi^+_s} < 0. \tag{12}$$

That is, an increase (decrease) in the rate of new HIV infections of the more educated (i.e., λ_{s+1}) relative to the less educated (i.e., λ_s) increases (decreases) the HIV-Education gradient. Further, within the HIV-positive population, an increase (decrease) in the survival rate of the more educated (i.e., γ^+_{s+1}) relative to the less educated (i.e., γ^+_s) implies an increase (decrease) in the HIV-Education gradient. Finally, increases (decreases) in the number of more-educated individuals in the education composition of HIV-positive individuals who become adults (i.e., ϕ^+_{s+1}) relative to the number of less educated (i.e., ϕ^+_s) implies increases (decreases) in the HIV-Education gradient.

The main finding of our paper is that the HIV-Education gradient displays a U-shaped pattern with the evolution epidemic (see Result 1 in Subsection 5.2). Following the implications of our epidemiological model, the initial decline in the HIV-Education gradient between Stage 0 and Stage 3 must due to either declines in the relative rate of new HIV infections of more-educated individuals, declines in the relative survival rates of the more educated, or declines in the number of individuals who enter adulthood being both better educated and HIV positive. In this context, analyzing these mechanisms one by one, first, we find that the more rapid adoption of safer sexual practices by the more educated—prominently, reductions in the number of extramarital partners as documented in Subsection 6.2—is strongly related to the decline in

[41] Precisely, one additional year of schooling is associated with a change in the probability of HIV infection by

$$G(\mu_{s+1,\tau}, \mu_{s,\tau})(\lambda_{s+1}, \lambda_s, \gamma_{s+1}, \gamma_s, \phi_{s+1,\tau}, \phi_{s,\tau}) = \frac{\frac{\gamma^+_{s+1}\ \mu^+_{s+1,\tau-1} + \lambda_s+1\ \gamma^-_{s+1}\ \mu^-_{s+1,\tau-1} + \phi^+_{s+1,\tau-1}}{\gamma^+_{s+1}\ \mu^+_{s+1,\tau-1} + \gamma^-_{s+1}\ \mu^-_{s+1,\tau-1} + \phi^+_{s+1,\tau-1} + \phi^-_{s+1,\tau-1}}}{\frac{\gamma^+_s\ \mu^+_{s,\tau-1} + \lambda_s\ \gamma^-_s\ \mu^-_{s,\tau-1} + \phi^+_{s,\tau-1}}{\gamma^+_s\ \mu^+_{s,\tau-1} + \gamma^-_s\ \mu^-_{s,\tau-1} + \phi^+_{s,\tau-1} + \phi^-_{s,\tau-1}}} - 1,$$

and note that $\gamma^-_s \mu^-_{s,\tau-1} + \phi^+_{s,\tau-1} + \phi^-_{s,\tau-1} > \lambda_s \gamma^-_s \mu^-_{s,\tau-1} + \phi^+_{s,\tau-1}$.

the HIV-Education gradient. Precisely, the association between the Partners-Education gradient and the HIV-Education gradient is high and positive as depicted in Figure 6 and is summarized by a correlation coefficient of .92 between Stages 0 and 3. That is, declines in the HIV-Education gradient are strongly associated with declines in the number of extramarital partners of the more-educated relative to the less-educated population.

Second, in our exercise we cannot provide direct evidence of the role of educational disparities in survival rates because, unfortunately, DHS do not collect data on ARV use—a source of information that, if available, would be extraordinarily valuable for our purposes. However, explaining the decline in the education gradient of HIV between Stages 0 and 3 using differences in survival rates would require the educational gap in survival rates—in particular, within the HIV infected population—to narrow during the epidemic, and this would not be the case if more-educated individuals have earlier access to ARV treatments and use them. Indeed, from anecdotal evidence we expect more-educated individuals to have greater access to ARV treatments for several obvious reasons—they (i) are more likely to live in the city (e.g., in Malawi, anyone who has a university degree is likely to live in the two largest cities, Lilongwe or Blantyre, where the ARV drugs are available), (ii) have better transportation (do not have to walk several miles to refill prescriptions), or (iii) have access to someone in a hospital who can help them gain priority status when when necessary to obtain ARVs. This leaves little room for relating the initial decline in the HIV-Education gradient to ARVs.

Third, a decline in the number of HIV-positive individuals who reach adulthood with more years of schooling could also be related to declines in the HIV-Education gradient. To determine whether this is the case requires an evaluation of the effects of HIV on schooling investments of children and teenagers—that is, future entrants into adulthoood—and an analysis by education groups of the probability of entering adulthood infected with HIV. While the latter is still an open a question, regarding the former Fortson (2011) suggests a significant negative effect of HIV on investment in children in a model where agents explicitly consider mortality risk when making human capital decisions. Here we argue that these effects of HIV on education will appear at later rather than at earlier stages of the epidemic for the simple reason that it takes years to reach adulthood—our study sample. At the same time, the number of individuals reaching adulthood with respect to the total adult population is likely to be small, and so the effects on the HIV-Education gradient until the accumulation of new entrants to adulthood whose human capital choices have been endogenously affected is large enough.[42] For these reasons, we do not expect

[42] Using a large teenage population in rural Kenya between 2003 and 2010, Duflo et al. (2011) find experimental evidence that schooling subsidies and HIV information curricula are both needed to generate incentives toward safer sexual behavior—that is, this combined policy of school subsidies and HIV information implies that there is less likelihood that current teenage populations reach adulthood HIV positive and with more education. Using

changes in the education composition of the new-entrant adult population to play a major role in the initial decline of the HIV-Education gradient as the effects on the human capital of the adult population will most likely appear in the long run and beyond that last stage of the epidemic for which we have information.

Our discussion has thus far focused on the initial decline of the education gradient of HIV. However, in the most advanced stages, this gradient reverts to positive. We find that this last increase cannot be related and hence attributed to changes in sexual behavior because neither the number of extramarital partners nor the frequency of condom use show significant educational differentials at these latest stages.[43] This implies that we need to look for the reasons for the increase in the education gradient of HIV elsewhere (other than sexual behavior change across education groups). Here, by contradiction (i.e., elimination of the other factors), we argue that it must be the case that the rise of the education gradient of HIV in the most advanced stages of the epidemic is related to education differentials in survival rates enhanced by earlier access and use of ARVs by the more-educated individuals.

Note that our exercise explores dynamic relationships between the evolution of the education gradient of HIV and some potential explanations, but not causal mechanisms. To actually explore the causal mechanisms that generate the gradient requires endogenization of the three exogenous elements poised in (9): sexual behavior (i.e., infection rates, λ_s), use of ARVs (i.e., survival rates, γ_s^+), and human capital accumulation (i.e., schooling and HIV status of new adults, ϕ_s). In this context, our discussion provides guidance on how to do so. We suggest that endogenizing risky sexual behavior (in particular, the number of extramaritals partners in the last 12 months) is potentially a useful approach to deal with the initial decline in the HIV-Education gradient, but not in dealing with the increase in the gradient documented for the most advanced stages of the epidemic, which is more likely related to education differentials in ARV use.

8 Further Insights

The set of HIV infections attributed to females is larger for countries with mature epidemics (see UNAIDS (2010)). However, our exploration of HIV infection risk by education groups has

our definition of the Stages of the HIV epidemic, in 2003 Kenya was in Stage 3 but is in the most advanced stage, Stage 4, in 2010. Hence, following our results, the finding in Duflo et al. (2011) implies that their combined policy could potentially—if implemented nationally—reduce the positive HIV-Education gradient associated with Stage 4 of the Kenya's epidemic.

[43] This is also the case for age at first marriage and the formation rate of first marriages (see Appendix A). Further, we find that more-educated individuals increase the age of sexual debut in the most advanced stages of the epidemic relative to the earliest stages, which potentially leads to a decline in the HIV-Education gradient in the last stages.

focused on the male sample. Our choice is guided by the typically asymmetric scenario in SSA in which women lack power in the choice of sexual practices inside and outside marriage.[44] Here, we explicitly study the education differentials of the probability of HIV infection and the number of extramarital partners separately for women. Further, marriage has been hypothesized as a potentially useful strategy to reduce risk of HIV infection. In this context, it is important to conduct a sensitivity exercise on marital status. Finally, there has been a recent large increase in the aggregate ARV treatment coverage in several SSA countries.[45] Since treatment mitigates the consequences of HIV infection—mainly, its development to AIDS and death—it is natural to think that the HIV-Education gradient is potentially related to educational disparities in the use of ARVs, hence, survival rates to the disease across education groups. There are two further offsetting mechanisms by which education differentials in ARV use can influence the HIV-Education gradient by changing the rate of new infections due to (i) increases in the sexual activity among individuals infected with HIV who take ARVs and (ii) reductions in the HIV infectiousness of HIV-positive individuals who take ARVs. Unfortunately, individual data on the use of ARVs are not available from DHS, and hence we cannot directly control for educational disparities on ARV treatment. We do, however, control for country-specific aggregate measures of ARV coverage (from the WPP, 2009) at the period of DHS data collection.

8.1 The Female Sample

As is the case for males, the HIV-Education gradient for females is significantly nonstationary over the stages of the HIV epidemic (see column 1 in Table 6).[46] It displays a U-shaped pattern similar to the counterpart for men but is more sizable. Precisely, at Stage 0, an additional year of schooling is associated with a 2.3% rise in the female probability of HIV infection—that is, an increase more than two times as great than for males per year of schooling, 1.1%. Thereafter, the female HIV-Education gradient substantially declines reaching 2.3-1.8=.5% in Stage 2. The gradient stays that low for Stages 2 and 3 (indeed, the coefficients are not significantly different among Stages 1, 2 and 3). In Stage 4, the HIV-Education gradient for females significantly increases (as it did for males) with a 1.7% more chances of HIV infection per additional schooling

[44] For example, using 1,799 subjects in Lusaka (Zambia), Ashraf et al. (2010) provide experimental evidence on the fact that the arrival of information about contraceptive methods increases the use of these methods (i.e., aimed at decreasing unwanted births) only when the household bargaining power of women is large. In this context, it is natural to think that there is a relationship between the bargaining power of women within marriage and the ability of women to protect themselves against HIV and other sexually transmitted diseases.

[45] While rising, the ARV treatment coverage (as a percentage of total population infected) is rather heterogenous across SSA countries ranging from less than 5% in Malawi, Zimbabwe, Lesotho, Rwanda, and Ghana, to more than 25% in Zambia, Swaziland and Mali at the time of DHS data collection.

[46] As we did for males, we restrict our attention to HIV-tested adults women 15-49 years old who reported their schooling achievement and are sexually active. Our sample consists of a total of 74,528 observations.

year. Regarding the actual sexual behavior of females, the Partners-Education gradient is smaller (about half) than that of males but it shows a similar decline over stages of the epidemic (see column 2 in Table 6). One potential caveat of this analysis is the possibility of underreporting of risky sexual behavior for women pointed out in Smith (1992) and Gersovitz et al. (1998).[47] However, as long as occurs systematically across all stages of the epidemic, the shape of the Partners-Education gradient for women over stages is robust to this problem.

8.2 Marital Status

Here, we explicitly incorporate marital status in our econometric analysis with a set of dummy variables on married and formerly married individuals where never-married individuals form our reference group. This approach allows us to explore whether marital status reshapes the relationship between education and the probability of HIV infection documented in Subsection 5.2. First, we find that the HIV-Education gradient follows the same significant and nonstationary (positive-zero-positive) U-shape pattern as our benchmark specification (see column 3 in Table 6). Further, the behavior of the Partners-Education gradient is almost identical to our previous results in Subsection 6.2 (see column 4 in Table 6). That is, conditioning on marital status does not alter our main results. Second, we find that being married is associated with an increase in the probability of infection by 2.3% (a nonnegligible number given that the average HIV prevalence for our DHS sample is 3.8%; see Table 1), and that being formerly married (i.e., currenlty widowed, divorced, or remarried) is associated with about a 6.9% increase in the probability of infection.[48]

8.3 Antiretroviral Treatment

Using longitudinal data from Kosirai Division in western Kenya, Goldstein et al. (2010) estimate increases in sexual activity without significant increases in the use of condoms soon after the beginning of individual ARV treatment. In this context, more-educated individuals, who are more likely to receive ARV treatment, have larger incentives to continue risky sexual behavior rather than take precautionary measures. Further, in a study of HIV-positive Ugandans using ARVs over 3 years, Apondi et al. (2011) suggest that, despite increases in sexual activity, the estimated risk of HIV transmission can remain lower than baseline levels due to reductions in the levels of infectiousness. While individual data on the use of ARVs are not available, we can control for country-specific aggregate measures of ARV coverage. When we do so, first, we find results in the

[47] More recently, using the sexual network data from Likoma Island, Malawi, Helleringer et al. (2009) find that men and women are equally likely to underreport risky sexual behavior.

[48] These results for the relationship between individual HIV status and marital status assume stationarity between these two variables. Alternatively, exploring changes over time in this same relationship, Reniers (2008) reports an increase in union-based risk-avoidance strategies in rural Malawi during the period HIV.

sign and size of the nonstationary HIV-Education and Partners-Education gradients (see columns 5 and 6 in Table 6) extraordinarily similar to those in our benchmark results in Subsection 5.2. Second, not surprisingly, we find that aggregate ART coverage reduces the probability of individual HIV infection.

9 Conclusion

The mixed evidence in the literature investigating the relationship between education and the probability of being HIV-positive in SSA suggests that finding which type of individuals are at greater risk of HIV infection is not an easy task. We proposed a fresh look to this question that consists of explicitly introducing the stages of the HIV epidemic into the analysis. Using nationally representative data for 18 SSA countries to exploit variation across stages of the HIV epidemic, we showed that the relationship between completed educational attainment and individual HIV status (i.e., the HIV-Education gradient) is dynamic and significantly evolves with the epidemic. At early stages of the epidemic more-educated individuals are more likely to be infected; however, this relationship strongly decreases as the epidemic evolves and eventually reaches a stage where education and the probability of being HIV-positive are no longer significantly correlated. Interestingly, in the most advanced stages of the epidemic, the education gradient of HIV returns to being high and positive. In light of our findings, we call for frameworks of HIV policy evaluation that incorporate the dynamic relationship between education and HIV status that we documented.

Further, we showed that the diffusion of knowledge on HIV prevention arrives earlier to the more-educated individuals and that the initial decline in the education gradient of HIV is strongly related to an earlier adoption of safer sexual practices by the more-educated individuals implemented mostly through reductions in the number of extramarital partners in the past year. Other margins of risky sexual behavior (e.g., condom use, age at first marriage, formation rate of first marriages, and age at sexual debut) do not display significant changes with the epidemic across education groups. However, in the most advanced stages of the epidemic, HIV knowledge, as well as actual risky sexual behavior, is quite uniform across education groups. This lack of educational disparities in knowledge and sexual behavior in the most advanced stages suggests that the increase in the education gradient of HIV documented for those stages represents indirect evidence of higher survival rates of the more-educated individuals, which are likely enhanced by their earlier use of ARV treatments. Therefore, we argue that a useful framework to understand the dynamic relationship between education and the probability of being infected with HIV requires the endogeneization of both (i) prevention through sexual behavior choices (in particular, in the number of extramarital partners) and (ii) access and use of ARV treatment.

References

Aizer, A. and Stroud, L. (2010). Education, knowledge and the evolution of disparities in health. NBER Working Paper, 15840.

Apondi, R., Bunnell, R., Ekwaru, J. P., Moore, D., Bechange, S., Khana, K., King, R., Campbell, J., Tappero, J., and Mermin, J. (2011). Sexual behavior and hiv transmission risk of ugandan adults taking antiretroviral therapy: 3 year follow-up. *AIDS*, 10:1317–1327.

Ashraf, N., Field, E., and Lee, J. (2010). Household bargaining and excess fertility: An experimental study in zambia. Mimeo, Harvard University and Duke University.

Asiimwe-Okiror, G., Opio, A. A., Musinguzi, J., Madraa, E., Tembo, G., and Caraël, M. (1997). Change in sexual behaviour and decline in hiv infection among young pregnant women in urban uganda. *AIDS*, 11(14):1757–63.

Bärnighausen, T., Hosegood, V., Timaeus, I. M., and Newell, M.-L. (2007). The socioeconomic determinants of hiv incidence: Evidence from a longitudinal, population-based study in rural south africa. *AIDS*, 21(7):S29–S38.

Beegle, K. and de Walque, D. (2009). Demographic and socioeconomic patterns of hiv/aids prevalence in africa. World Bank Policy Research Working Paper, 5076.

Bongaarts, J., Buettner, T., Heilig, G., and Pelletier, F. (2008). Has the hiv epidemic peaked? *Population and Development Review*, 34(2):199–224.

Case, A., Lubotsky, D., and Paxon, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92(5):1308–34.

Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The determinants of mortality. *Journal of Economic Perspectives*, 20(3):97–120.

de Walque, D. (2004). Education, information, and smoking decisions: evidence from smoking histories, 1940-2000. World Bank Policy Research Working Paper, 3362.

de Walque, D. (2007). How does the impact of an hiv/aids information campaign vary with educational attainment? evidence from rural uganda. *Journal of Development Economics*, 84(2):686–714.

de Walque, D. (2009). Does education affect hiv status? evidence from five african countries. *The World Bank Economic Review*, 23(2):209–233.

de Walque, D., Nakiyingi-Miiro, J. S., Busingye, J., and Whitworth, J. A. (2005). Changing association between schooling levels and hiv-1 infection over 11 years in rural populations cohort in south-west uganda. *Tropical Medicine and International Health*, 10(10):993–1001.

Dinkelman, T., Levinsohn, J. A., and Majelantle, R. (2006). When knowledge is not enough: Hiv/aids information and risky behavior in botswana. NBER Working Paper, 12418.

Duflo, E., Dupas, P., and Kremer, M. (2011). Education, hiv and early fertility: Experimental evidence from kenya. Mimeo MIT, Stanford University, and Harvard University.

Dupas, P. (2011). Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *American Economic Journal: Applied Economics*, 3(1):1–36.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology*, 35:553–572.

Elo, I. T. and Preston, S. H. (1996). Educational differentials in morality in the united states 1979-1985. *Social Science Medicine*, 42:47–57.

Feldman, J. J., Makuc, D. M., Kleinman, J. C., and Cornoni-Huntley, J. (1989). National trends in educational differences in mortality. *Americal Journal of Epidemiology*, 10:214–23.

Fortson, J. G. (2008). The gradient in sub-saharan africa: Socioeconomic status and hiv/aids,. *Demography*, 45(2):303–322.

Fortson, J. G. (2011). Mortality risk and human capital investment: The impact of hiv/aids in sub-saharan africa. *Review of Economics and Statistics*, 93(1):1–15.

Gersovitz, M., Jacoby, H. G., Dedy, F. S., and Tap, A. G. (1998). The balance of self-reported heterosexual activity in kap surveys and the aids epidemic in africa. *Journal of the American Statistical Association*, 93:875–883.

Glynn, J. R., Caraël, M., Buvé, A., Anagonou, S., Zenkeng, L., Kahindo, M., and Musonada, R. (2002). Does increased general schooling protect against hiv infection? a study in four african cities. *Tropical Medicine and International Health*, 9:4–14.

Goldman, D. and Smith, J. P. (2005). Socioeconomic differences in the adoption of medical technology. *American Economic Review*, 95:234–37.

Goldstein, M., Zivin, J. G., Habyarimana, J., Pop-Eleches, C., and Thirumurthy, H. (2010). Aids treatment programs and sexual behavior. Mimeo, Columbia University.

Gollin, D., Parente, S. L., and Rogerson, R. (2002). The role of agriculture in development. *American Economic Review*, 92(2):160–164.

Gollin, D., Parente, S. L., and Rogerson, R. (2007). The food problem and the evolution of international income levels. *Journal of Monetary Economics*, 54(4):1230–850.

Hansen, G. D. and Prescott, E. C. (2002). Malthus to solow. *American Economic Review*, 92(4):1205–1217.

Hargreaves, J. R., Bonell, C. P., Morison, L. A., Kim, J. C., Phetla, G., Porter, J. D., Watts, C., and Pronyk, P. M. (2007). Explaining continued high hiv prevalence in south africa: Socioeconomic factors, hiv incidence and sexual behavior change among a rural cohort, 2001-2004. *AIDS*, 21(7):S39–S48.

Hargreaves, J. R. and Glynn, J. R. (2002). Education attainment and hiv-1 infection in developing countries: A systematic review. *Tropical Medicine and International Health*, 7(6):489–498.

Helleringer, S., Kohler, H.-P., Chimbiri, A., Chatonda, P., and Mkandawire, J. (2009). The likoma network study: Context, data collection, and initial results. *Demography Research*, 21:427–468.

Herrendorf, B., Rogerson, R., and Valentinyi, A. (2009). Two perspectives on preferences and structural transformation. NBER Working Paper, 15416.

Kates, J., Wexler, A., Lief, E., Avila, C., and Gobet, B. (2011). Financing the response to aids in low- and middle- income countries: International assistance from donor governments in 2010. Joint United Nations Programme on HIV/AIDS and The Henry J. Kaiser Foundation.

Kohler, I. V., Martikainen, P., Smith, K. P., and Elo, I. T. (2008). Education differences in all-cause mortality by marital status—evidence from bulgaria, finland and the united states. *Demography Research*, 19:2011–42.

Link, B. G., Northridge, M. E., Phelan, J. C., and Ganz, M. L. (1998). Social epidemiology and the fundamental cause concept: On the structuring of effective cancer screens by socioeconomic status. *Milbank Quarterly*, 76(3):375–402.

Link, B. G. and Phelan, J. C. (1995). Social conditions as the fundamental causes of disease. *Journal of Health and Social Behavior*, 35:80–94. Extra Issue: Forty Years of Medical Sociology: The State of the Art and Directions for the Future.

Lleras-Muney, A. (2004). The relationship between education and adult mortality in the united states. *Review of Economic Studies*, 72(1):189–221.

Lopman, B., Lewis, J., Nyamukapa, C., Mushati, P., Chandiwana, S., and Gregson, S. (2007). Hiv incidence and poverty in manicaland, zimbabwe: Is hiv becoming the disease of the poor? *AIDS*, 21(7):S57–S66.

Mishra, V. and Bignami-Van-Assche, S. (2009). Concurrent sexual partnerships and hiv infection: Evidence from national population-based surveys. DHS Working Papers, 62.

Mishra, V., Bignami-Van-Assche, S., Greener, R., Vaessen, M., Hong, R., Ghys, P. D., Boerma, J. T., Van-Assche, A., Shane-Khan, and Rutstein, S. (2007). Hiv infection does not disproportionately affect the poorer in sub-saharan africa. *AIDS*, 21(7):S17–S29.

Preston, S. H. and Elo, I. T. (1995). Are educational differentials in adult mortality increasing in the u.s.? *Journal of Aging Health*, 7(4):476–496.

Reniers, G. (2008). Marital strategies for regulating exposure to hiv. *Demography*, 45(2):417–438.

Santaeulàlia-Llopis, R. (2011). Aggregate effects of aids on development. Mimeo, Washington University in St. Louis.

Smith, T. W. (1992). A methodological analysis of the sexual behavior questions on the general social surveys. *Journal of Official Statistics*, 8:309–325.

Stoneburner, R. L. and Low-Beer, D. (2004). Population-level hiv declines and behavioral risk avoidance in uganda. *Science*, 304:714–8.

Strauss, J. and Thomas, D. (2007). Health over the life course. In Schultz, T. P. and Strauss, J., editors, *Handbook of Development Economics*, volume 4, chapter 54, pages 3375–3474. Amsterdam, North-Holland.

UNAIDS (1997). A measure of success in uganda: the value of monitoring both hiv prevalence and sexual behavior. Geneva: UNAIDS.

UNAIDS (2005). Evidence for hiv decline in zimbabwe: A comprehensive review of the epidemiological data. Geneva: UNAIDS.

UNAIDS (2010). Report on the global aids epidemic. Joint United Nations Programme on HIV/AIDS, New York.

Wojcicki, J. M. (2005). Socioeconomic status as a risk factor for hiv infection in women in east, central and southern africa: A systematic review. *Journal of Biosocial Science*, 37(1):1–36.

Table 1: DHS Sample Characteristics (across Countries): Males (15-49 years)

	Mean	Median	Min.	Max.	Gini
HIV Prevalence (%)	4.5	2.3	1.0	27.8	0.44
Years of Schooling	3.6	3.6	1.3	6.3	0.22
Age	31.0	31.0	28.6	33.3	0.03
Urban (%)	31	27	13	59	0.23
Extramarital Partners, n	0.47	0.48	0.08	1.17	0.30
$n = 0$ (%)	68	66	42	93	0.12
$n = 1$ (%)	24	27	7	40	0.20
$n = 2$ (%)	5	5	1	13	0.34
$n \geq 3$ (%)	2	2	0	10	0.46
Frequency of Condom Use (%)	16	15	4	48	0.28
$n = 0$ (%)	5	3	1	28	0.34
$n > 0$ (%)	38	40	20	62	0.13
$n > 1$ (%)	42	44	27	70	0.12
Noncasual Sexual Partner (%)	5	3	1	28	0.35
Casual Sexual Partner (%)	45	46	19	78	0.17
Stability I:					
Age at First Marriage					
Sample: Newlyweds (1st Marriage) in current Stage j	23.9	23.6	22.7	26.3	0.01
Sample: All	23.1	22.9	22.0	25.2	0.01
Stability II:					
Newlyweds (1st Marriage) in current Stage j (%)					
Sample: Eligibles (1st Marriage) in current Stage j (%)	65	54	33	91	0.16
Sample: All (%)	67	66	41	85	0.10
Age at Sexual Debut					
Sample: Sexual Debutants in current Stage j	18.3	17.6	16.2	20.3	0.04
Sample: All	18.1	17.6	16.1	19.9	0.04

Notes: The computation of these statistics is performed by first using individual HIV weights provided by the DHS to compute averages for each country, and then country-specific population weights (i.e., the population size of each country provided by WPP, 2009) to compute the statistics across countries.

Table 2: The Evolution of the HIV Epidemic across Sub-Saharan Countries: The DHS Sample

									Speed	
	U.N. Peak		DHS year						To Peak	From Peak
Country	HIV^i_*	t^i_*	HIV^i_{DHS}	t^i_{DHS}	$\frac{HIV^i_{DHS}}{HIV^i_*}$	$t^i_{DHS} - t^i_*$	$t^i_{-.5}$	$t^i_{.5}$	$.5/(t^i_* - t^i_{-.5})$	$.5/(t^i_{.5} - t^i_*)$
Burkina Faso	2.1	1991	1.6	2003	.76	12	1986	2045	.10	.01
Cameroon	6.2	1999	5.8	2004	.94	5	1993	2040	.08	.01
Cote d'Ivorie	6.3	1997	5.2	2005	.83	8	1991	2023	.08	.02
D. R. of Congo	1.6	1988	1.3	2007	.81	19	1985	2047	.17	.01
Ethiopia	2.6	1996	2.1	2005	.81	9	1992	2043	.13	.01
Ghana	2.5	1997	2.1	2003	.84	6	1994	2044	.17	.01
Guinea	1.7	2010	1.6	2005	.94	-5	1998	2037	.04	.02
Kenya	10.3	1996	8.6	2003	.83	7	1991	2040	.10	.01
Lesotho	23.8	2000	23.4	2004	.98	4	1995	-	.10	-
Liberia	1.8	2002	1.7	2007	.94	5	1994	2047	.06	.01
Malawi	13.8	1998	12.3	2004	.89	6	1994	2050	.13	.01
Mali	1.6	2007	1.6	2006	1.00	-1	1995	2048	.04	.01
Niger	.8	2006	.8	2006	1.00	10	1996	-	.05	-
Senegal	-	-	0.7	2005	-	-	-	-	-	-
Rwanda	7.1	1993	3.3	2005	.46	12	1987	2003	.08	.05
Swaziland	26.8	2002	26.8	2006	1.00	4	1995	-	.07	-
U. R. of Tanzania	7.5	1996	6.5	2003	.87	7	1989	2045	.07	.01
U. R. of Tanzania	*idem*	*idem*	6.3	2007	.84	9	*idem*	*idem*	*idem*	*idem*
Zambia	16.4	1995	15.1	2006	.92	11	1991	2048	.13	.01
Zimbabwe	29.1	1997	19.2	2005	.66	8	1991	2009	.08	.04

Notes: HIV^i_* is the peak prevalence rate for country i; t^i_* is the year country i reaches its HIV prevalence peak; HIV^i_{DHS} is the prevalence rate for country i the year of DHS data collection; t^i_{DHS} is the year of DHS data collection for country i; $t^i_{-.5}$ is the year country i reaches half of its peak on the rising side of the epidemic; and $t^i_{.5}$ is the year country i reaches half of its peak on the declining side of the epidemic. Sources: United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects: The 2008 Revision, Medium-Variant Estimation and Projection.

Table 3: The HIV-Education Gradient

HIV Status	Stationary	Nonstationary		
	(1)	(2)	(3)	(4)
Education	0.004***	0.011***	0.014***	0.015***
	(0.000)	(0.001)	(0.001)	(0.001)
Education * Stage 1		-0.006***	-0.006***	-0.014***
		(0.002)	(0.001)	(0.002)
Education * Stage 2		-0.007***	-0.013***	-0.011***
		(0.002)	(0.001)	(0.002)
Education * Stage 3		-0.010***	-0.015***	-0.015***
		(0.002)	(0.001)	(0.002)
Education * Stage 4		-0.001	0.003*	-0.002
		(0.002)	(0.002)	(0.002)
Age	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Urban Area	0.024***	0.018***	0.023***	0.022***
	(0.003)	(0.003)	(0.003)	(0.003)
Stage 1		0.036***		0.047***
		(0.006)		(0.006)
Stage 2		0.015***		-0.011***
		(0.004)		(0.004)
Stage 3		-0.002		-0.001
		(0.006)		(0.006)
Stage 4		0.019***		0.024***
		(0.005)		(0.005)
Agricultural Share		-0.003***		
		(0.000)		
Output per Capita		-0.000***		
		(0.000)		
Constant	-0.032***	0.073***	-0.031***	-0.039***
	(0.005)	(0.016)	(0.004)	(0.005)
Sample Size	53,588	53,588	53,588	53,588
R^2	0.01	0.04	0.02	0.03

Notes: Sexually active men with fewer than 20 sex partners (other than spouses) in the past 12 months. Robust SE in parentheses. *Significant at 10%; ** significant at 5%; and *** significant at 1%.

Table 4: The Knowledge-Education Gradients

Knowledge on HIV Prevention Strategies	Partners Reduction	Condom Use		
	All	n>0	n>1	Casual
	(1)	(2)	(3)	(4)
Education	0.004	-0.004	-0.010	0.002
	(0.003)	(0.005)	(0.011)	(0.013)
Education * Stage 1	0.005	0.011*	0.010	0.010
	(0.003)	(0.006)	(0.012)	(0.014)
Education * Stage 2	0.011***	0.027***	0.017	0.031**
	(0.004)	(0.007)	(0.014)	(0.016)
Education * Stage 3	0.008**	0.019***	0.023*	0.005
	(0.004)	(0.007)	(0.014)	(0.017)
Education * Stage 4	-0.010***	0.004	0.011	-0.007
	(0.004)	(0.006)	(0.013)	(0.016)
Age	0.001***	0.000	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.002)
Urban Area	0.012***	0.032***	0.015	0.041
	(0.005)	(0.010)	(0.017)	(0.031)
Stage 1	0.042***	-0.035	-0.007	-0.090*
	(0.014)	(0.022)	(0.052)	(0.054)
Stage 2	0.096***	-0.100***	-0.001	-0.204***
	(0.013)	(0.030)	(0.064)	(0.068)
Stage 3	0.001***	-0.076***	-0.042	-0.107
	(0.000)	(0.026)	(0.057)	(0.069)
Stage 4	0.000***	0.031	0.056	0.069
	(0.000)	(0.026)	(0.053)	(0.057)
Agricultural Share	0.040***	0.002***	0.002**	0.003
	(0.013)	(0.001)	(0.001)	(0.002)
Output per Capita	0.019	0.000***	0.000***	0.000**
	(0.013)	(0.000)	(0.000)	(0.000)
Constant	0.678***	0.662***	0.660***	0.638***
	(0.023)	(0.042)	(0.081)	(0.150)
Sample Size	52,228	19,406	4,711	2,548
R^2	0.01	0.02	0.02	0.04

Notes: Sexually active men with fewer than 20 sex partners (other than spouses) in the past 12 months. Robust SE in parentheses. *Significant at 10%; ** significant at 5%; and *** significant at 1%.

Table 5: The Risky Sex-Education Gradient

	Extramarital Partners	Condom Use		
	All	n>0	n>1	Casual
	(1)	(2)	(3)	(4)
Education	0.057***	0.021**	0.017	-0.001
	(0.009)	(0.009)	(0.018)	(0.023)
Education * Stage 1	-0.048***	0.003	0.003	0.034
	(0.010)	(0.009)	(0.019)	(0.024)
Education * Stage 2	-0.030***	-0.008	-0.007	0.016
	(0.010)	(0.010)	(0.021)	(0.025)
Education * Stage 3	-0.080***	-0.011	-0.000	0.015
	(0.012)	(0.010)	(0.020)	(0.025)
Education * Stage 4	-0.053***	-0.016	-0.007	-0.002
	(0.010)	(0.011)	(0.023)	(0.028)
Age	-0.028***	-0.010***	-0.007***	0.004**
	(0.001)	(0.001)	(0.001)	(0.002)
Urban Area	0.180***	0.201***	0.221***	0.284***
	(0.016)	(0.011)	(0.023)	(0.031)
Stage 1	0.262***	0.059*	0.098	0.020
	(0.028)	(0.031)	(0.066)	(0.076)
Stage 2	-0.021	0.179***	0.209***	0.117
	(0.024)	(0.035)	(0.079)	(0.081)
Stage 3	0.328***	0.071**	0.115	0.065
	(0.042)	(0.033)	(0.070)	(0.084)
Stage 4	-0.002	0.323***	0.381***	0.375***
	(0.024)	(0.037)	(0.084)	(0.092)
Agricultural Share	-0.005***	0.002**	0.002	0.006**
	(0.001)	(0.001)	(0.002)	(0.003)
Output per Capita	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	1.252***	0.178***	0.089	-0.378**
	(0.069)	(0.061)	(0.133)	(0.173)
Sample Size	53,588	18,872	4,498	2,570
R^2	0.10	0.11	0.10	0.17

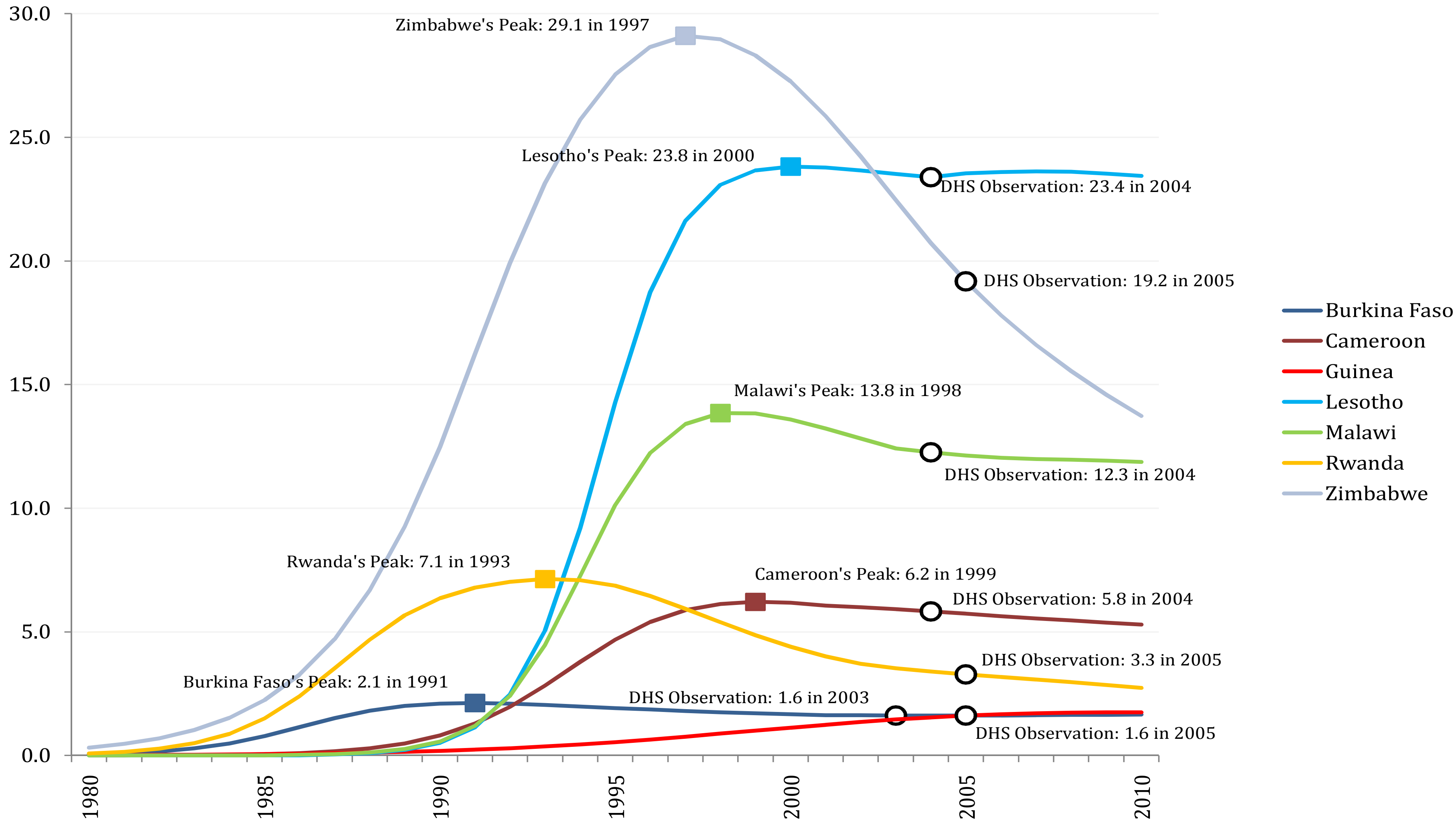
Notes: Sexually active men with fewer than 20 sex partners (other than spouses) in the past 12 months. Robust SE in parentheses. *Significant at 10%; ** significant at 5%; and *** significant at 1%.

Table 6: Further Insights on the HIV-Education and Partners-Education Gradients

	Females		ARVs Covarage		Marital Status	
	HIV Status	Partners	HIV	Partners	HIV	Partners
	(1)	(2)	(3)	(4)	(5)	(6)
Education	0.023***	0.025***	0.010***	0.058***	0.011***	0.048***
	(0.002)	(0.003)	(0.001)	(0.009)	(0.001)	(0.009)
Education * Stage 1	-0.018***	-0.016***	-0.007***	-0.048***	-0.007***	-0.044***
	(0.002)	(0.004)	(0.002)	(0.010)	(0.002)	(0.010)
Education * Stage 2	-0.017***	-0.008**	-0.006***	-0.031***	-0.007***	-0.030***
	(0.002)	(0.004)	(0.002)	(0.010)	(0.002)	(0.009)
Education * Stage 3	-0.016***	-0.035***	-0.010***	-0.081***	-0.010***	-0.068***
	(0.002)	(0.004)	(0.002)	(0.013)	(0.002)	(0.012)
Education * Stage 4	-0.006***	-0.022***	0.001	-0.055***	-0.001	-0.050***
	(0.002)	(0.004)	(0.002)	(0.011)	(0.002)	(0.010)
Age	0.001***	-0.009***	0.002***	-0.028***	0.001***	-0.008***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Urban Area	0.034***	0.105***	0.019***	0.178***	0.019***	0.101***
	(0.003)	(0.006)	(0.003)	(0.016)	(0.003)	(0.015)
Stage 1	0.045***	0.103***	0.015**	0.294***	0.036***	0.216***
	(0.005)	(0.015)	(0.007)	(0.032)	(0.006)	(0.027)
Stage 2	0.023***	0.024*	-0.010	0.019	0.016***	-0.044**
	(0.004)	(0.013)	(0.007)	(0.035)	(0.004)	(0.022)
Stage 3	-0.012***	0.194***	-0.022***	0.358***	0.001	0.211***
	(0.004)	(0.017)	(0.007)	(0.048)	(0.006)	(0.039)
Stage 4	0.020***	0.018	-0.007	0.037	0.022***	-0.051**
	(0.004)	(0.014)	(0.007)	(0.031)	(0.005)	(0.023)
Agricultural Share	-0.003***	0.001**	-0.003***	-0.004***	-0.003***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Output per Capita	0.000	0.000***	-0.000***	0.000***	-0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Antiretroviral Coverage			-0.001***	0.001*		
			(0.000)	(0.001)		
Married					0.023***	-0.723***
					(0.003)	(0.022)
Formerly Married					0.067***	-0.002
					(0.007)	(0.048)
Constant	0.085***	0.212***	0.124***	1.174***	0.072***	1.196***
	(0.014)	(0.028)	(0.014)	(0.055)	(0.016)	(0.066)
Sample Size	74,525	74,525	53,588	53,588	53,585	53,585
R^2	0.05	0.08	0.04	0.10	0.04	0.17

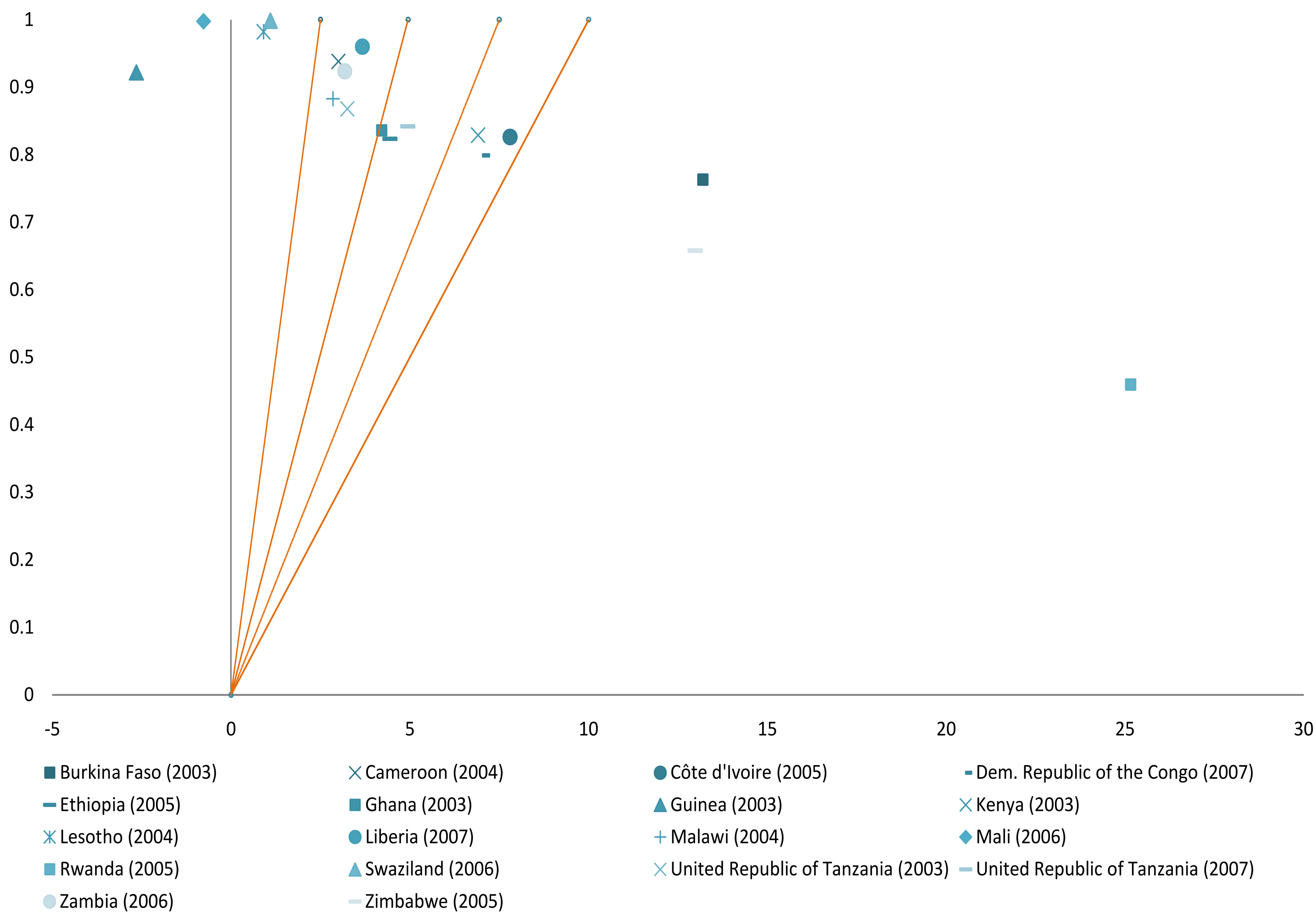
Notes: Sexually active men with fewer than 20 sex partners (other than spouses) in the past 12 months. Robust SE in parentheses. *Significant at 10%; ** significant at 5%; and *** significant at 1%.

Figure 1: Challenges for the Definition of Stages of the HIV: Epidemic. The Evolution of the HIV Epidemic for a DHS Subsample.



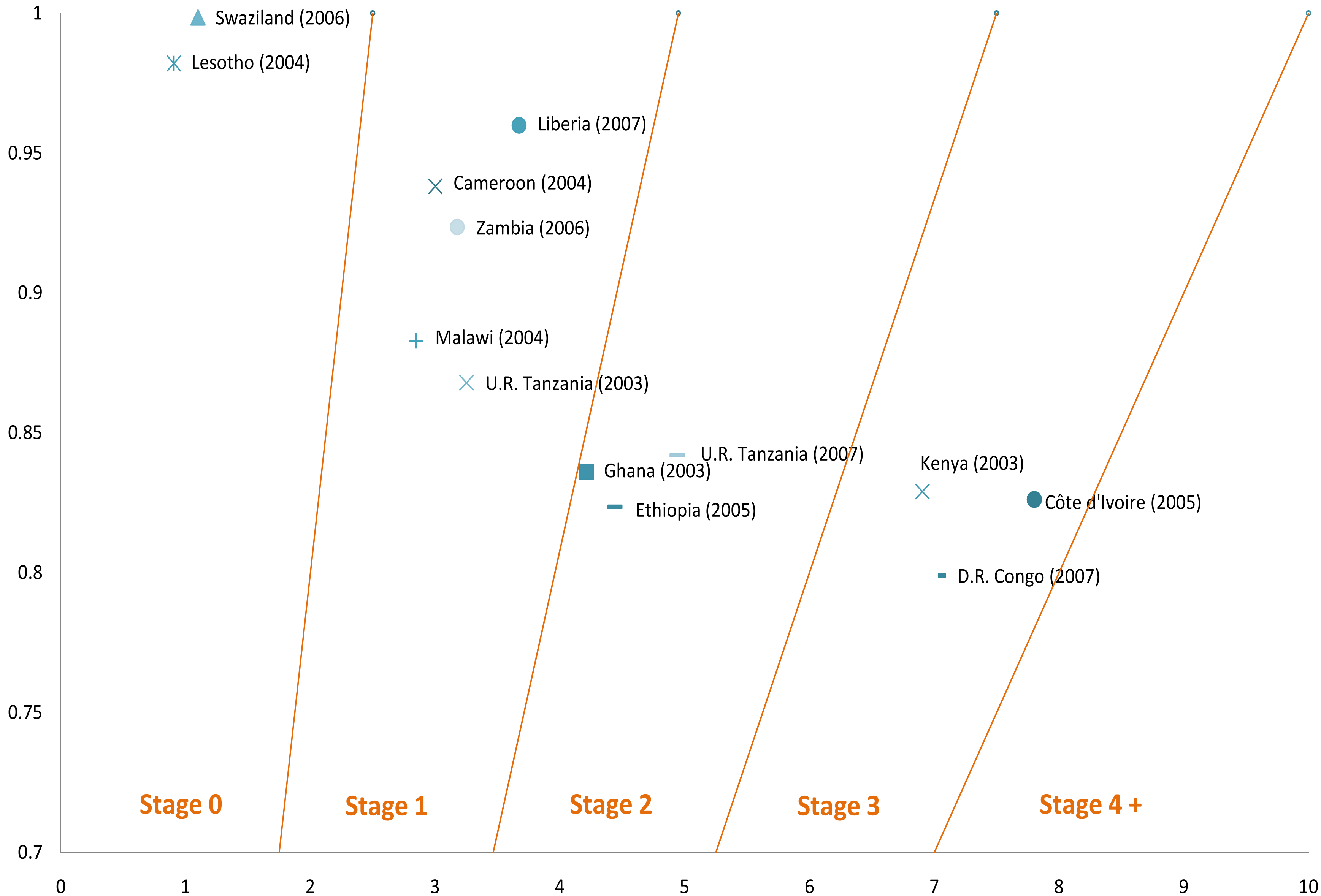
Source: United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects: The 2008 Revision, Medium-Variant Estimation. *Notes*: The solid square on each HIV time path displays the HIV prevalence at the peak year, and the open black circle on each HIV time path displays the HIV prevalence at the year that the DHS data were collected.

Figure 2: **Stages of the HIV Epidemic and Location of SSA Countries (DHS Sample) on the 2D-Normalized Space at the Time of DHS Data Collection.** Normalized HIV-Prevalence $\in [0,1]$ (vertical axes) and normalized time τ (horizontal axes).



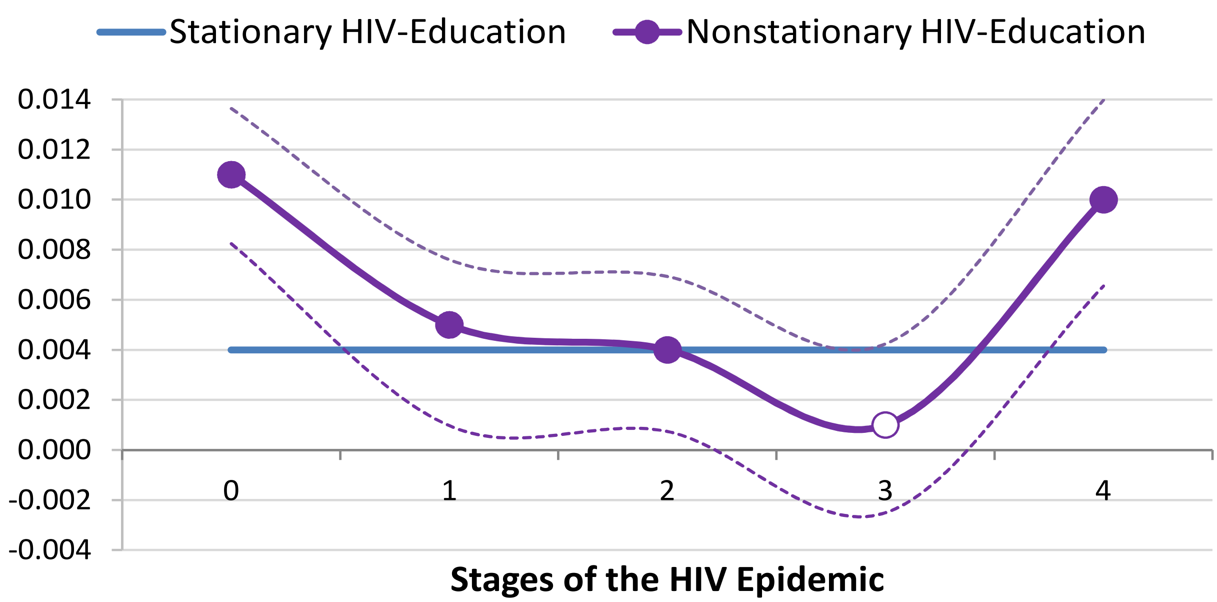
Source: Outcome of our 2D-normalization algorithm (Subsection 4.2) implemented using WPP, 2009, data. *Notes*: Each point in the scatterplot represents a DHS dataset.

Figure 3: **Stages of the HIV Epidemic and Location of SSA Countries (DHS Sample) on the 2D Normalized Space at the Time of DHS Data Collection.** Normalized HIV Prevalence $\in [.7, 1]$ (vertical axes) and normalized time τ (horizontal axes).



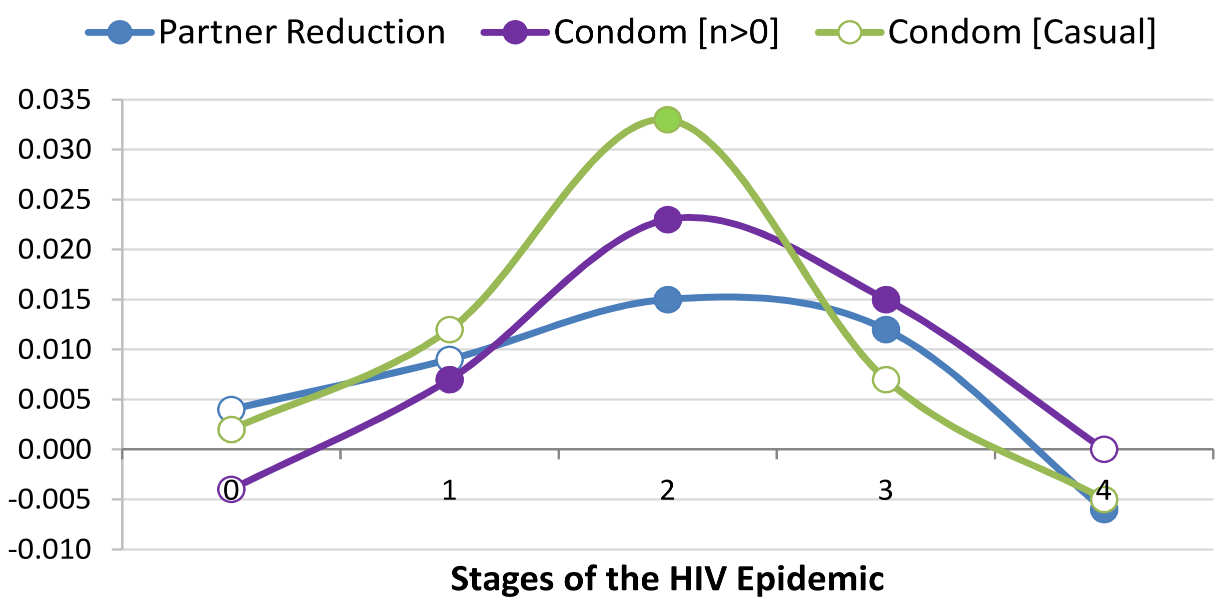
Source: Outcome of our 2D normalization algorithm (Subection 4.2) implemented using WPP, 2009, data. *Notes*: Each point in the scatterplot represents a DHS dataset.

Figure 4: The Dynamics of Education and Probability of Infection: The HIV-Education Gradient



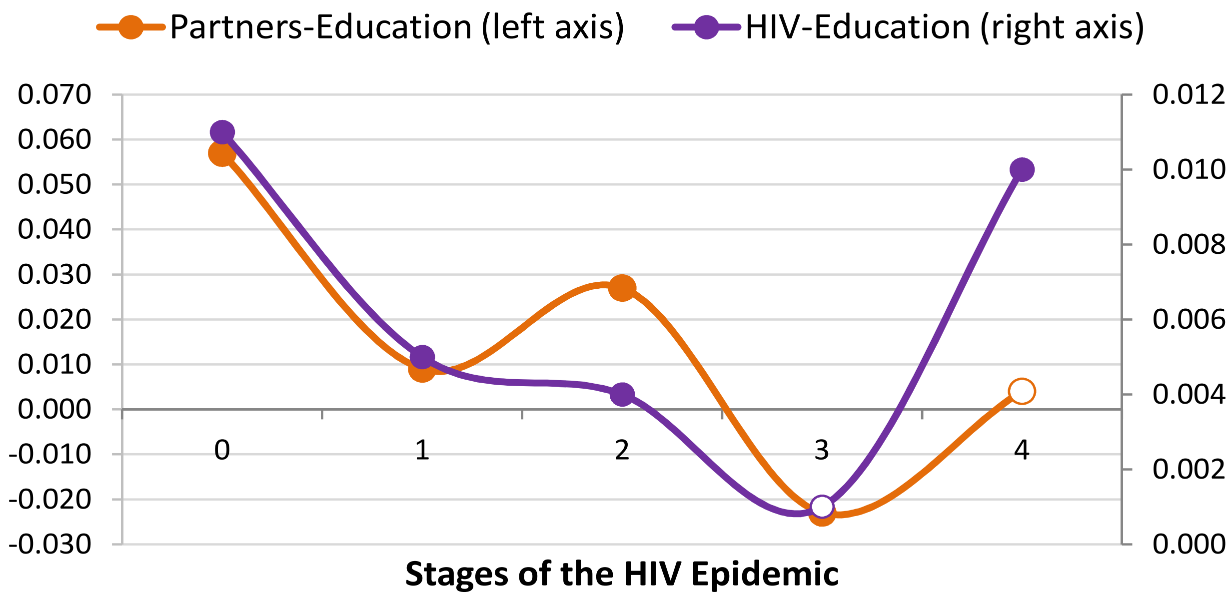
Source: Estimates of $\left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j\right)$ for each Stage j from columns 1 and 2 in Table 3.
Notes: Significance at 1%, 5%, and 10% is represented by, respectively, markers with solid fill, markers with medium transparency fill, and markers with light transparency fill. Markers with open circles are not significant at the 10% level. Dashed lines represent 95% confidence intervals.

Figure 5: The Diffusion Patterns of HIV Knowledge: The Knowledge-Education Gradient



Source: Estimates of $\left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j\right)$ for each Stage j from columns 1, 2 and 4 in Table 4.
Notes: Significance at 1%, 5%, and 10% is represented by, respectively, markers with solid fill, markers with medium transparency fill, and markers with light transparency fill. Markers with open circles are not significant at the 10% level.

Figure 6: The Partners-Education Gradient and the HIV-Education Gradient



Source: Estimates of $\left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j\right)$ for each Stage j from column 2 in Table 3 and column 1 in Table 5 .
Notes: Significance at 1%, 5%, and 10% is represented by, respectively, markers with solid fill, markers with medium transparency fill, and markers with light transparency fill. Markers with open circles are not significant at the 10% level.

A Cross-Country Heterogeneity on the Evolution of the HIV Epidemic

This section discusses in detail evidence of the large cross-country heterogeneity on the evolution of the HIV epidemic. Our discussion here extends the analysis of Subsection 4.1 and builds on Figure 1 and on the statistics reported in Table 2 for several SSA countries with available DHS datasets.

1. **The peak of HIV prevalence substantially differs across SSA countries.** As reported in the first column in Table 2, the highest HIV peak was achieved by Zimbabwe (29.1%), followed by Swaziland (26.8%) and Lesotho (23.8%). About half way of the largest peaks we have Zambia (16.4%), Malawi (13.8%) and Kenya (10.3%). Above 5% we still have the countries of Cameroon (6.2%), Cote d'Ivoire (6.3%) and Tanzania (7.5%). There are 5 countries below 5% of HIV prevalence: Burkina Faso (2.1%), Democratic Republic of Congo (1.6%), Guinea (1.7%), Liberia (1.8%) and Mali (1.6%).

2. **The HIV prevalence levels largely differ across SSA countries at the time of DHS collection.** See the third column in Table 2 and Figure 1. For example, in 2006 data were collected for Mali, Swaziland, and Zambia, which at that time had very different HIV prevalences, respectively, of 1.6%, 26.8%, and 15.1%. A similar phenomenon occurs in all other years, for example, in 2005, Guinea had a prevalence of 1.6%, Rwanda 3.3%, Cote d'Ivoire 5.2% and Zimbabwe 19.2%.

3. **There exists a large degree of heterogeneity on the relative size of the HIV prevalence across SSA countries.** The fifth column of Table 2 shows the relative HIV prevalence of the DHS observations over their corresponding peaks. While there are four countries at (or very close to) their peak when the DHS data were collected (Lesotho in 2004, Mali in 2006, Niger in 2006, and Swaziland in 2002), most SSA countries have a relative HIV prevalence range between .75 and .95. The countries with the lowest relative HIV prevalence are Rwanda (.46) and Zimbabwe (.66).

4. **There exists a large degree of heterogeneity on the HIV peak year across SSA countries.** The earliest country to reach its peak in the DHS sample is the Democratic Republic of Congo in 1988 followed in early 1990s by Burkina Faso in 1991 and Rwanda in 1993 (see the second column in Table 2). Most SSA countries peak in the later 1990s and early 2000s: Zambia in 1995; Ethiopia, Kenya, and Tanzania in 1996; Cote d'Ivoire, Ghana, and Zimbabwe in 1997; Malawi in 1998; Cameroon in 1999; Lestotho in 2000; and Liberia and Swaziland in 2002. The last four countries in the DHS sample to reach their respective peak are Niger in 2006, Mali in 2007 and Guinea in 2010, whereas Senegal is still on the rising side of the epidemic and has not reached its peak yet.[49]

5. **DHS data were collected at different years for different countries.** See the fourth column in Table 2 and Figure 1. For instance, data from Burkina Faso were collected in 2003, Malawi in 2004, Zimbabwe in 2005, Mali in 2006, and Liberia in 2007.

6. **There exists a large degree of heterogeneity on the time distance between the HIV peak year and the year DHS data were collected across SSA countries.** See the sixth column in Table 2. For example, in 2006 data were collected for Mali, Swaziland and Zambia in 2006. At that time, Mali had not reached its peak yet (it did one year later, 1.6% in 2007); however,

[49]Guinea's HIV peak is the one projected by the United Nations. We do not have analogous projections for Niger and Senegal but data from 1980 to 2007. Niger attains its HIV peak during that interval, in 2006, while Senegal is still rising in 2007.

Swaziland and Zambia had already passed their peaks, respectively, of 26.8% in 2002 and 16.4% in 1995. That is, Swaziland and Zambia passed their respective peaks 4 years and 11 years before data collection. A similar phenomenon occurs in all other years.

7. **There exists a large degree of heterogeneity on the speed at which SSA countries reach their HIV peak.** The rising side of the HIV epidemic has very different paces across SSA countries. To see this in great detail see the seventh and ninth columns in Table 2. The ninth column shows how quickly a country that is halfway from its peak on the rising side of the epidemic moves toward its peak. For instance, a value of .10 indicates that every year Burkina Faso raises its HIV prevalence—relative to its peak—by one-tenth. This means that, at this average speed of .10, Burkina Faso moves from halfway, $t^{BF}_{-.5}$, to its peak, t^{BF}_{*}, in about 5 years. The highest speeds are those of the Democratic Republic of Congo and Ghana .17, which moved from half way to their respective peaks in 3 years. [50] The slowest countries to peak are Guinea and Mali, .04. Finally, Senegal is still rising.

8. **There exists heterogeneity on the speed SSA countries move away from their HIV peak.** Countries tend to move slowly from their HIV peak, see the eigth and tenth columns in Table 2. The average reduction from the peak, t^i_*, to halfway, t^i_*, in terms of relative HIV prevalence, is about .02 per year. This implies an average of 25 years to reach halfway to the peak prevalence. Further, the distribution of the speed of this reduction is very skewed, with only two countries above average, Rwanda (.05) and Zimbabwe (.04), and two countries about average Cote d'Ivoire and Guinea, both 02. The rest are below .02. Finally, note that for all countries the pace at which they approach their respective peak is always faster than when they move away from the peak. The average speed toward the peak is .10, which is about 5 times higher than the average speed of the decline, .02. This gives rise to the standard asymmetric bellshape of the HIV epidemic.

These observations are even more transparent for larger sets of countries, such as for all SSA and world countries with available estimates for HIV (see the nonnormalized figures at our companion site "http://rsantaeulalia.wustl.edu/Stages-of-HIV-Epidemic.html").

[50] Speeds below .17 and above or equal to .10 are those of Burkina Faso, Ethiopia, Kenya, Lesotho, Malawi, and Zambia. Below .10 and $\geq$ .05 we have Cameroon, Cote d'Ivoire, Guinea, Liberia, Mali, Niger, Rwanda Swaziland, Tanzania, and Zimbabwe.

Online Appendix

*Not for publication

A Further Margins of Risky Sexual Behavior

In Subsection 6.2 we documented the dynamic relationship between education and the number of sex partners other than spouses, as well as condom use, over stages of the epidemic. This section further describes the relationship between education and additional margins of risky sexual behavior (age at first marriage, formation rate of first marriages, and age at sexual debut) over stages of the epidemic, and discusses its association with the HIV-Education gradient. First, we discuss some properties of the risky sexual behavior variables that we study, including the ones used in the main text. Second, we discuss our results.

A.1 Some Descriptive Statistics

The number of sex partners (i.e., the extensive margin) other than spouses in the past 12 months is nontrivial. On average, this number is about .47, ranging from .08 partners in Ethiopia to 1.17 partners in Cameroon. Further, about 32% of our respondents reported having had at least one sex partner other than spouses in the past 12 months, about 24% reported only one extramarital partner, about 5% reported two extramarital partners, and about 2% reported three or more extramarital partners in the past 12 months. While the proportion of individuals without extramarital partners is very similar across countries—a Gini index of .12—differences start to increase when we look at the proportion of individuals with one or more extramarital partners, reaching a Gini index up to .46 when we focus on individuals with at least three extramarital partners.

The frequency of condom use (i.e., the intensive margin, quality) is on average rather low, around 16% with some dispersion in its use, from 4% in Ethiopia to 48% in Swaziland. Further, while the frequency of condom use is on average about 5% among individuals without extramarital partners, this frequency of use substantially rises to 38% for individuals with one extramarital partner and to 42% for individuals with two or more extramarital partners. Conditioning on the type of sex partner, individuals who unambiguously had last sexual intercourse with a noncasual sex partner (spouse or live-in partner) report a frequency of condom use of about 5%, while individuals who unambiguously had last sexual intercourse with a casual sex partner (commercial sex or casual acquaintance) report a frequency of condom use of about 45%.[51] That is, condom use is substantially higher (about nine times) for individuals who engage in risky sexual activities. Note that the similarity in condom use across countries also increases for individuals who engage in risky sexual behavior; that is, the Gini index of condom use when last sex partner is noncasual is .35, and about half of it, .17, when the last sex partner is casual.[52]

As for the stability of sexual parternships, we focus on two alternative measures related to first marriages. Our first measure of stability of sex partnerships is the age at first marriage (Stability I).[53] The notion of increased stability underlying the age at first marriage is that it establishes the formation

[51]We omit from our analysis sex partners whom we cannot unambiguously allocate to either casual or noncasual partner groups. This includes girlfriends, fianceés, relatives (other than spouses), friends, and others.

[52]While an accurate measure of the intensive margin is the frequency of sexual intercourse, unfortunately, DHS do not report such data. Therefore, our measure on the intensive margin—whether our respondent used a condom or not during last intercourse—is more accurately described as a quality measure of the intensive margin of sexual behavior. In this regard, we think economic research would benefit greatly if DHS were to ask questions on the frequency of sex and frequency of sex by types of partners in future surveys.

[53]Note, however, that we cannot consider age at first marriage as a statistic of partnership duration because not all individuals are still married since they first got married and further, even in the event that they are married, DHS do not provide information about whether they are still married to the first spouse or they are remarried (hence, married to someone else).

of stable sex partnerships over the life cycle. Since not all individuals in the each DHS dataset have married for the first time recently (but rather earlier or much earlier), they may inherit behavior from the pre-HIV era or from previous stages of the HIV epidemic. For this reason, we find it convenient to focus on the sample of individuals who were first married (newlyweds, first marriage) within the country-specific current stage of the HIV epidemic j at which the data were collected. This implies an age at first marriage of 23.9 years with few differences across countries, with a minimum at 22.7 in Malawi, a maximum 26.3 in Swaziland, and a Gini index of .01. The age at first marriage for the enitre sample (including those who did not marry for the first time in the current stage of the epidemic) is 23.1, suggesting a small rise in the age of first marriage for younger cohorts.[54] Our second measure of stability of sex partnerships is the formation rate of first marriages (Stability II). To evaluate this measure we construct for each individual a dummy that is equal to 1 if individuals were married for the first time within the country-specific current stage of the HIV epidemic j at which the data were collected (newlyweds, first marriage), and equal to zero otherwise. Our sample is formed by individuals who enter the stage j of epidemic as eligible for marrying for the first time. This includes individuals who report they were never married at the time of data collection and who entered their current country-specific stage j at age greater than 14 plus the newlyweds (first marriage) in stage j. Our findings suggest that, on average, about 65% of the individuals eligible for first marriage do marry in the current stage of the epidemic. This figure ranges from 33% in Swaziland to 91% in Tanzania with a Gini index of .16. The results of the formation of first marriages for the full sample (including those who did not marry for the first time in the current stage) are similar but less dispersed with a Gini index of .10.[55]

Finally, the average age at sexual debut (participation margin) is about 18.3 years among those individuals with sexual debut in the current country-specific stage j of the epidemic at which the DHS data were collected, which implies a gap of about 5.6 years on average between the age at sexual debut and the age at first marriage. There is little dispersion across countries (a Gini index of .04), with a minimum debut age of 16.2 in Kenya and a maximum debut age of 20.3 in Ethiopia. For the overall sample, including those with sexual debut in earlier stages of the epidemic, we find the age of sexual debut is virtually identical for the current debutants, 18.1 years.

[54] Precisely, for each individual k we know the country i he/she belongs to, his/her current age a_{ki}, and the age of at which he/she first married a^m_{ki}. Further, for each country we know the calendar year t^i_{DHS} at which DHS data were collected, the stage of the HIV epidemic j of the country at the time of data collection, and the calendar year t^{ij}_0 at which country i entered the stage j of the HIV epidemic. Hence, we can infer that individuals have married for the first time during the current epidemiological stage if $t^i_{DHS} - t^{ij}_0 \geq a_{ki} - a^m_{ki}$, and the opposite otherwise.

[55] The fraction of total first-time newlyweds over total eligible individuals for first marriage defines the rate at which sex partnerships of this kind—first marriages—form. The rate of formation of first marriages, unfortunately, exhausts what can be done using current DHS data in terms of computing statistics of sex partnership formation and destruction. In this context, a valuable set of information that we urge for future DHS collection is the duration of sex partnerships—specifically, the initial and terminal dates of each sex partnership. These are the minimum requirements to compute rates of sex partnership formation (e.g., within noncasual partnerships: first marriage and remarriages rates) and destruction (e.g., within noncasual partnerships: divorce and death rates). We argue these rates of partnership formation and destruction are the most useful factors (as opposed to the cross-sectional distribution of sex partnerships) to define the stability of sex partnerships because they uniquely identify the population processes (for example, within the noncasual partnerships, these processes are first marriage/divorce/remarriage/death) which gives rise to the cross-sectional distribution of sex partnerships (within the noncasual these partnerships are never married/married/divorced/widowed/remarried) that we observe in the data, while the opposite is not true. That is, there are infinitely many different population processes (hence, sexual behavior choices in terms of partnership formation and destruction) that can give rise to the same cross-sectional distribution of sex partnerships observed in the data.

A.2 Further Risky Sex-Education Gradients

This section shows the results of the Risky Sex-Education gradients associated with the stability measures of sexual partnerships and age at sexual debut.

The Stability-Education Gradients Increasing the stability of sex partnerships has been proposed as one major policy for the reduction of the sexual transmission of HIV (see UNAIDS (2010)). This section focuses on two alternative measures of stability: age at first marriage (Stability I) and the formation rate of first marriages (Stability II). Table A-1 reports the results for the nonstationary specification in equation (8) for the age at first marriage (henceforth, the Stability I-Education gradient) in column 1 and for the formation rate of first marriages (henceforth, the Stability II-Education gradient) in column 2. Our findings are summarized by the following:

Result: The Stability I-Education gradient is not significant over stages of the HIV epidemic.

This suggests that SSA populations do not show educational disparities in the age at first marriage in response to the HIV epidemic. However, the Stability II-Education gradient is significant in all stages except the last, Stage 4, but its behavior is rather erratic over the evolution of the epidemic. Specifically, we find that the Stability II-Education gradient always changes sign across two consecutive stages. In particular we find:

Result: The evolution of the Debut-Education gradient is significant but erratic with a correlation between the gradient at stage j and its lag, stage $j-1$, equal to -.89.

This way, while the formation of first marriages of more-educated individuals relative to less-educated individuals changes over stages of the epidemic, the behavior of the Stability II-Education gradient only weakly tracks the nonstationary behavior of the HIV-Education gradient. Indeed, the correlation coefficient between the Stability II-Education gradient and the HIV-Education gradient is -.28.

The Debut-Education Gradient The delay of first sexual intercourse is one pursued policy in the prevention of HIV infection. Here, we investigate the behavior of sexual debutants. Specifically, we explore whether the age of sexual debut for individuals who have had their first sexual intercourse in the country-specific stage j changes across educational groups and over epidemiological stages (henceforth, the Debut-Education gradient).[56] In other words, are more-educated individuals deferring (with respect to less-educated individuals) their age of sexual debut with the course of the epidemic? Column 3 in Table A-1 shows our results for the Debut-Education gradient. While at the earliest Stage 0, 5 more years of schooling are associated with a lower age of sexual debut of about eight months (-.141*5= -.70 of a year), at further stages of the epidemic the relationship between the age of sexual debut and education displays an erratic behavior that can be summrized by the following:

Result: The evolution of the Debut-Education gradient is significant but erratic with a correlation between the gradient at stage j and its lag, stage $j-1$, equal to -.78.

Indeed, there is a small association between the Debut-Education gradient and the HIV-Education gradient summarized by a a negative correlation coefficient of -.22.

[56] Precisely, for each individual k we know the country i he/she belongs to, his/her current age a_{ki}, and the age of his/her sexual debut a^d_{ki}. For each country we also know the calendar year t^i_{DHS} at which DHS data were collected, the stage of the HIV epidemic j of the country i at the time of data collection, and the calendar year t^{ij}_0 at which country i entered the stage j of the epidemic. Hence, we can infer whether individuals are sexual debutants in the current epidemiological stage in which the data are collected by checking whether $t^i - t^{ij}_0 \geq a_{ki} - a^d_{ki}$.

Table A-1: Further Risky Sex-Education Gradients

	Stability I	*Stability II*	*Debut*
Dependent Variable:	*Age at First Marriage*	*Newlyweds, First Marriage*	*Age at Sexual Debut*
Sample:	*Newlyweds, First Marriage*	*Eligible for First Marriage*	
	(1)	(2)	(3)
Education	-0.015	-0.010***	-0.141***
	(0.042)	(0.004)	(0.050)
Education * Stage 1	0.098**	0.021***	0.172***
	(0.045)	(0.004)	(0.052)
Education * Stage 2	-0.006	-0.008*	-0.010
	(0.047)	(0.004)	(0.055)
Education * Stage 3	0.090*	0.013***	0.120**
	(0.052)	(0.005)	(0.056)
Education * Stage 4	-0.092*	0.006	0.132**
	(0.048)	(0.005)	(0.054)
Age	0.658***	0.038***	0.504***
	(0.006)	(0.000)	(0.007)
Urban Area	0.510***	-0.122***	0.140**
	(0.080)	(0.007)	(0.059)
Stage 1	-0.592***	-0.051***	-1.263***
	(0.180)	(0.016)	(0.227)
Stage 2	-0.516***	-0.072***	-0.508**
	(0.187)	(0.016)	(0.239)
Stage 3	-0.284	-0.186***	-1.674***
	(0.206)	(0.018)	(0.243)
Stage 4	0.407**	-0.103***	-0.219
	(0.182)	(0.017)	(0.228)
Agricultural Share	0.044***	0.005***	0.024***
	(0.006)	(0.000)	(0.003)
Output per Capita	0.001***	-0.000*	0.000***
	(0.000)	(0.000)	(0.000)
Constant	1.563***	-0.422***	5.775***
	(0.375)	(0.030)	(0.318)
Sample Size	22,554	37,673	27,317
R^2	0.63	0.36	0.52

Notes: Sexually active men with fewer than 20 sex partners (other than spouses) in the past 12 months. Robust SE in parentheses. *Significant at 10%; ** significant at 5%; and *** significant at 1%.

Figure A-1: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 1980**. Normalized HIV Prevalence $\in [0, 1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).

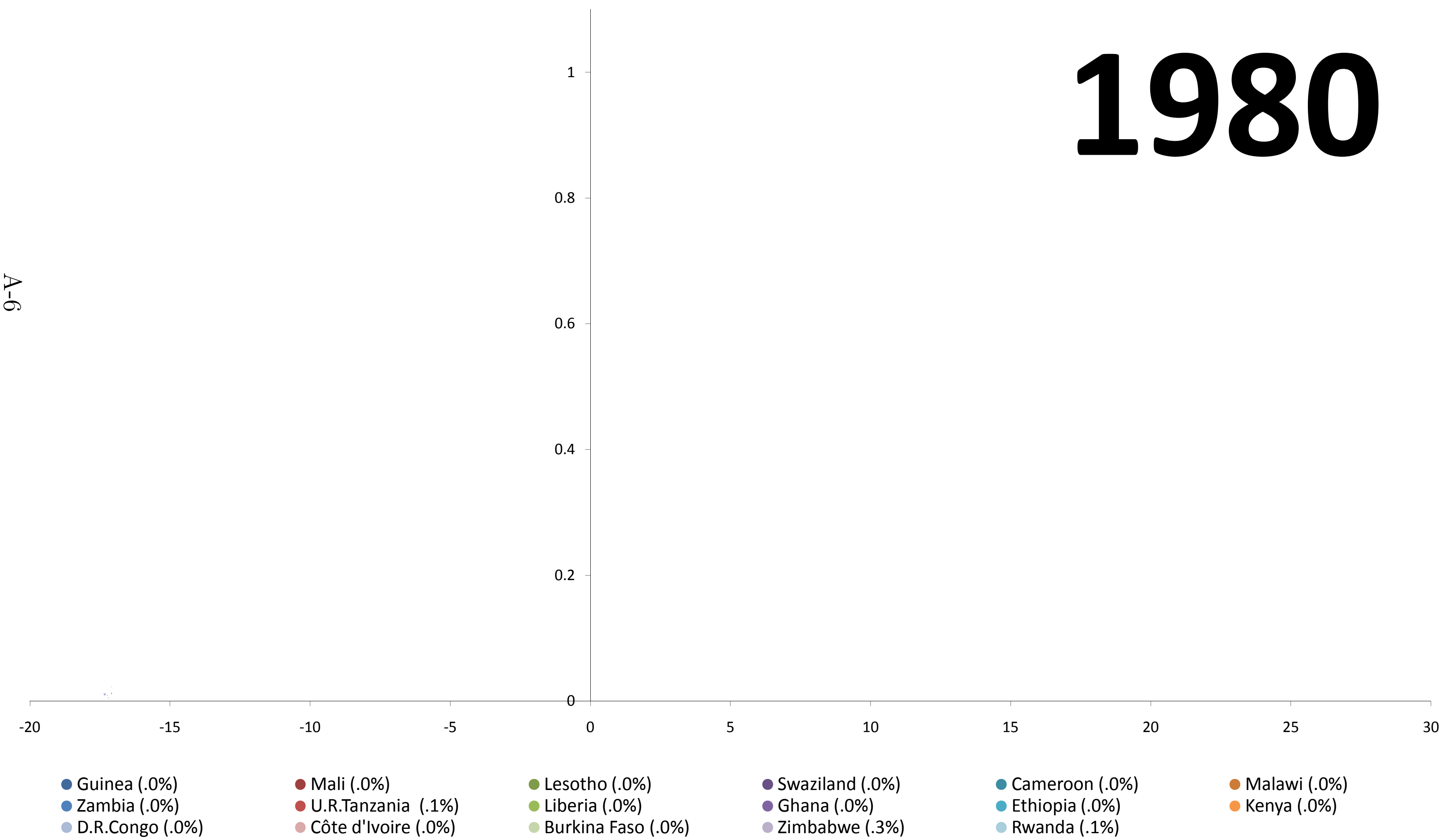


Figure A-2: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 1985**. Normalized HIV Prevalence $\in [0,1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).

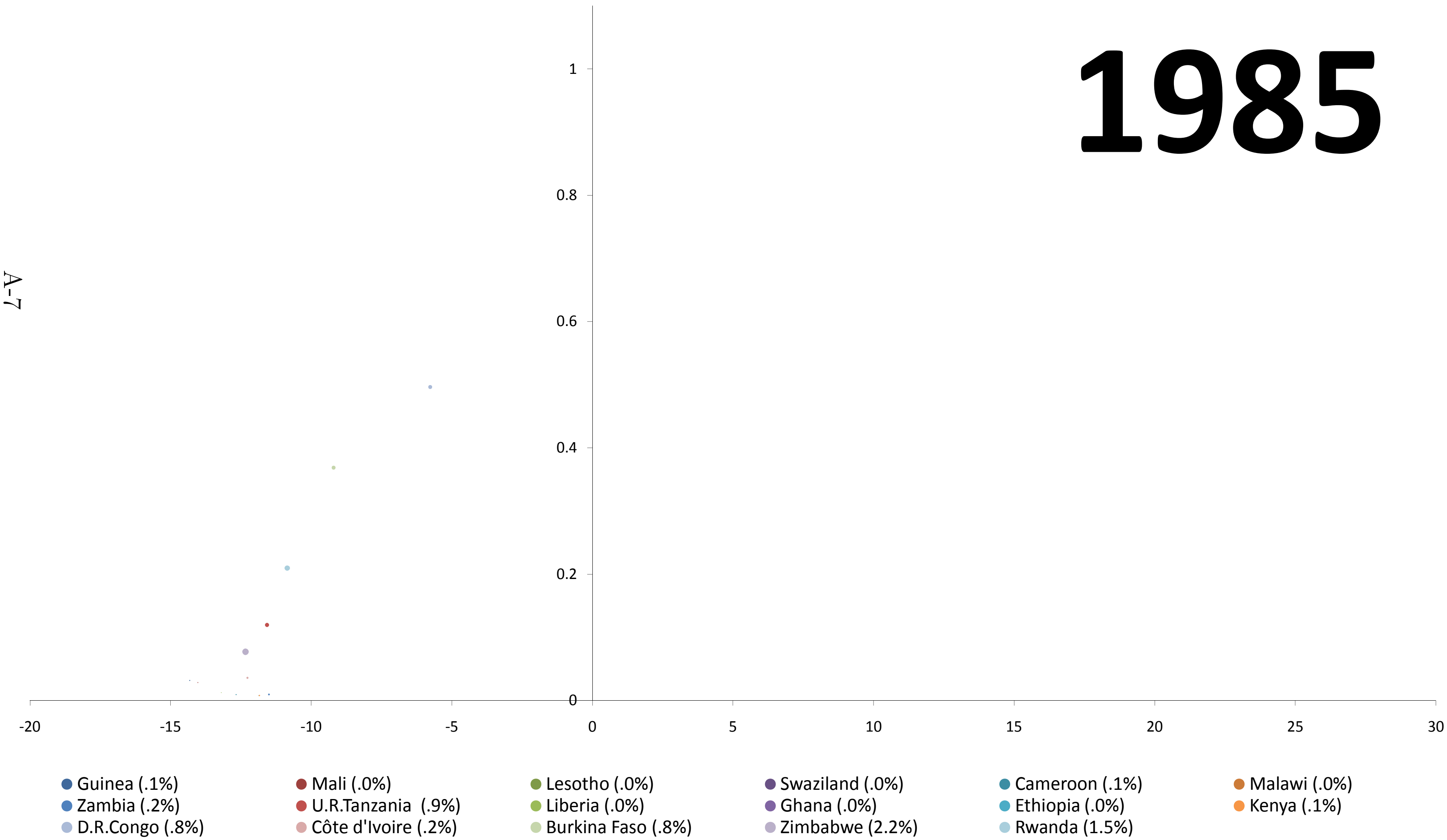


Figure A-3: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 1990**. Normalized HIV Prevalence $\in [0, 1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).

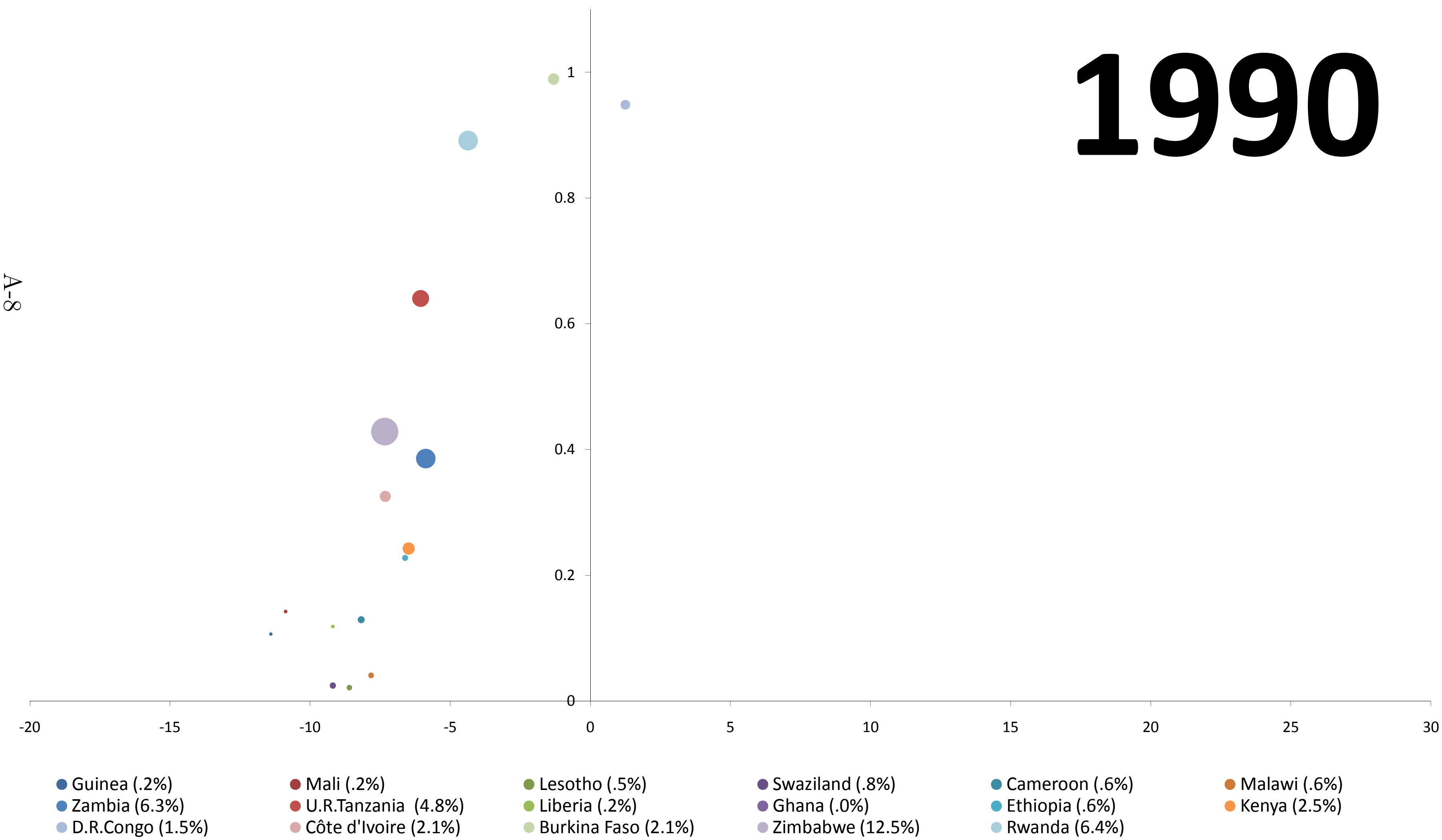


Figure A-4: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 1995**. Normalized HIV Prevalence $\in [0, 1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).

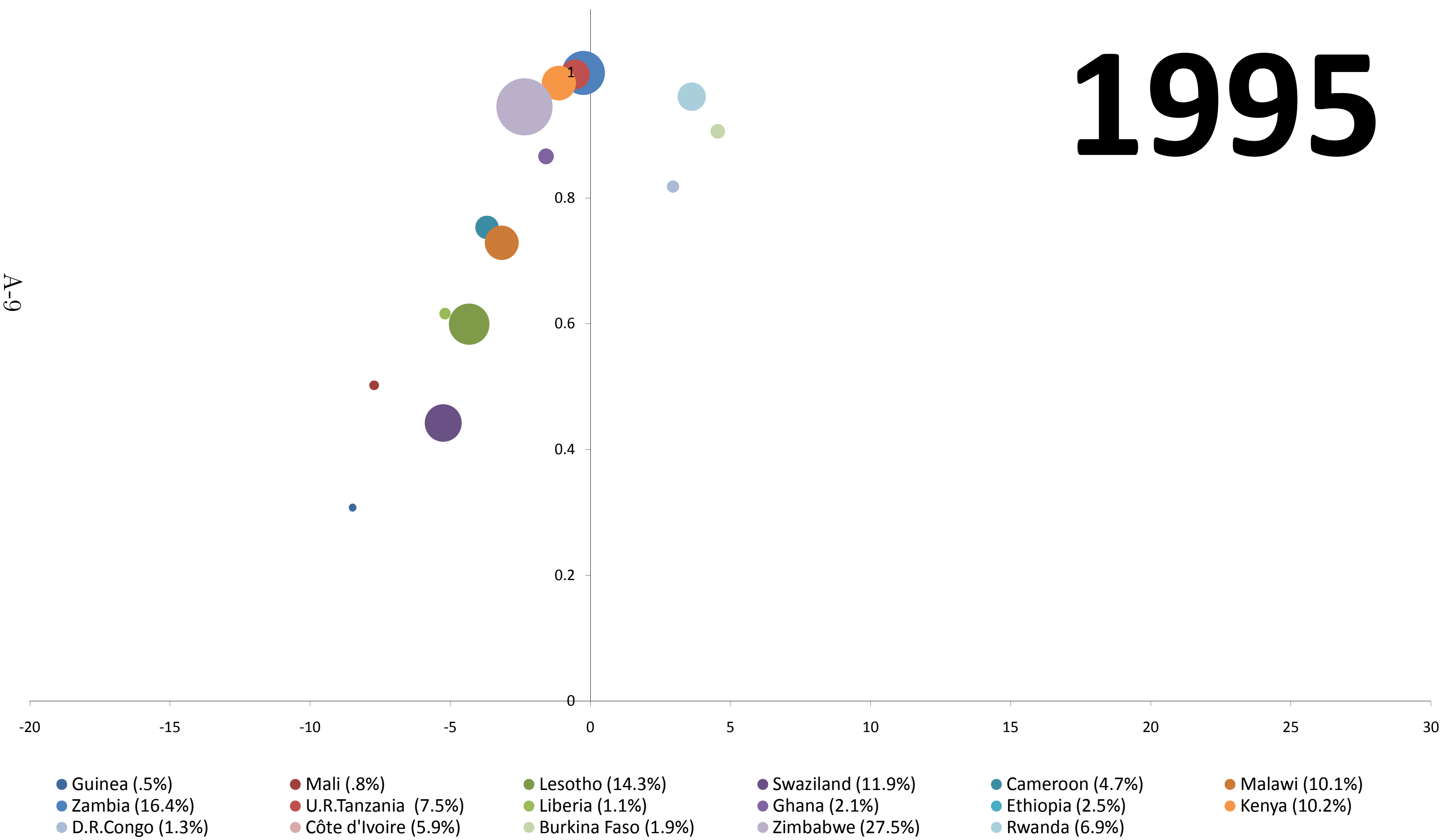


Figure A-5: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 2000**. Normalized HIV Prevalence $\in [0,1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).

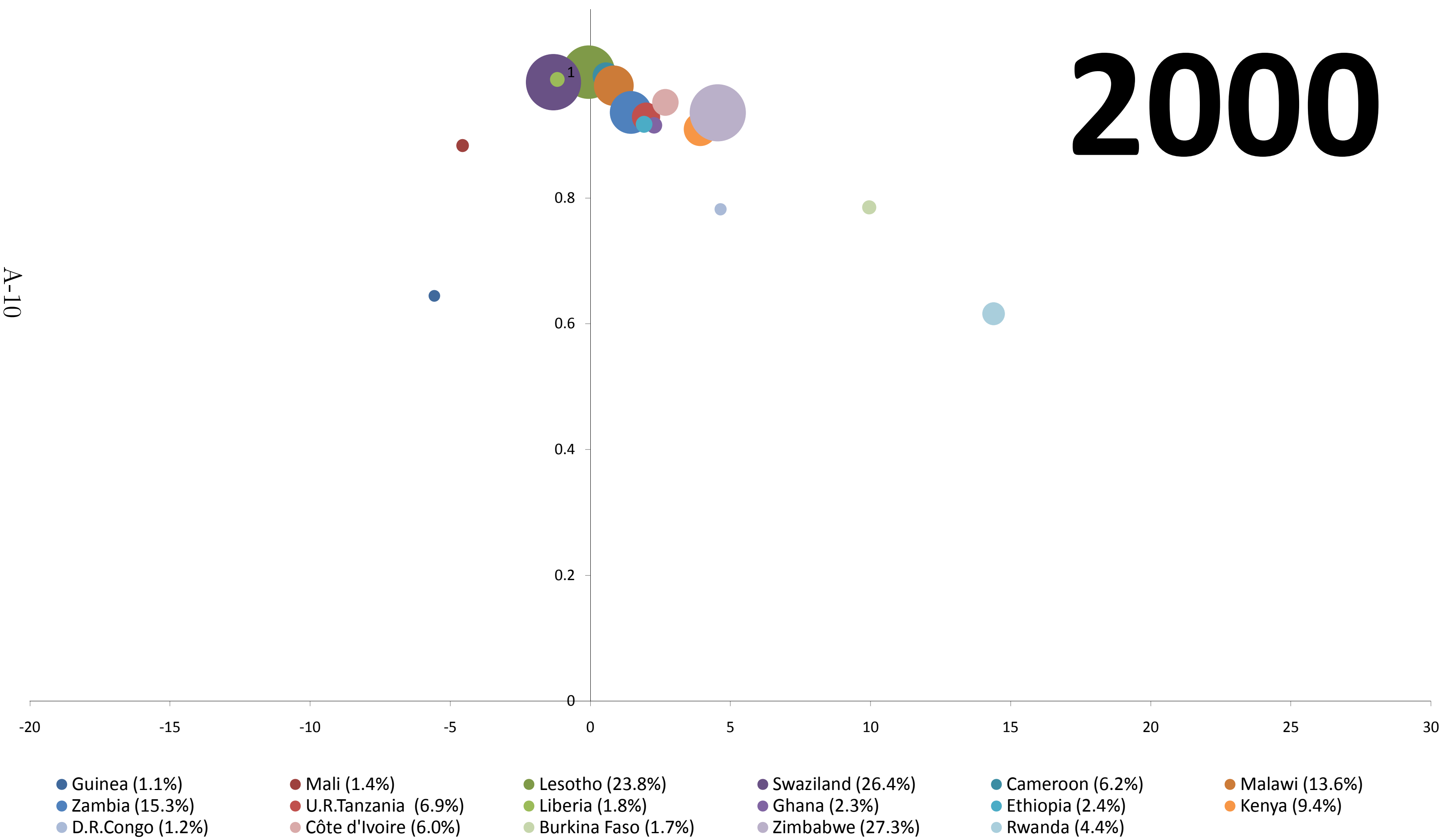


Figure A-6: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 2005**. Normalized HIV Prevalence $\in [0, 1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).

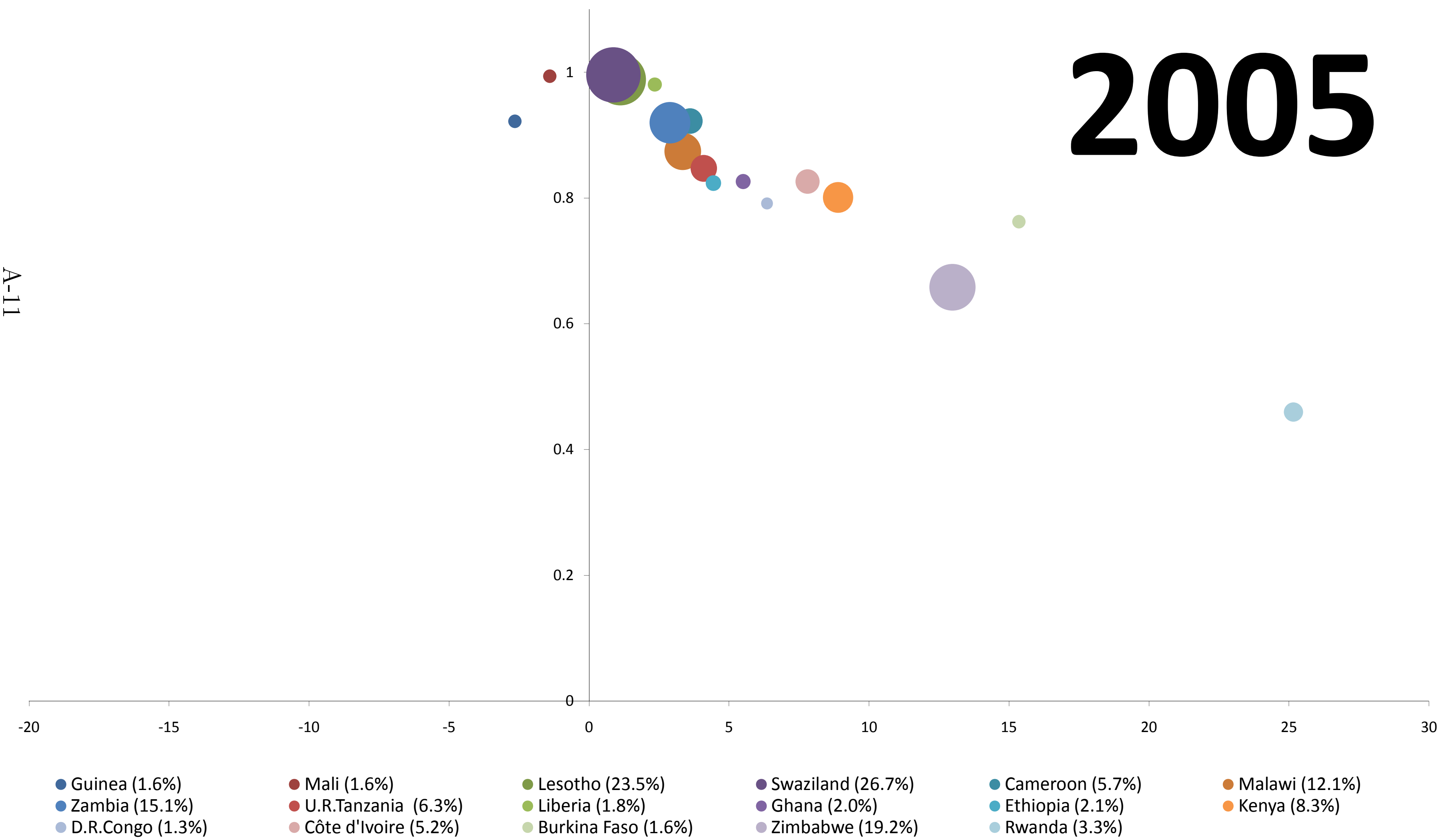


Figure A-7: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 2010**. Normalized HIV Prevalence $\in [0, 1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).

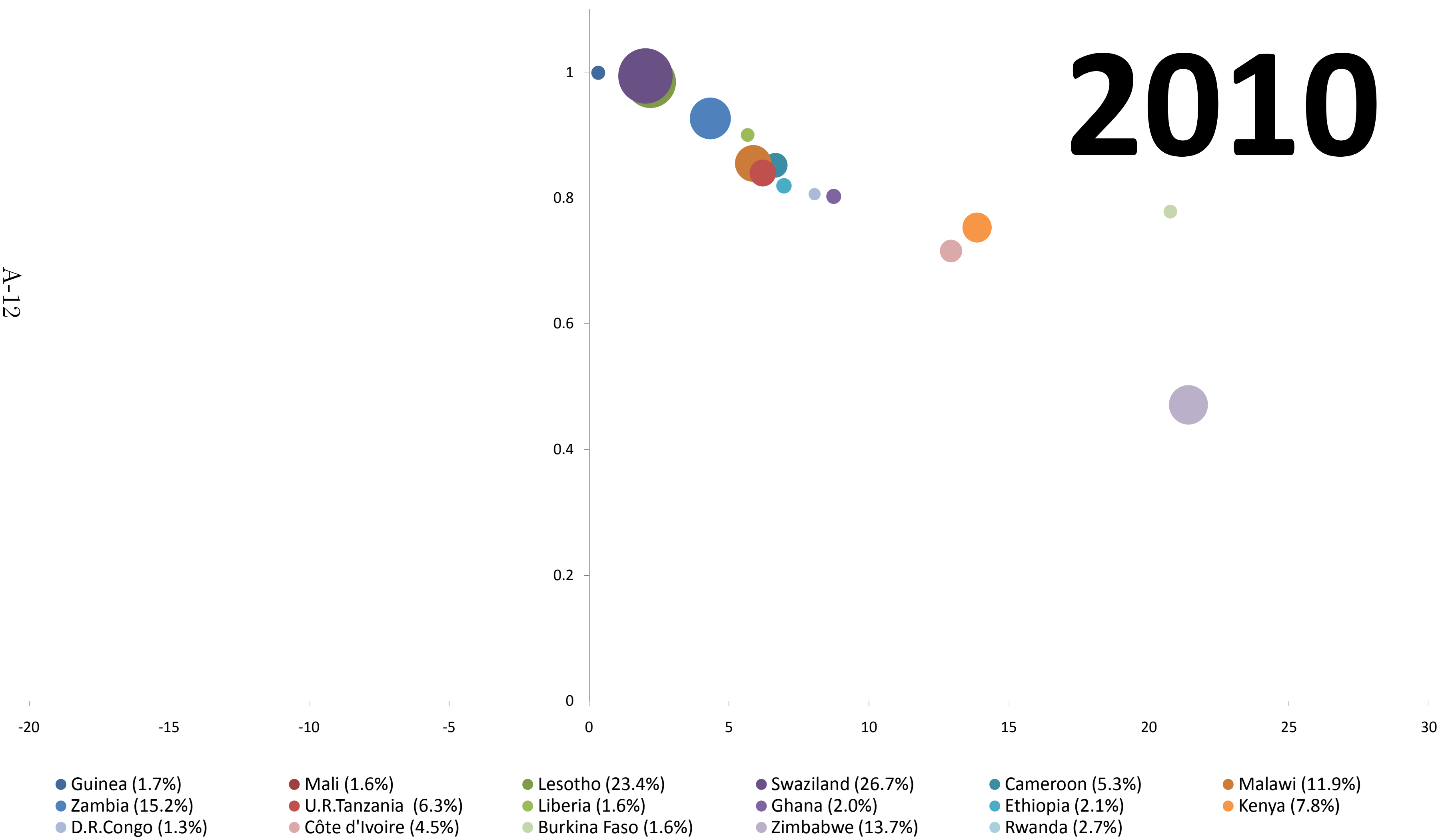


Figure A-8: **2D-Normalized HIV Epidemic: SSA Countries (DHS Sample), 2015**. Normalized HIV Prevalence $\in [0, 1]$ (vertical axes) and normalized time τ (horizontal axes). Bubble size: Country-specific current HIV prevalence (also in brackets in the legend).