WPS6343

Policy Research Working Paper 6343

How Subjective Beliefs about HIV Infection Affect Life-Cycle Fertility

Evidence from Rural Malawi

Gil Shapira

The World Bank
Development Research Group
Human Development and Public Services Team
January 2013

Policy Research Working Paper 6343

Abstract

This paper studies the effect of subjective beliefs about HIV infection on fertility decisions in a context of high HIV prevalence and simulates the impact of different policy interventions, such as HIV testing programs and prevention of mother-to-child transmission, on fertility and child mortality. It develops a model of women's life-cycle, in which women make sequential fertility decisions. Expectations about the life horizon and child survival depend on women's perceived exposure to HIV infection, which is allowed to differ from the actual exposure. In the model, women form beliefs about their HIV status and about their own and their children's survival in future periods. Women update their beliefs with survival to each additional period as well as when their HIV status is revealed by an HIV test. Model parameters are estimated by maximum likelihood with longitudinal data from the Malawi Diffusion and Ideational Change Project, which contain family rosters, information on HIV testing, and measures of subjective beliefs about own HIV status. The model successfully fits the fertility patterns in the data, as well as the distribution of reported beliefs about own HIV status. The analysis uses the model to assess the effect of HIV on fertility by simulating behavior in an environment without HIV. The results show that the presence of HIV reduces the average number of births a woman has during her life-cycle by 0.15. The paper also finds that HIV testing can reduce the fertility of infected women, leading to a reduction of child mortality and orphan-hood.

This paper is a product of the Human Development and Public Services Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org.
The author may be contacted at gshapira@worldbank.org.

How Subjective Beliefs about HIV Infection Affect Life-Cycle Fertility: Evidence from Rural Malawi*

Gil Shapira†

January 30, 2013

JEL classification codes: O15, I10, J13

Keywords: fertility, HIV/AIDS, subjective beliefs

Sector board: Health, Nutrition and Population (HNP)

*I thank my dissertation committee members Jere Behrman, Áureo de Paula, Kenneth Wolpin, and especially my advisor Petra Todd for their dedication and support. I am also grateful to Clement Joubert, Nirav Mehta, Shalini Roy, and participants of seminars at the University of Pennsylvania, International Food Policy Research Institute, The World Bank, Tel Aviv University, Haifa University, Hebrew University, Ben Gurion University, Interdisciplinary Center Herzliya, and University of Washington for helpful comments. I gratefully acknowledge financial support from the William and Flora Hewlett Foundation.
The findings, interpretations and conclusions expressed in this paper are entirely those of the author, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

†Development Research Group, The World Bank. Email: gshapira@worldbank.org

1 Introduction

Both fertility and HIV prevalence rates in Malawi are among the highest in the world, with the total fertility rate at 5.7 births per woman and the HIV prevalence rate at 10.6 percent.[1] Malawian women make fertility decisions in an environment characterized by high adult and child mortality, exacerbated by mother-to-child HIV transmission. Out of a population of about 15 million, it is estimated that 68,000 die annually from AIDS and that 560,000 children under the age of 17 have lost at least one parent to the disease.[2]

There are many policy interventions aimed at reducing HIV in Malawi and other Sub-Saharan African countries. These include HIV testing programs, information campaigns, and antiviral distribution programs. Evaluating the effects of such policies on outcomes such as number of births, child mortality, and orphan-hood requires an understanding of how women's fertility decisions are affected by the presence of HIV.

An important aspect of the environment in Malawi is that women are typically uncertain regarding their own HIV status. An infected person can live for many years with no symptoms, and testing was not widely available until relatively recently. The median survival time after infection, without treatment, is about 10.4 years.[3] During most of this time, an infected person is in a clinical latency stage and experiences few or no symptoms.[4]

In addition to being uncertain about own HIV status, women often express beliefs about HIV risk that differ substantially from actual risk. Studies using the Malawi Diffusion and Ideational Change Project data show that individuals in rural Malawi tend to overestimate both the probability of being HIV-infected (Anglewicz and Kohler, 2009) and the HIV prevalence in their community (Anglewicz, 2007). Anglewicz and Kohler (2009) attribute these high risk assessments to overestimated probabilities of transmission. More than 95 percent of respondents believe that transmission from a single instance of unprotected intercourse

[1] Malawi Demographic and Health Surveys (2010)

[2] UNSAID/WHO/UNICEF Epidemiological Fact Sheets (2008)

[3] Todd et al. (2007). Estimate for infected adults in Eastern and Southern Africa

[4] Morgan et al. (2002), for example, find median time from infection to AIDS to be 9.4 years and median time from AIDS to death to be 9.2 months in rural Uganda.

with an infected person is highly likely or certain; however, studies estimate that it can be as low as 1 per 1,000 encounters in the absence of an increased viral load (Gray et al., 2001).[5]

Women's perceptions of HIV risk and of their own HIV status affect beliefs about their own and their children's life expectancy, which in turn may influence life-cycle fertility choices. In this paper, I study the determinants of women's reproductive decisions in Malawi, taking into account uncertainty about HIV status and differences between perceived and actual HIV infection risk. I investigate how HIV affects fertility and simulate the impact of different policy interventions, such as HIV testing programs and prevention of mother-to-child transmission, on fertility and child mortality.

To this end, I develop a dynamic discrete-choice life-cycle fertility model in which expectations about the life horizon and child survival depend on a perceived infection hazard. A woman makes annual pregnancy decisions from the time of marriage until she becomes infecund. She maximizes utility, which depends on her number of children, household consumption, and pregnancies, subject to a per-period budget constraint. The woman faces uncertainty regarding future income, HIV status, and the survival of herself and her children in future periods.

A woman's perceived infection hazard is allowed to differ from her actual infection hazard to reflect the misperceptions about HIV risk observed empirically. The perceived hazard rate for each period is a function of a woman's characteristics, such as her age, region of residence, marital status, and schooling level. To account for unobservable factors, the hazard rate also incorporates heterogeneity in the form of a discrete number of unobserved types.[6]

Given that HIV is initially asymptomatic, the model assumes that a woman does not

[5]The viral load is high in the few weeks following infection and increases again as an infected person develops AIDS. Infectivity increases with viral load as well as with other conditions, such as the existence of other STDs. Powers et al. (2008) review the studies estimating HIV infectivity and discuss different factors that increase infectivity. Note however that even a low transmission rate such as 0.001 can translate into a nontrivial probability of infection during a year of partnership. Gray et al. (2001) find average frequency of intercourse to be about 106 acts per year, implying about a 10 percent chance of transmission given this transmission rate.

[6]The probability of being a certain type is a function of the woman's characteristics and follows a multinomial logit specification.

observe realizations of the infection process and therefore does not know her HIV status. Assuming she knows the mortality process associated with HIV infection, however, survival during each additional period gives her information about her status. Specifically, she reduces her subjective probabilities of having become infected in each of the past periods based on the fact that she is still alive. According to these probabilities and given the mortality process, she updates her survival expectations. HIV infection also increases child mortality probabilities through mother-to-child transmission. In the model, the woman also updates expectations about the survival of each of her children depending on the probability assigned to her having been infected at the time of birth.

The dynamic fertility model is estimated using the Malawi Diffusion and Ideational Change Project (MDICP) dataset, a rich longitudinal dataset collected in rural areas of three different districts of the country. The data contain extensive information on more than 4,000 individuals at the individual and household level. The three sampled regions vary significantly in several aspects that are potentially relevant for the analysis, such as HIV prevalence rates, polygamy rates, and schooling levels. A unique feature of the MDICP data is that they include measures of subjective expectations regarding a range of outcomes, including the likelihood respondents assigned to being HIV-infected at the time of the interview. The expectations data were collected using a novel bean-counting method, developed by Delavande and Kohler (2009), which is appropriate for populations with low levels of numeracy. Delavande and Kohler (2009) find that the reported subjective expectations follow basic properties of probabilities and that the assessments of HIV-infection vary meaningfully with observable characteristics associated with different levels of HIV prevalence. I use a subsample of 1006 married women who were interviewed at least once during the 2006 and 2008 rounds, when family rosters and subjective expectations were collected.

Since 2004, each round of data collection included HIV testing of respondents and the prevalence rate was found to be about seven percent.[7] Although individuals who received

[7]HIV prevalence is lower in rural areas than in urban areas. The prevalence in the MDICP sample is lower than the rural 2004 DHS prevalence because the sample does not include peri-urban areas such as

positive test results assign significantly higher likelihood to being infected two years later relative to individuals who received negative test results, some individuals who tested positive later assign a probability of less than one to being infected. I therefore assume that a woman assigns a probability to the accuracy of the test. Given the test result, the woman updates the probabilities assigned to infection and the corresponding probabilities of survival.

I structurally estimate the model parameters using maximum likelihood and then use the estimated model to perform several counterfactual simulations. First, I simulate life-cycle fertility in an environment with no HIV exposure. The results indicate that the presence of HIV has an average negative effect on fertility. Overall, women in the no-HIV environment have on average 0.15 more births over their life-cycle. However, there is heterogeneity in the effect of the presence of HIV on fertility, with some women decreasing the number of births and fewer women increasing it. I also simulate the effects of prevention of mother-to-child transmission and HIV testing programs on fertility outcomes. Although these programs are not necessarily intended to influence fertility, they have the potential to affect reproductive choices by altering beliefs about HIV status, life expectancy, and child survival. Such policy interventions are implemented in Malawi and are continuously expanding in scope. I find these programs to have negligible effects on the overall number of births. However, elimination of mother-to-child transmission will reduce fertility by infected women who respond to child mortality by additional pregnancies either because they have high marginal utility of an additional child or because they assign low likelihood to being infected. I also find that HIV testing can reduce fertility by HIV-positive women, leading to a reduction in child mortality and orphan hood. However, the effect of testing is limited by the partial updating of beliefs.

trading centers (Obare et al., 2009).

1.1 Related Literature

Several studies analyzed the response of fertility in Sub-Saharan Africa to the HIV/AIDS epidemic and reached mixed conclusions. Young (2005) studies the effect of the HIV/AIDS epidemic on the welfare of future African generations. He concludes that thanks to a negative effect on fertility, the epidemic, on net, will increase future per capita consumption.[8] He attributes the reduction in fertility to an unwillingness to engage in unprotected sexual activity and increasing labor opportunities for women because of scarcity of labor. He shows empirically a negative relationship between fertility and HIV prevalence using retrospective fertility histories and seroprevalence in antenatal clinics in South Africa. Using time series cross-country data on fertility and HIV prevalence rates, Young (2007) finds a similar effect. Kalemli-Ozcan (2006), using similar types of data, shows that regressing total fertility rate on HIV/AIDS prevalence can yield both positive and negative effects with different estimation strategies and different measures of HIV prevalence.

Later studies use the cross-country data from the latest rounds of the Demographic and Health Surveys (DHS). These surveys are nationally representative and contain results of HIV testing of respondents. Fortson (2009) and Juhn et al. (2008) find lower fertility rates of HIV-infected women than of HIV-uninfected women; however, they find no effect of local prevalence rates on the fertility of uninfected women and an overall insignificant aggregate effect of HIV on fertility. Fink and Linnemayr (2009), linking historical data from World Fertility Surveys (WFS) with the DHS data, argue that while HIV does not have a significant effect on aggregate fertility levels, it affects differently women depending on their educational attainment. They find that, in the presence of HIV, more educated women reduce fertility more than uneducated women. Analyzing the Malawi DHS surveys from 2000 and 2004, together with HIV rates obtained from antenatal clinics, Durevall and Lindskog (2011) conclude that although the HIV/AIDS epidemic has small impact on the number of

[8] According to Young (2005), the positive effect of reduces fertility dominates a negative effect of reduced human capital accumulation by orphans.

births a woman experiences it affects timing of fertility. Women in districts with higher HIV prevalence are more likely to give birth at younger ages and are less likely to do so when they are over 29 years.

My analysis differs from these studies both in the empirical approach and in the type of data used in the analysis. The data on individuals' subjective beliefs allow me to exploit heterogeneity within communities, which the use of data on local HIV prevalence only does not allow. The structural estimation of a model enables me to perform counterfactual simulations to assess fertility patterns in different environments.

Many economists have studied the determinants of fertility in different environments.[9] My analysis is most closely related to studies that model fertility decision-making as a sequential process (Heckman and Willis (1974)) and studies of fertility in environments with non-negligible infant and child mortality risk (Wolpin, 1984; Sah, 1991; Mira, 2007). Wolpin (1984) presents an estimable dynamic discrete-choice fertility model in an environment where infant survival is uncertain and uses the model to study the response of fertility choices to experienced infant mortality. Mira (2007) uses a similar modeling framework and extends it by introducing heterogeneity in infant mortality risk across mothers. Parents in his model learn about a family-specific component of infant mortality risk throughout their life-cycle. Fertility choices are influenced by how the parents adapt to the information received from infant survival and mortality.

2 Model

2.1 General Setup

I develop a dynamic discrete-choice life-cycle model of woman's fertility decisions in an environment of exposure to HIV infection. Women maximize subjective expected utility by making sequential binary fertility choices in a framework similar to that of Wolpin (1984) and

[9] Joseph Hotz et al. (1997), Schultz (1997), and Wolpin (1997) survey the literature on fertility.

Mira (2007). Specifically, a woman makes annual decisions of whether to become pregnant beginning at the age of her marriage and ending when she becomes infecund at a fixed age F (assumed to be 45). A woman gives birth in the period following the one in which she became pregnant. If never infected, a woman survives with certainty to age T (assumed to be 60). An infected woman might die prior to reaching the terminal model period. HIV infection of a mother at time of birth also increases mortality probabilities of children of ages zero to three. Given that HIV infection is asymptomatic for the majority of the infection duration, women cannot observe their HIV status.

Women are heterogeneous with respect to a group of characteristics that are treated as exogenous and constant determinants of their choices. These characteristics include region of residence, completed schooling level, the size of the household's land plot, age of marriage, and whether a woman is married to a polygamous husband. Women are also of different discrete unobserved types, which are incorporated in the model to account for unobservable permanent factors which might affect preferences as well as exposure to HIV.

2.2 Preferences

Each period, a woman receives a utility flow from household consumption (C), her number of children (N), pregnancy status (p), and a time-varying preference shock (ϵ_p) which is iid across time and women. The per-period utility function is given by

$$U(t) = \frac{C(t)^{\phi}}{\phi}\left[1 + \exp\left(\lambda_1 N(t)\right)\right] + \lambda_{2,r,e,m,\mu} N(t) + \lambda_{3,r,e,m,\mu} N(t)^2 - \left(\lambda_{4,t} + \epsilon_p(t)\right) p(t) - (\lambda_5 + \lambda_6 t) p(t) p(t-1),$$

$$\epsilon_p(t) \sim iidN\left(0,\ \sigma_p^2\right).$$

The utility function exhibits constant relative risk aversion (CRRA) in consumption and includes an interaction term between household consumption and the number of children to reflect consumption being divided among more individuals as the household size increases. The utility is quadratic in the number of children. The parameters related to preference for children, λ_2 and λ_3, are allowed to vary with region of residence (r), schooling level (e),

polygamy (m), and unobserved type (μ).[10] The utility function also incorporates a non-pecuniary cost (or benefit) associated with being pregnant. This cost includes a stochastic preference shock as well as a deterministic age-dependant element ($\lambda_{4,t}$). The cost of pregnancy changes if a woman was pregnant in the previous period (λ_5). The cost of consecutive births is allowed to change with the age of women (λ_6).

The specification of the model implies perfect control over conception and contraception. I could have instead introduced a cost of contraceptives, a probability of conception conditional on not wanting to become pregnant and a probability of not conceiving conditional on trying. Using data on births only, however, I cannot separately identify these elements. Instead, these elements will be absorbed into the cost of pregnancy (both deterministic and stochastic) and the preference for children. For example, having the cost of pregnancy change with age captures physiological factors which vary the propensity to conceive during different stages of a woman's life-cycle. The added cost of consecutive births enables the model to generate patterns of birth spacing.

2.3 Income and Consumption

It is assumed that households cannot borrow or save, which implies household consumption (C) equals the household's income (Y). The household's income is exogenous and stochastic. I specify a parsimonious household income function that is appropriate for the context of subsistence agriculture. I assume that the logarithm of income is distributed as

$$\ln\left(Y(t)\right) = \theta_1 + \theta_2 \text{Balaka} + \theta_3 \text{Mchinji} + \theta_3 \text{Land High} + \theta_4 N(t) + \theta_5 t + \theta_6 t^2 + \epsilon_y(t), \tag{1}$$

$$\epsilon_y(t) \sim iidN\left(0,\ \sigma_y^2\right).$$

Income depends on region of residence (Balaka, Mchinji), size of household's land plot (Land High), the age of the woman (t), the number of children (N), and a time-varying income

[10] The majority of respondents in each of the three sites of the MDICP dataset are of different tribal groups. The tribes differ in their practices of lineage and residence after marriage which might affect preferences for children.

shock (ϵ_y).[11] Realization of the income shock occurs after the fertility decision is made. Therefore, pregnancy decisions are based on expected income.

2.4 Perceived Infection Hazard

A woman's perceived exposure to HIV infection is modeled as a hazard process.[12] Let $h(t)$ be the probability a woman assigns to getting infected at period t, conditional on being HIV-negative until then. The perceived hazard rate for period t is given by

$$h(t) = \frac{1}{1 + \exp\left(-x(t)'\beta\right)}, \tag{2}$$

where $x(t)$ is a vector containing the woman's characteristics, the duration of her marriage, her age and age squared, and a constant. The parameters related to the constant, age and age squared are allowed to differ for unobserved types.[13]

The perceived unconditional probability of getting infected at period t, $P(t)$, is given by

$$P(t) = h(t) \prod_{k=1}^{t-1} \left(1 - h(k)\right). \tag{3}$$

2.5 Survival Expectations

A woman is assumed to know the mortality processes associated with HIV infection (for both adults and children). Survival to each additional period provides her with information about her HIV status. Specifically, she reduces the probability of having gotten infected in each past period. Given these probabilities and the mortality process, the woman assigns probabilities to survival to each future period. The woman also updates expectations about the survival of children of ages zero to three in future periods. Conditional on the probabilities she assigns

[11] The Malawi Multiple Indicator Cluster Survey 2006 estimates that 27.5 percent of children aged 5 to 11 are involved in at least one hour of economic work or 28 hours of domestic work per week. A more natural specification of the income function might take into account the ages of the children; however, I do not keep track of the ages of children above age 3 in the state space to reduce computational burden.

[12] Perceived and actual exposures to HIV infection are treated as exogenous to women's behavior because there is no variation in self reports of sexual behavior by women in the data. All but very few married women report having sex with their husbands and not having extramarital partners.

[13] The detailed specification of the perceived infection hazard function is presented in the appendix.

to having been infected at times of giving birth, she assigns probabilities to the survival of each of her children to future periods.

Women's beliefs are also updated by receiving an HIV test result. MDICP respondents did not anticipate being offered HIV testing, and almost all of the respondents agreed to get tested. Because of these features of the data, I abstract from modeling the decision to get tested. Instead, I treat the HIV testing as an unanticipated revelation of HIV status. I first present the updating of expectations about life horizon without testing and then proceed to discuss the updating with testing.

2.5.1 Updating Survival Expectations without HIV Testing

Updating probabilities assigned to infection in different periods:

Let $I(\tau,\ t)$ be the probability a woman assigns to getting infected at period τ, conditional on being alive at t. The probability assigned to having gotten infected at a *past or present* period τ is given by

$$I\left(\tau,\ t\right) = \frac{\Pr\left(\text{got infected at } \tau \text{ \& alive at } t\right)}{\Pr\left(\text{alive at } t\right)} = \frac{P(\tau)S\left(\tau,\ t\right)}{1-\sum_{k=1}^{t} P(k)\left(1-S(k,\ t)\right)}, \quad t \geq \tau, \tag{4}$$

where $S\left(\tau,\ t\right)$ is the probability a woman who gets infected at period τ survives to period t. The probability assigned at period t to being HIV-positive is given by the summation of the probabilities assigned to infection happening in all periods up to t:

$$B(t) = \sum_{k=1}^{t} I\left(k,\ t\right). \tag{5}$$

Probabilities assigned to survival in future periods:

The probability a woman assigns at time t to being alive at the following period is given by

$$\begin{aligned}\pi(t,\ t+1) &= \left(\textstyle\sum_{k=1}^{t} \Pr(\text{got infected at k}) \Pr(\text{alive at } t+1 \mid \text{ got infected at } k \ \&\ \text{ alive at } t)\right) \\ &\quad + \Pr(\text{not infected at } t) \\ &= \textstyle\sum_{k=1}^{t} \left(I(k,\ t)\frac{S(k,\tau)}{S(k,t)}\right) + 1 - B(t).\end{aligned} \tag{6}$$

Probabilities assigned to child survival:

Let $S_c^-(j)$ ($S_c^+(j)$) be the probability that a child born to an HIV negative (positive) woman survives to age j. The probability a woman assigns at period t for a child of age a to survive to the following period depends on the probability she assigns to having been infected at the time of the child's birth and is given by

$$\begin{aligned}\pi_c^a(t,\ t+1) &= \Pr(\text{mother was HIV-positive at birth}) \\ &\quad \times \Pr(\text{child alive at } t+1 \mid \text{mother was HIV-pos at birth } \&\ \text{ child alive at } t) \\ &\quad + \Pr(\text{mother was HIV-negative at birth}) \\ &\quad \times \Pr(\text{child alive at } t+1 \mid \text{mother was HIV-neg at birth } \&\ \text{ child alive at } t) \\ &= \left(\textstyle\sum_{k=1}^{t-a} I(k,\ t)\right)\frac{S_c^+(a+1)}{S_c^+(a)} + \left(1 - \textstyle\sum_{k=1}^{t-a} I(k,\ t)\right)\frac{S_c^-(a+1)}{S_c^-(a)}.\end{aligned} \tag{7}$$

2.5.2 Updating Survival Expectations with HIV Testing

HIV testing provides a woman information about her infection status at the time of the test; however, it does not provide her with any new information about when she might have gotten infected. I assume women do not use the test result to update their perceived infection

hazard process.

Given that some of the women who received positive test results assigned some likelihood to not being infected after the test, I assume that women assign a probability to the accuracy of the test result. Specifically, she assigns probability p_{test} to the test result being her actual status and probability $1 - p_{test}$ to the test result being uninformative. Therefore, the probability a woman assigns to being HIV-positive when receiving a test result at period t_{test}, denoted as $\hat{B}(t_{test})$, is

$$\hat{B}(t_{test}) = (1 - p_{test})\, B(t_{test}) + p_{test} \mathbf{1}\{\text{positive test result}\},$$

where $B(t_{test})$ is the probability assigned at period t_{test} to being HIV-positive without having been tested.

To compute the subjective life expectancies in periods after t_{test}, it is necessary to recover the post-test probabilities women assign to infection in each period. I define $\hat{P}(t)$ to be the post-test update of $P(t)$ (defined in Equation 3). These probabilities can be recovered using the fact that a test result does not provide the woman with any new information about when she might have gotten infected. Let $G(\tau,\ t_{test})$ be the probability assigned to having gotten infected at τ, conditional on being HIV-positive at t_{test}. This probability does not change after learning the test result. The after-test probability assigned at t_{test} to having gotten infected in a past period τ can be written as

$$\begin{aligned} \hat{I}(\tau,\ t_{test}) &= \Pr(\text{got infected at } \tau | \text{ HIV-positive at } t) \Pr(\text{HIV-positive at } t) \\ &= G(\tau,\ t_{test})\, \hat{B}(t_{test}) = \tfrac{I(\tau,\ t)}{B(t)} \hat{B}(t_{test}), \quad \tau = 1, ..., t_{test}. \end{aligned} \tag{8}$$

In addition, similar to Equation 4, the after-test probability assigned at t_{test} to having gotten infected in a past period τ is also given by

$$\hat{I}(\tau,\ t_{test}) = \frac{\hat{P}(\tau)\, S(\tau,\ t)}{1 - \sum_{k=1}^{t} \hat{P}(k)\,(1 - S(k,\ t))}, \quad \tau = 1, ..., t_{test}. \tag{9}$$

By (8) and (9), I get

$$\frac{\hat{P}(\tau)\, S(\tau,\, t)}{1-\sum_{k=1}^{t}\hat{P}(k)\,(1-S(k,\, t))} = \frac{I(\tau,\, t)}{B(t)}\hat{B}\,(t_{test}), \quad \tau = 1, ..., t_{test}. \tag{10}$$

I can solve for $\hat{P}(1), ..., \hat{P}(t_{test})$ by solving the system of t_{test} equations with t_{test} unknowns (the solution of the system of equations is presented in the appendix). The probability assigned to infection at a future period t is given by

$$\hat{P}(t) = h(t)\prod_{k=1}^{t-1}\left(1-\hat{P}(k)\right), \quad t > t_{test}.$$

Given the vector $\hat{P} = \left(\hat{P}(1), ..., \hat{P}(T-1)\right)$, the perceived probabilities of infection and survival in periods following the test are constructed as in equations (4) to (7).

2.6 Model Solution

The woman's problem can be formulated as a discrete-choice discrete-time stochastic dynamic program. Let $\Omega(t)$ be the state space at time t, consisting of all of the information relevant to decision-making that the woman has available at that time. Specifically, it contains the realized preference shock, the number of living children, ages of young children at risk of dying, whether she was pregnant in the previous period, and her testing history (period and result of each test). It also contains her fixed characteristics: her region of residence, schooling level, age of marriage, polygamy status of her husband, and the land owned by the household. Note that life horizon expectations are fully determined by $\Omega(t)$; therefore, there is no need to include the expectations in the state space.

Let $V\,(\Omega(t),\, t)$ be the value function, that is, the maximized present discounted value of lifetime utility. Let $V^{f}\,(\Omega(t),\, t)$ be the alternative-specific value function, that is, the value if choice f is taken, with f indicating pregnancy status. The Bellman equation of the

optimization problem is

$$V\left(\Omega(t),\ t\right) = \begin{cases} \max\left[V^0\left(\Omega(t),\ t\right),\ V^1\left(\Omega(t),\ t\right)\right], & t = 1, ..., F-1 \\ \mathrm{E}U^0\left(t,\ \Omega(t)\right) + \delta\pi\left(t,\ t+1\right)\mathrm{E}\left(V\left(\Omega(t+1),\ t+1 \mid p(t)=0,\ \Omega(t)\right)\right), & t = F, ..., T-1 \\ \mathrm{E}U^0\left(T,\ \Omega(T)\right), & t = T \end{cases}$$

$$V^f\left(\Omega(t),\ t\right) = \mathrm{E}U^f\left(t,\ \Omega(t)\right) + \delta\pi\left(t,\ t+1\right)\mathrm{E}\left(V\left(\Omega(t+1),\ t+1 \mid p(t)=f,\ \Omega(t)\right)\right), \quad f = 0, 1, \quad t = 1, ..., F-1,$$

where $U^1\left(t,\ \Omega(t)\right)$ represents a utility flow at state $\Omega(t)$ with pregnancy, and $U^0\left(t,\ \Omega(t)\right)$ represents the utility flow without pregnancy. The expectation associated with flow utility is taken over the present income shock. The expectation associated with the next-period value function is taken over future income and preference shocks as well as over child survival.

$V^1\left(\Omega(t),\ t\right)$ is strictly increasing in $\epsilon_p(t)$; however, $V^0\left(\Omega(t),\ t\right)$ is constant in $\epsilon_p(t)$ because the preference shock enters the utility flow only if the woman becomes pregnant. Let $\Omega^d\left(t\right)$ be the set of deterministic elements of the state space, that is, the set without $\epsilon_p(t)$. For any $\Omega^d(t)$ there is a unique critical value $\epsilon^*(\Omega^d(t),\ t)$ such that $V^1\left(\Omega(t),\ t\right) = V^0\left(\Omega(t),\ t\right)$. The solution to the woman's optimization problem is to become pregnant only if the preference shock is bigger than the corresponding critical value.

2.7 HIV/AIDS and Fertility Outcomes

There are several channels through which the presence of the HIV/AIDS epidemic can affect fertility outcomes in the framework presented above. These channels are driven by both actual and perceived exposure to infection. They operate simultaneously and can potentially have opposing influences on the timing and number of births as well as on the experienced child mortality. In terms of the effect of actual HIV infection, the shorter life-span of women's

life-cycle will reduce the number of periods during which they can get pregnant and will therefore have an obvious negative effect on the overall average number of births. The increase in child mortality, on the other hand, can increase fertility through a replacement behavior. If there is a decreasing marginal utility from an additional child, the loss of a child will increase fertility in the periods after the loss.

The impact of the subjective beliefs on fertility outcomes is even more ambiguous. As described in the previous subsection, a woman compares the expected value of lifetime utility with and without a current pregnancy. With higher likelihood assigned to being infected, a woman assigns less likelihood to her and her child's survival and therefore perceives fewer periods during which she would get utility flows from having the additional child. Given that the non-pecuniary cost of pregnancy is not affected by beliefs about HIV status, this implies that the woman would be less likely to get pregnant if she assigns higher probability to being infected. On the other hand, if a woman has high valuation for children, she might choose to increase her fertility when she assigns some probability to being infected. She would do so in expectation that a larger share of her children will not survive. In addition, even without a change in the total number of births a woman gives, the presence of HIV might change the timing of her pregnancies. A woman might assign no or small likelihood to being infected in the present but decide to give birth at younger ages if she perceives a higher likelihood to being infected in the future. This will depend on her perceived hazard rates throughout the fertile stage of her life-cycle.

3 Data

3.1 Malawi Diffusion and Ideational Change Project

Malawi is a landlocked country located in Southeast Africa. It has a population of about 15 million, comprised of different ethnic and religious groups. 81 percent of the population

lives in rural areas and relies mostly on subsistence agriculture.[14] The Malawi Diffusion and Ideational Change Project (MDICP) data have been collected since 1998 in rural areas of three districts in Malawi: Balaka in the south of the country, Mchinji in the center, and Rumphi in the north.[15] The different rounds of the longitudinal dataset contain extensive information on more than 4,000 men and women, at the individual, household and community levels.

A unique feature of the MDICP dataset is that it includes measures of subjective likelihoods respondents assign to being HIV-positive. The subjective expectations were collected using an elicitation methodology developed by Delavande and Kohler (2009) for a developing country context with low levels of literacy and numeracy.[16] Respondents were provided with ten beans and a plate. They were asked to allocate different number of beans on the plate to express the likelihood that different events will be realized. The respondents were instructed that zero beans reflect certainty that an event will not happen, more beans reflect higher likelihood that an event happens, and that ten beans imply certainty about the event happening. The likelihood assigned to being HIV-positive is measured by asking: "Pick the number of beans that reflects how likely you think it is that you are infected with HIV/AIDS now." Delavande and Kohler (2009) find that reported subjective expectations follow basic properties of probabilities and that the assessments of HIV-infection vary meaningfully with observable characteristics associated with different levels of HIV prevalence.

Since 2004, HIV testing has been offered to all respondents during data collection. The take-up rates of the tests were high, above 90 percent in all waves. The HIV prevalence rate in the sample was 6.9 percent Obare et al. (2009). In 2004, test results were available five to seven weeks after testing. The testing component of the survey was linked to an experiment that is described and analyzed by Thornton (2008). Respondents were assigned vouchers

[14] data.worldbank.org

[15] Detailed information on the Malawi Diffusion and Ideational Change Project can be obtained at http://www.malawi.pop.upenn.edu/.

[16] Attanasio (2009) and Delavande et al. (2010) review existing subjective expectations data from developing countries.

for a monetary reward, redeemable upon return to temporary Voluntary Consulting and Testing (VCT) sites where results where provided. The VCT sites were set up such that all respondents' homes are within five kilometers distance from at least one site. Approximately 70 percent of those tested chose to pick up their results. In 2006 and 2008, rapid blood tests were adopted, eliminating the time delay between testing and provision of results.

For my analysis, I am using a subsample of married women in their first marriage, who did not become pregnant from a relationship with men other than their future husbands prior to marriage. I restrict the sample to women in unbroken first marriages because I abstract from modeling any decisions related to marriage or partnership. I also restrict the sample to include only women who were interviewed in at least one of the 2006 and 2008 rounds. Data collected in these rounds include family rosters, containing information on all children of respondents, and the elicitation of subjective expectations. I exclude women who were born before 1960, because the last age of fertility in the model is assumed to be 44 and 2004 is the earliest year included for fertility outcomes. After excluding additional women for missing information, the estimation sample consists of 1006 women.

Tables 1 to 4 and Figures 1 and 2 provide descriptive statistics for the variables used in the analysis. The average age of women in the sample is 26.6 in 2004. 29 percent of the sample resides in Balaka (south), 35 percent in Mchinji (center), and 36 percent in Rumphi (north). The median number of years of schooling is 5. The women in Balaka have the lowest schooling levels with 31 percent of women never having attended school and only three percent having attended some secondary school. The women in Rumphi have the highest schooling levels, with 99 percent of women ever having attended school and 26 percent some secondary school. The average age of marriage in the overall sample is 17.5 and is similar across regions. Polygamy is most prevalent in Rumphi, with 33 percent of the women in the sample from that region married to a polygamous husband. The share of polygamous women is 27 percent in Mchinji and 20 percent in Balaka. As shown in Table 2, HIV testing take-up rates were high in all the rounds: 87.9 percent in 2004, 93.3 percent

in 2006, and 94.6 percent in 2008. The percentage of women who tested positive was 2.9 percent in 2004, 3.1 percent in 2006, and 3.7 percent in 2008.

Figure 1 depicts the distribution of the likelihood women assigned to being HIV-infected in the 2006 and 2008 rounds (pooled), measured by beans on a scale of zero to ten. Fifty percent of the reports were of zero beans. The percentage of women who chose each category decreases with the number of beans, except for the 5-bean category. Fewer than one percent chose the 10-bean category. Table 3 shows the average number of beans. The overall sample average is 1.52 beans. The average number of beans chosen in each region conforms to the ranking of HIV prevalence in the general MDICP sample. The averages are 1.78 beans in Balaka, 1.69 in Mchinji, and 1.16 in Rumphi. The 2004 HIV prevalence rates for these regions are 7.9 percent in Balaka, 6.4 percent in Mchinji, and 4.4 percent in Rumphi. The table also shows that average beliefs decrease with schooling and are higher for women married to a polygamous husband.

Figure 2 depicts the distribution of the likelihood women assigned to being HIV-infected after receiving HIV test results. Because the testing component of the survey was conducted after the interview components, these are beliefs reported two years after the tests (the subsequent round of data collection). Because of low HIV prevalence and high attrition of infected women, there are only 18 women who report their beliefs two years after receiving a positive test result. The average number of beans allocated by these women is 4.67 beans, which is significantly higher then the average of 1.5 beans reported by women who received a negative test result. Most of the women who received a positive test result assign at least some likelihood to not being HIV infected.

I obtain information about births from family rosters that were collected in 2006 and 2008. Respondents were asked to list all of the children ever born to them, but year of birth is missing for children who died more than two years before the interview. I therefore use only birth outcomes reported for the years 2004 to 2007. Depending on the year of marriage, year of birth, and the rounds in which the woman participated, I observe between

1 and 4 potential fertility years for women of ages 16 to 45. In total, I observe 3,148 years of potential fertility with births in 920 of them (29.2 percent). Table 4 shows the annual birth probabilities of married women by 5-year age groups. The annual birth probabilities decrease from 0.405 for ages 16 to 20 to 0.06 for ages 41 to 45.[17]

3.2 Malawi Second Integrated Household Survey

The MDICP data do not include detailed measures of household consumption. For this reason, I use the Malawi Second Integrated Household Survey (IHS-2) to impute a better measure of consumption of households in the MDICP sample. The dataset, gathered by the Malawi National Statistics Office in 2004-2005, is part of the World Bank's Living Standards Measurement Study program. It includes comprehensive data on consumption and expenditures of households, as well as on local prices. Specifically, the dataset includes a measure of annual consumption expenditure aggregates in real value. The survey was fielded in 26 out of Malawi's 27 districts, including all three MDICP districts. After restricting the sample to households in the districts covered by MDICP and excluding households without a woman as a head of the household or as a wife of the head, the estimation sample includes 612 households. 220 of these households are from Balaka, 188 from Mchinji, and 204 from Rumphi.[18]

3.3 Mortality Statistics

I supplement the use of the MDICP data with statistics about survival after infection and child mortality from other studies that use data from samples with repeated HIV testing and frequent follow-ups. These data provide more precise information on times of infection and death than MDICP does. Hallett et al. (2008) estimate probabilities of survival after infection by fitting a Weibull distribution to survival data presented by Todd et al. (2007)

[17] The annual birth probability is defined as the proportion of alive and married women of specific age that give birth.

[18] Detailed information on the Malawi Second Integrated Household Survey can be obtained from the World Bank's website at http://econ.worldbank.org.

from 5 studies in Eastern and Southern Africa before highly active antiviral therapy. The probability of survival to year t conditional on getting infected at year τ is estimated as

$$S(\tau,\ t) = \exp\left(-\left[\frac{t-\tau}{\psi_\tau}\right]^2\right),$$

with ψ_τ reducing with age of infection. (Parameter estimates and median survival time are presented in Table 5.) The mortality hazard increases with both duration and age of infection. The median survival time of an individual infected at ages 15 to 19 is 13.3 years, while it is 8.4 years for an individual who gets infected at ages 40 to 44.

Data from different longitudinal studies with repeated assessments of HIV status of adults show higher mortality rates of children born to infected women. In these studies, the HIV status of the children was generally not available. Controlling for different background characteristics, mortality rates of children born to HIV-infected mothers were estimated to be about three times higher than those for children born to uninfected mothers, with the effect lasting throughout childhood years (Newell et al., 2004). Crampin et al. (2003) report child mortality rates by status of mother at birth from a retrospective cohort study with more than 10 years of follow-up in Karonga district in Northern Malawi. The rates are reported in Table 6.

4 Estimation

The main estimation sample contains data on 1006 women from the MDICP dataset. The information on the ith woman consists of up to 4 years of pregnancy choices $\left(p_i(t),\ t = \underline{t_{pi}}, ..., \overline{t_{pi}}\right)$, up to 2 reports of subjective assessments of HIV status $\left(b_i(t),\ t = \underline{t_{bi}}, ..., \overline{t_{bi}}\right)$, up to 3 HIV test results (collected in the 2004, 2006 and 2008 rounds), and a vector of fixed characteristics: region of residence, schooling level, year of birth, age of marriage, polygamy status of husband, and household's land.

I also use an auxiliary sample containing data on 612 households from the IHS-2 dataset.

The information on the jth household consists of household annual aggregate consumption expenditure (y_j), and a vector of household characteristics: region of residence, age of woman (head or wife of the head of the household), number of children, and household's land.

The first step of the econometric implementation involves estimating the parameters of the income function (Equation 1) by ordinary least squares regression using the auxiliary sample. The rest of the model parameters are estimated using maximum likelihood, taking the parameters of the income and survival functions as given. The likelihood function contains the following elements: (1) belief reports probabilities; (2) fertility outcomes probabilities; (3) actual HIV status and survival; and (4) unobserved type probabilities. I proceed by describing the contribution of each of these elements to the likelihood.

4.1 Beliefs

The perceived infection hazard is estimated using the data on subjective assessments of HIV-status. Let $B_i(t)$ be woman i's perceived probability of being HIV-positive at period t and $b_i(t)$ be the number of beans she allocates to being HIV-infected in the expectations elicitation exercise. I make the following two assumptions regarding the relationship between a woman's belief and the recorded number of beans. First, I assume that beliefs are reported with some noise. Specifically, the reporting error, $\epsilon_{bi}(t)$, is assumed to be iid across time and women:

$$\epsilon_{bi}(t) \sim iidN\left(0, \sigma_b\right).$$

Second, I assume that each discrete "bean category" corresponds to a probability interval.[19] A respondent reports the number of beans corresponding to the interval in which her belief added to the reporting error falls. The reports are assumed to be made according to the

[19] As in Delavande and Kohler (2009)

following rule:[20]

$$b_i(t) = 0, 1 \quad \text{if } B_i(t) + \epsilon_{bi}(t) \leq 0.15,$$

$$b_i(t) = 2 \quad \text{if } 0.15 < B_i(t) + \epsilon_{bi}(t) \leq 0.25,$$

$$\vdots$$

$$b_i(t) = 10 \quad \text{if } 0.95 < B_i(t) + \epsilon_{bi}(t).$$

Let $\underline{\Omega}_i^d$ be the set of initial conditions for woman i. It contains the deterministic (and observable) elements of the state space. That is, it contains her permanent characteristics that enter the perceived hazard function. I define it to also contain the woman's testing variables (timing and results of tests). Although the woman does not forecast getting tested, I treat this information as part of the initial conditions for the econometric implementation. The woman's sequence of life-cycle beliefs are determined given these initial conditions and her unobserved type, and can be written as $B\left(t \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right)$. The probability of observing $b_i(t)$, conditional on the set of woman i's initial conditions and her unobserved type is given by

$$\Pr\left(b_i(t) = 0, 1 \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right) = \Phi\left(\frac{0.15 - B\left(t \mid \underline{\Omega}_i^d, \text{type}_i = j\right)}{\sigma_b}\right),$$

$$\Pr\left(b_i(t) = 2 \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right) = \Phi\left(\frac{0.25 - B\left(t \mid \underline{\Omega}_i^d, \text{type}_i = j\right)}{\sigma_b}\right) - \Phi\left(\frac{0.15 - B\left(t \mid \underline{\Omega}_i^d, \text{type}_i = j\right)}{\sigma_b}\right),$$

$$\vdots$$

$$\Pr\left(b_i(t) = 10 \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right) = 1 - \Phi\left(\frac{0.95 - B\left(t \mid \underline{\Omega}_i^d, \text{type}_i = j\right)}{\sigma_b}\right),$$

where Φ is the cumulative distribution function of the the standard normal distribution.

As shown in Figure 1, the percentage of women reporting each bean category generally decreases with the number of beans. The percentage of women who report five beans, 8.78 percent , is higher than the two categories below (3.41 percent report four beans and 6.52

[20] Figure 3 shows the distributions of reported beliefs in the 2006 and 2008 rounds. There is a big drop in the percentage of women choosing the 0-bean category and the number of women who choose the 1-bean category doubles. My model would not be able to generate this shift. For my empirical analysis, I treat the 0 and 1-bean categories as a single category.

percent report three) and is also significantly higher than the next category (1.59 percent report six). My model is not likely to capture this pattern. I therefore assume that the probability interval corresponding to the five bean category is larger, implying that some of the women who would allocate 4 or 6 beans to the likelihood of being HIV-positive under the rule described above, report instead 5 beans. The probability of observing 5 beans is assumed to be

$$\Pr\left(b(t)=5 \mid \underline{\Omega}_i^d,\ \mu_i=j\right)=\Phi\left(\frac{0.65-B\left(t \mid \underline{\Omega}_i^d,\ type_i=j\right)}{\sigma_b}\right)-\Phi\left(\frac{0.35-B\left(t \mid \underline{\Omega}_i^d,\ type_i=j\right)}{\sigma_b}\right).$$

4.2 Fertility

The parameters of the utility function are estimated using the data on pregnancy outcomes. As described in Section 3.6, the solution to a woman's optimization problem is to become pregnant at period t if the preference shock, $\epsilon_{pi}(t)$, is bigger than the critical value $\epsilon^*(\Omega_i^d(t),\ t,\ \text{type}_i)$. Conditional on the woman's type and the observable elements of the woman's state space at time t, the probability of the woman's choice to become pregnant is given by

$$\Pr\left(p_i(t)=1 \mid \Omega_i^d(t),\ \text{type}_i=j\right)=1-\Phi\left(\frac{\epsilon^*(\Omega_i^d(t),\ t,\ \text{type}_i)}{\sigma_p}\right),$$

$$\Pr\left(p_i(t)=0 \mid \Omega_i^d(t),\ \text{type}_i=j\right)=\Phi\left(\frac{\epsilon^*(\Omega_i^d(t),\ t,\ \text{type}_i)}{\sigma_p}\right).$$

Conditional on the set of initial conditions and the unobserved type, the probability of observing a sequence of belief reports is independent from the probability of observing a sequence of fertility outcomes. This is because the set of initial conditions and type map deterministically into the sequence of life-cycle beliefs.

4.3 Actual Infection Process

Given the assumption of no symptoms, the state space does not contain actual HIV status beyond any HIV test results that women have received. It is therefore not necessary to recover the actual hazard process for estimation of the decision-model parameters. For each possible

state, the estimated model generates a probability of becoming pregnant conditional on being alive at that state. I do need an estimate of the actual hazard process for my counterfactual analysis, however. To study the effect of different policy interventions on outcomes of life-cycle fertility and child mortality, I include the infection and mortality processes in the simulations.

I assume an actual HIV infection hazard rate with a functional form similar to that of the perceived infection hazard described in equation (2). Information about the hazard process is contained in the HIV test results, the age in which the tests were taken, and the ages in which a woman is last observed (regardless of testing histories). I define H_i to be a vector of a woman i's testing and survival information: the oldest age in which she had a negative test result (if ever), the earliest age at which she received a positive test result (if ever), and the latest age in which she is observed in the data. The probability of observing a woman with testing survival history H_i, conditional on her type and observable elements of her initial state is given by

$$\Pr\left(H_i \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right).$$

I present the details of how this probability is computed in the appendix.

4.4 Type Distribution

A multinomial logit specification is used for the type probabilities. The probability that woman i is of type j is given by

$$\Pr\left(\text{type}_i = j \mid \underline{\Omega}_i^d\right) = \frac{\exp\left(w_{\underline{ti}}'\gamma_j\right)}{1+\sum_{k=1}^{J}\exp\left(w_{\underline{ti}}'\gamma_k\right)}, \quad j = 1, ..., J,$$

$$\Pr\left(\text{type}_i = 0 \mid \underline{\Omega}_i^d\right) = \frac{1}{1+\sum_{k=1}^{J}\exp\left(w_{\underline{ti}}'\gamma_k\right)},$$

where $w_{\underline{ti}}$ is a vector of woman's initial conditions, including region of residence, schooling level, age of marriage, whether she is married to a polygamous husband, year of birth, and the number of children she had in the first period observed interacted with the age she was

when first observed. The last term is included to take into account the fact that not all the women are observed from the time of their marriage. The state in which they are first observed depends on past decisions and therefore on their unobserved type.

4.5 Likelihood Function

The contribution of woman i to the sample likelihood is given by

$$\mathcal{L}_i = \sum_{j=0}^{J} \left(\prod_{t=\underline{t_p}}^{\overline{t_p}} \Pr\left(p_i(t) \mid \Omega_i^d(t),\ \text{type}_i = j\right) \right) \left(\prod_{t=\underline{t_b}}^{\overline{t_b}} \Pr\left(b_i(t) \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right) \right)$$

$$\times \left(\Pr\left(H_i \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right) \right) \Pr\left(\text{type} = j \mid \underline{\Omega}_i^d\right).$$

5 Estimation Results

5.1 Parameter Estimates

The model is fit with four unobserved types.[21] Recall that the types can differ with respect to their preferences for children, the assigned probabilities to the accuracy of HIV test results, and the perceived and actual exposure to HIV infection. The four types have distinctively different beliefs and characteristics. As seen in Table 12, type 0 women, who represent 28 percent of the sample, have the largest share of women with some secondary education and the lowest share of women married to polygamist men. They assign no likelihood to being infected throughout their life cycle. Type 1s, comprising 23 percent of the sample, have the highest share of women with no schooling and the youngest age of marriage. Type 2s, comprising only 2.1 percent of the sample, perceive the highest exposure to HIV infection. On average, women of that group assign probabilities of 0.59 and 0.69 to being infected when

[21] There were significant improvements in model fit beyond three types. I did not attempt to fit the model with more types because of computational burden.

they are 20 and 40 years old respectively. Type 3 women are the biggest group, representing 47 percent of the sample.

Tables 7 to 11 report the parameter estimates and their standard errors. As can be seen in Table 7, the marginal utilities of additional children are positive for the first child and are decreasing with the stock of surviving children for all women. However, the profile of these marginal utilities varies with region of residence, schooling level, polygamy status and unobserved type. Relative to women with lower levels of education, women who attended secondary school have the highest marginal utility for the first child but the marginal utility declines at the fastest rate with the number of children. The same is true for women from Rumphi district relative to women from the other two districts and for type 1 women relative to the other types.

The childbearing costs, which are assumed to depend only on a woman's age, are estimated to be positive and increasing with age. The cost for the youngest age group (less than 20 years) is 25 percent of that for women of ages 20 to 24 and only about one percent of the cost for women of ages 40 to 44. The cost of consecutive pregnancies, on the other hand, is estimated to be decreasing with age.

5.2 Model Fit

To assess model fit, I compare the model's prediction of the distributions of reported beliefs and pregnancy probabilities to the distributions of actual beliefs and pregnancies observed in the data. I simulate each observed woman 100 times. The simulation starts from the age a woman is first observed and takes as given the observable elements of the state space (her constant characteristics, pregnancy in previous period, and the number and ages of children younger than four years). For each simulation, I draw an unobservable type from the type distribution and preferences shocks.

Figures 4 and 5 and Table 13 compare the fertility and belief reporting outcomes predicted by the model to the actual outcomes observed in the data. Figure 4 shows that the model

is able to generate the shape of the reported beliefs distribution. It under-predicts the proportion of women in the lowest bean category by about 4 percentage points and over-predicts the proportions in the 2 and 3-bean categories. As described in section 4, I assume that some of the respondents round and report five beans instead of four and six. 13.8 percent of belief reports in the data are of the four to six-bean categories; the model predicts 13.6 percent.

Figure 5 depicts the actual and predicted annual pregnancy probabilities for different age groups. The model captures the decline in pregnancy probabilities with age. This pattern is generated by the diminishing marginal utility of additional children and childbearing costs, which increase with age. Table 13 shows the actual and predicted annual pregnancy probabilities by region, schooling level, and polygamy status of the husband.

6 Counterfactual Analysis

Having estimated the structural model parameters, I use the model to perform counterfactual experiments to quantify the effect of the HIV/AIDS epidemic on fertility outcomes in the given environment and to evaluate the impacts of different policy interventions. I do so by simulating the women's life-cycle fertility decisions in different environments. As with the previous simulations, performed to assess goodness of fit, each woman is simulated 100 times. However, in the counterfactual simulations I also incorporate infection and mortality. That is, women are exposed to HIV infection according to the estimated actual infection hazard process, and, once infected, they are exposed to the mortality process. Survival probabilities of children of ages zero to three depend on the HIV status of the mother at time of birth according to the rates in Table 6.

For the counterfactual simulations, I use a younger subsample of the 509 women who were born after 1978. I exclude the older women to have the composition of the sample, in terms of characteristics and types, minimally affected by survival to the time in which women are

first observed. The characteristics of the sample used in the counterfactuals is presented in Table 14. I define the baseline environment to be with the levels of perceived and actual hazards of infection estimated by the model, but without any HIV tests. Results of the baseline environment simulation are presented in Panel A of Table15. The table includes the average beliefs and infection rates at different points of the women's life-cycle, number of life-cycle births as well as the average number of child mortality experienced by a woman. The results are shown for the full sample as well as by type.

To assess how fertility outcomes would have been different were there no HIV, I simulate fertility in an environment with no HIV (and no beliefs about HIV). These results are presented in Panel B of Table 15. Comparison of fertility outcomes in the two environments indicates that HIV has a negative average effect on fertility for the given group of women. The average number of births during a woman's life in the no-HIV environment is 7.22 in comparison to 7.07 in the baseline environment.[22] Although the total amount of births is lower in the baseline environment, the total incidents of child mortality is higher by seven percent as the percentage of children not surviving to age beyond age 4 increases to 17.5 percent from 15.8 percent in the presence of HIV.

Eighty-five percent of the women will experience the same number of births in both environments, three percent will have higher fertility in the HIV environment and twelve percent will experience fewer pregnancies. The women who have higher fertility in the presence of HIV are of two groups. The first group includes women who are infected with HIV but assign relatively low probability to being infected. They respond to experienced child mortality (due to mother to child transmission) by having additional pregnancies. The second group is of women who have high valuation of children and would increase fertility early in their life-cycle in anticipation that some of their children might die. However, the overall average reduction in fertility in the presence of HIV is largely due to the shorter life-span of women infected with HIV as well as women that choose to reduce fertility as

[22]The life-cycle number of births is higher than the national total fertility rate estimated at 6 pregnancies per woman because fertility rates are higher in rural areas of Malawi.

they assign lower probabilities to their own as well as their children survival in the future.

There is heterogeneity in the effect of HIV on fertility by women's characteristics. The average difference for type 0 women is only 0.03. Because they do not assign any likelihood to being infected, it is completely due to the increase in mother and child mortality by infected women. Type 1s experience an average of 0.17 fewer pregnancies in the presence of HIV although this group has the highest share of women (8 percent) who actually increase fertility in that environment. Type 3s are the most affected, with a difference of 0.2 pregnancies. Women from Balaka, with the highest HIV prevalence and highest average likelihood women assign to being infected, will have on average 0.31 more births in the no-HIV environment while women from Mchinji and Rumphi will have on average 0.08 and 0.1 more births respectively.

The fertility outcomes are affected by perceived as well as actual exposure to HIV. As expected, the difference in the number of births in the two environments depends not only on whether a woman got infected during the fertile stage of her life cycle but also on when in the life cycle she got infected. Women who got infected by age 25 (5.1 percent of the sample) experience 1.57 less births in the baseline environment relative to the no-HIV environment. Women who got infected by age 45 (11.6 percent of the sample) experience a reduction of 0.84 births and women who did not get infected by the end of their fertile stage, experience an average difference of 0.06 births.

To separate the effect of women's shorter life-span on the number of births from the other channels through which HIV can affect fertility (child mortality and survival expectations), I count pregnancies in the no-HIV environment which occurred only in periods in which women would have been alive in the baseline environment. This brings the average number of births down to 7.1, only 0.03 higher than the number in the baseline environment. During these periods, women of types 0 and 1 experience fewer births in the no-HIV environment, by 0.01 and 0.08 respectively. Type 2s and 3s experience more births. Type 2s, who have the highest beliefs about being infected, experience on average 0.19 more births and Type 3s,

the largest group in the sample, experience 0.1 more births. Regardless of whether women decrease or increase the number of births in the presence of HIV, most of the divergence in fertility rates happens in the second half of the women's fertile stage of the life-cycle. This is because older women assign higher likelihood to being infected, they are more likely to actually be infected, and the marginal utility from additional child reduces with the number of children.

In Panel C of Table 15 I present results from a simulation of an environment in which there is no mother-to-child transmission. This simulation can be thought of as an assessment of the potential effect of provision of antiviral treatment that prevents such transmission.[23] In the simulation, child mortality probabilities of children born to HIV-infected women are equated to those of children born to HIV-negative women. The average number of life-cycle births in this environment is 7.05, a slight difference from the 7.07 births experienced in the baseline environment. Although the reduction in expected and experienced child mortality, less than one percent of women will change the number of births they give.

To study the impact of HIV testing on fertility outcomes, I simulate life-cycle fertility in counterfactual scenarios in which the respondents are offered a single test in different points in their life cycle: at the time of their marriage (age 17 on average), age 25 and age 35. The impact of testing on outcomes depends on how women update their beliefs after learning their test result. This updating depends on the discrepancy between the beliefs and actual HIV status and on the accuracy women assign to test results. The extent to which testing will affect outcomes is expected to be limited given the estimated low probabilities most women assign to the accuracy of test results (presented in Table10). In addition, the impact of testing also depends on the extent to which the updated beliefs alter the relative valuation of women's choices. This can vary with age and the number of children a woman has as life horizon shortens and marginal utility from additional children reduced. It is also important to keep in mind that while the test result gives information about one's status

[23] In high-income countries, the rate of mother-to-child transmission has been reduced to less than one percent (unaids.org).

in the present, women's decisions are also affected by perceived exposure to infection in the future. A woman receiving a negative test result, for example, might not change by much her survival expectations if she believes she is in high risk of getting infected in the periods just after the test.

The results of simulating testing at age of marriage and ages 25 and 35 are presented in Panels D, E, and F respectively. 1.3 percent of women are simulated to be infected at the time of their marriage, 4.1 percent at age 25 and 5.7 percent at age 35. At ages 25 and 35, women have already given 3.3 and 6.1 births on average. All three simulations result in a negligible effect on the average number of births and child mortality women experience over their life-cycle. As expected, type 0s and 1s, who assign probabilities of zero and 0.03 to the accuracy of the test result, do not alter their choices. Type 2s all receive negative test results. However, given the probability of 0.21 they assign to the accuracy of the test results and their high perceived risk of infection, the total average number of births they give does not change. Type 3s, who assign the highest probability of 0.76 to the accuracy of the test results, reduce their fertility after receiving a positive test result. The group that received a positive test result at the time of marriage reduced the average number of births by 0.42, from 4.98 to 4.56. The women of type 3 who receive a positive test result at ages 25 and 35 have 0.39 and 0.21 fewer births. The smaller effect at age 35, relative to that when tests were given in earlier ages, is partially due to the fact that mortality from HIV accelerates with age, leaving the women fewer periods during which to make fertility choices.

Finally, in Panel G of Table 15 I present a simulation of a scenario in which women are offered a test at age 25. Unlike the former testing simulations, women perceive the test result as accurate and fully update their beliefs after receiving a test result. There is a negligible effect on the overall average number of births and experienced child mortality. There is very little impact on the fertility behavior of women who receive a negative test result. Women who receive a positive test result, however, see an average reduction of 0.19 in the number of births. HIV-positive women of type 0 and 1, who do not reduce fertility in the results

presented in Panel E, have 0.07 and 0.05 less births in this environment. Type 3s who receive a positive test result have 0.57 fewer births than in the baseline environment.

7 Conclusion

In this paper, I specified and structurally estimated a dynamic model of fertility in an environment with high HIV prevalence. My analysis takes into account uncertainty about HIV status and the discrepancies between perceived and actual exposure to HIV infection. Women's perceptions about their exposure to HIV infection and about their own HIV status affect their beliefs about their own and their children's life expectancy. Beliefs about life expectancies, in turn, affect fertility choices by changing the profile of expected lifetime utilities associated with each choice. I estimate the model parameters by maximum likelihood with longitudinal data from the Malawi Diffusion and Ideational Change Project, which contain measures of subjective beliefs about own HIV status. The model fits well the fertility patterns in the data, as well as the distribution of reported beliefs about own HIV status.

Model simulations are informative about how HIV affects fertility and about the impact of policies aimed at reducing HIV, such as HIV testing programs and provision of antiviral therapy, on fertility and child mortality. Results show that the presence of HIV reduces fertility, both for women who are infected and who are not infected. The presence of HIV reduces the average number of life-cycle births by 0.15. I also find that HIV testing is effective at reducing fertility of infected women, leading to a reduction in child mortality and orphan hood; however, the partial updating of beliefs following a test mitigates this effect. I also find that prevention of mother-to-child transmission through dissemination of antiviral drugs would have a negligible effect on the average number of life-cycle births, although it reduces the incidence of child mortality.

References

Anglewicz, P., *Migration, risk perception, and HIV infection in Malawi*, Ph.D. thesis, University of Pennsylvania (2007).

Anglewicz, P. and H. Kohler, "Overestimating HIV infection: The construction and accuracy of subjective probabilities of HIV infection in rural Malawi," *Demographic research* 20 (2009), 65.

Attanasio, O., "Expectations and perceptions in developing countries: their measurement and their use," *The American Economic Review* (2009), 87–92.

Crampin, A., S. Floyd, J. Glynn, N. Madise, A. Nyondo, M. Khondowe, C. Njoka, H. Kanyongoloka, B. Ngwira, B. Zaba et al., "The long-term impact of HIV and orphanhood on the mortality and physical well-being of children in rural Malawi," *Aids* 17 (2003), 389.

Delavande, A., X. Giné and D. McKenzie, "Measuring subjective expectations in developing countries: a critical review and new evidence," *Journal of Development Economics* (2010).

Delavande, A. and H. Kohler, "Subjective Expectations in the Context of HIV/AIDS in Malawi," *Demographic research* 20 (2009), 817.

Durevall, D. and A. Lindskog, "Uncovering the impact of the HIV epidemic on fertility in Sub-Saharan Africa: the case of Malawi," *Journal of Population Economics* 24 (2011), 629–655.

Fink, G. and S. Linnemayr, "HIV Does Matter for Fertility: Human Capital, Mortality and Family Size," *Harvard Public School of Health, mimeo* (2009).

Fortson, J., "HIV/AIDS and fertility," *American Economic Journal: Applied Economics* 1 (2009), 170–94.

Gray, R., M. Wawer, R. Brookmeyer, N. Sewankambo, D. Serwadda, F. Wabwire-Mangen, T. Lutalo, X. Li, T. vancott and T. Quinn, "Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda," *The Lancet* 357 (2001), 1149–1153.

Hallett, T., B. Zaba, J. Todd, B. Lopman, W. Mwita, S. Biraro, S. Gregson and J. Boerma, "Estimating incidence from prevalence in generalised HIV epidemics: methods and validation," *PLoS Medicine* 5 (2008), e80.

Heckman, J. and R. Willis, "Estimation of a Stochastic Model of Reproduction: An Econometric Approach," (1974).

Joseph Hotz, V., J. Klerman and R. Willis, "The economics of fertility in developed countries," *Handbook of population and family economics* 1 (1997), 275–347.

Juhn, C., S. Kalemli-Ozcan and B. Turan, "HIV and fertility in Africa: First evidence from population based surveys," (2008).

Kalemli-Ozcan, S., "AIDS," Reversal" of the Demographic Transition and Economic Development: Evidence from Africa," (2006).

Mira, P., "UNCERTAIN INFANT MORTALITY, LEARNING, AND LIFE-CYCLE FERTILITY*," *International Economic Review* 48 (2007), 809–846.

Morgan, D., C. Mahe, B. Mayanja, J. Okongo, R. Lubega and J. Whitworth, "HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries?," *Aids* 16 (2002), 597.

Newell, M., H. Brahmbhatt and P. Ghys, "Child mortality and HIV infection in Africa: a review," *Aids* 18 (2004), S27.

Obare, F., P. Fleming, P. Anglewicz, R. Thornton, F. Martinson, A. Kapatuka, M. Poulin, S. Watkins and H. Kohler, "Acceptance of repeat population-

based voluntary counselling and testing for HIV in rural Malawi," *Sexually transmitted infections* 85 (2009), 139.

Powers, K., C. Poole, A. Pettifor and M. Cohen, "Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis," *The Lancet infectious diseases* 8 (2008), 553–563.

Sah, R., "The effects of child mortality changes on fertility choice and parental welfare," *Journal of Political Economy* (1991), 582–606.

Schultz, T., "Demand for children in low income countries," *Handbook of population and family economics* 1 (1997), 349–430.

Thornton, R., "The demand for, and impact of, learning HIV status," *The American economic review* (2008), 1829–1863.

Todd, J., J. Glynn, M. Marston, T. Lutalo, S. Biraro, W. Mwita, V. Suriyanon, R. Rangsin, K. Nelson, P. Sonnenberg et al., "Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy," *Aids* 21 (2007), S55.

Wolpin, K., "An estimable dynamic stochastic model of fertility and child mortality," *The Journal of Political Economy* (1984), 852–874.

———, "Determinants and Consequences of the Mortality and Health of Infants and Children," *Handbook of population and family economics* 1 (1997), 483–557.

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

———, "In sorrow to bring forth children: fertility amidst the plague of HIV," *Journal of economic growth* 12 (2007), 283–327.

A Tables and Figures

Table 1: Descriptive Statistics

Variable		All	Balaka (South)	Mchinji (Center)	Rumphi (North)
Region:	**Balaka**	0.29	-	-	-
	Mchinji	0.35	-	-	-
	Rumphi	0.36	-	-	-
Education:	**No school**	0.14	0.31	0.15	0.01
	Primary	0.73	0.66	0.8	0.73
	Secondary	0.12	0.03	0.05	0.26
Land>1 hectare		0.44	0.19	0.53	0.55
Polygamy		0.27	0.2	0.27	0.33
Age of Marriage		17.54	17.02	17.56	17.93
		(2.25)	(2.28)	(2.03)	(2.35)
Age 2004		26.63	25.99	25.75	28.01
		(8.09)	(8.44)	(7.21)	(8.45)
Number of observations		1006	355	288	363

Table 2: Test Take-up and Percentage Tested Positive by Year of Test

	2004		**2006**		**2008**	
	%	**N**	**%**	**N**	**%**	**N**
Took Test	87.9%	620[a]	93.28%	759[a]	94.6%	741[a]
Tested positive	2.94%	545[b]	3.11%	708[b]	3.71%	701[b]

[a] Number of women who were offered to take a HIV test for the given year

[b] Number of women tested for the given year

Table 3: Reported Belief about Own HIV Infection, Measured in Beans: 2006 and 2008 pooled

		Mean	sd	N
All		1.52	2.13	1640
Region	**Balaka**	1.78	2.16	462
	Mchinji	1.69	2.24	565
	Rumphi	1.16	1.96	613
Schooling	**No school**	1.77	2.2	230
	Primary	1.55	2.18	1214
	Secondary	1.06	1.67	196
Polygamy	**Mono**	1.35	1.35	1188
	Poly	1.95	2.44	452
Age	**$\leq$26**	1.41	2.02	658
	$>$26	1.59	2.21	982

Figure 1: Belief Distribution, 2006 and 2008 pooled

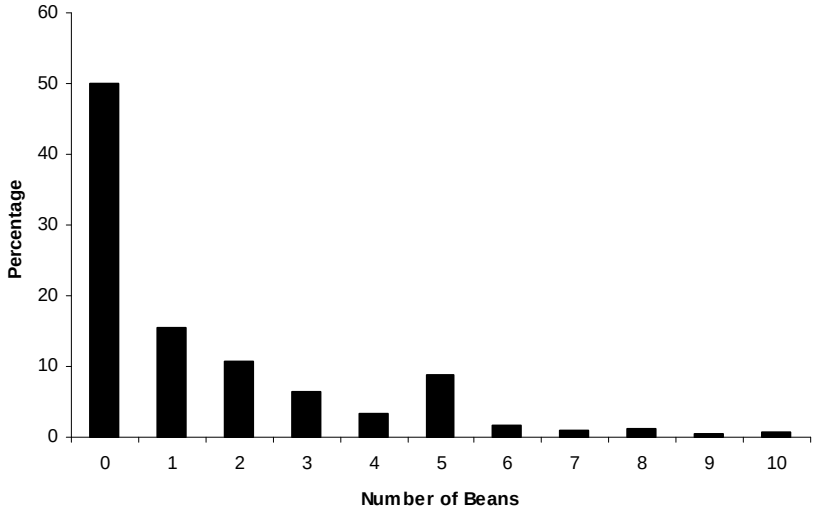

Figure 2: Beliefs by Test Result, 2006 and 2008 pooled

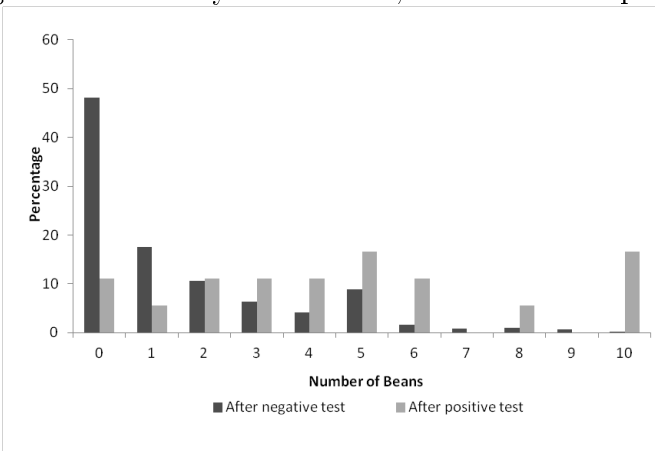

Table 4: Annual Pregnancy Probabilities

			Age Group					
			16-20	**21-25**	**26-30**	**31-35**	**36-40**	**41-45**
All		Prob.	0.405	0.377	0.307	0.273	0.211	0.06
		N	412	778	698	539	370	351
Region	**Balaka**	Prob.	0.414	0.398	0.323	0.31	0.281	0.086
		N	162	226	155	142	121	81
	Mchinji	Prob.	0.369	0.357	0.32	0.267	0.177	0.082
		N	130	319	291	172	113	85
	Rumphi	Prob.	0.433	0.382	0.282	0.253	0.176	0.038
		N	120	233	252	225	136	185
Schooling	**None**	Prob.	0.375	0.492	0.27	0.263	0.31	0.085
		N	24	61	89	95	84	82
	Primary	Prob.	0.395	0.366	0.315	0.281	0.183	0.053
		N	332	596	520	381	262	247
	Secondary	Prob.	0.482	0.372	0.292	0.238	0.167	0.045
		N	56	121	89	63	24	22
Polygamy	**Mono**	Prob.	0.412	0.388	0.309	0.285	0.23	0.077
		N	354	605	511	396	235	196
	Poly	Prob.	0.362	0.335	0.299	0.238	0.178	0.039
		N	58	173	187	143	135	155

Table 5: Parameter Estimates of the Survival-after-Infection Function

	Age Group						
	15-19	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**
ψ_τ, Weibull scale parameter	16.0	15.4	14.1	12.1	11.0	10.1	7.9
Median survival years after infection	13.3	12.8	11.7	10.0	9.1	8.4	6.6

Estimated by Hallett et al. (2008)

Table 6: Child Mortality Rates per 1000 Person-Year, by Mother HIV status

	Child's Age			
	0	**1**	**2**	**3-4**
Mother HIV-negative at birth	115	26	18	8
Mother HIV-positive at birth	331	128	87	41

Source: Crampin et al. (2003)

Figure 3: Belief Distribution, by Year

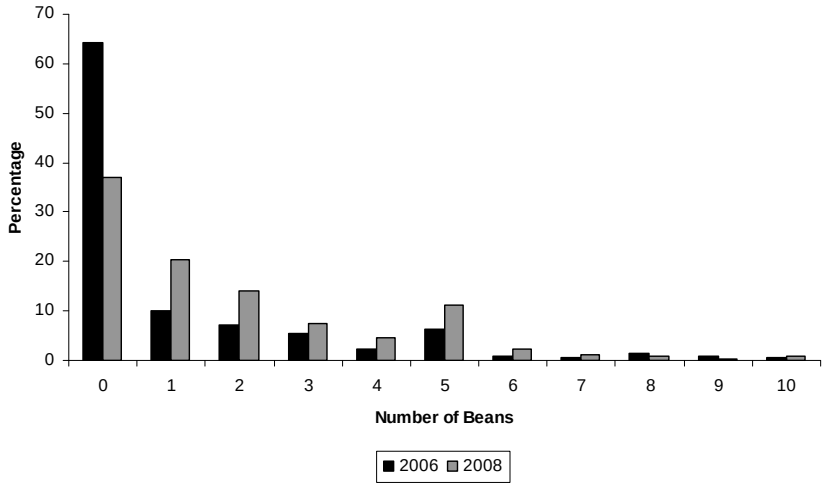

Table 7: Maximum-Likelihood Parameter Estimates: Preferences

Parameter	Description	Estimate	SE
ϕ	CRRA parameter	1.066	0.051
λ_1	Child-consumption interaction	-0.166	0.052
λ_2 - type 0	$N(t)$, Number of children, type 0	767.4	1282.58
λ_2 - type 1	$N(t)$, Number of children, type 1	1501	1150.40
λ_2 - type 2	$N(t)$, Number of children, type 2	698.8	3475.82
λ_2 - type 3	$N(t)$, Number of children, type 3	1082	1206.99
λ_2 - Balaka	$N(t)$, Balaka shifter	161	450.07
λ_2 - Rumphi	$N(t)$, Rumphi shifter	637	563.72
λ_2 - primary	$N(t)$, Primary shifter	-393.6	624.67
λ_2 - secondary	$N(t)$, Secondary shifter	196.6	785.85
λ_2 - poly	$N(t)$, Polygamy shifter	-251.5	420.21
λ_3 - type 0	$N(t)^2$, type 0 shifter	-55.08	98.12
λ_3 - type 1	$N(t)^2$, type 0 shifter	-136.9	85.09
λ_3 - type 2	$N(t)^2$, type 0 shifter	-24.78	302.23
λ_3 - type 3	$N(t)^2$, type 0 shifter	-24.76	89.28
λ_3 - balaka	$N(t)^2$, Balaka shifter	17.31	46.05
λ_3 - rumphi	$N(t)^2$, Rumphi shifter	-57.22	61.50
λ_3 - primary	$N(t)^2$, Primary shifter	6.743	64.29
λ_3 - secondary	$N(t)^2$, Secondary shifter	-72.04	86.07
λ_3 - poly	$N(t)^2$, Polygamy shifter	0.7026	43.84
λ_4	Pregnancy, $p(t)$	133.6	4966.52
λ_4 - age 20-24	$p(t)$, age 20-24 shifter	400.3	706.72
λ_4 - age 25-29	$p(t)$, age 25-29 shifter	2,432	679.71
λ_4 - age 30-34	$p(t)$, age 30-34 shifter	3,653	816.29
λ_4 - age 35-39	$p(t)$, age 35-39 shifter	5,624	1132.3
λ_4 - age 40-44	$p(t)$, age 40-44 shifter	10,680	2073.07
λ_5	Consecutive pregnancy, $p(t)p(t-1)$	8,511	1799.57
λ_6	Consecutive-age interaction, $p(t)p(t-1)t$	-218.7	92.89
σ_p	Standard deviation of preference shock	6,698	1291.56
δ	Discount factor	0.8675	0.073

Table 8: Ordinary Least Squares Parameter Estimates: Income Function

Parameter	Description	Estimate	SE
θ_1	Constant	10.619	0.135
θ_2	Balaka	-0.0717	0.046
θ_3	Mchinji	-0.092	0.049
θ_4	Land High	0.4295	0.039
θ_5	Number of children, $N(t)$	0.0499	0.011
θ_6	Period, t	0.0118	0.007
θ_7	Period squared, t^2	-0.0002	0.0001
σ_y	Standard deviation of income shocks	0.536407	0.065

Table 9: Maximum-Likelihood Parameter Estimates: Infection Hazard Function, Perceived and Actual

	Perceived		Actual	
Parameter	Estimate	SE	Estimate	SE
Constant, type 0	-8.05	1E+07	-12.6	2.448
Constant, type 1	-4.298	0.815	-9.611	1.652
Constant, type 2	0.1935	5.164	-17.51	1340.9
Constant, type 3	-8.311	2.420	-11.38	2.085
Period, type 0	-0.08	2E+07		
Period, type 1	-0.0414	0.123		
Period, type 2	-0.2406	0.281		
Period, type 3	0.1208	0.225		
Period squared, type 0	-1.88	9E+06		
Period squared, type 1	0.0004	0.004		
Period squared, type 2	0.0057	0.009		
Period squared, type 3	-0.001	0.006		
Period			0.1587	0.223
Period squared			-0.0053	0.007
Duration of marriage	0.0662	0.053	0.0171	0.099
Primary	0.6926	0.392	2.231	0.672
Secondary	0.5547	0.885	2.891	1.054
Land $>$ 1 hectare	0.1973	0.213	0.3524	0.441
Polygamy	0.3766	0.296	0.5221	0.571
Balaka	0.126	0.319	2.245	0.609
Rumphi	-2.48	5.180	1.824	1.072

Table 10: Maximum-Likelihood Parameter Estimates: Other Parameters Related to Beliefs

Parameter	Description	Estimate	SE
p_{test}, type 0	Test result accuracy, type 0	0	21.236
p_{test}, type 1	Test result accuracy, type 1	0.0301	0.115
p_{test}, type 2	Test result accuracy, type 2	0.2084	0.193
p_{test}, type 3	Test result accuracy, type 3	0.7668	0.222
σ_b	Standard deviation of reporting error	0.25	0.009

Table 11: Maximum Likelihood Parameter Estimates: Type Distribution Parameters

	Type 1		Type 2		Type 3	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
Constant	-1.748	2.614	-5.543	30.05	-3.207	2.593
Balaka	-1.275	1.297	-0.39	29.87	-1.685	1.324
Rumphi	-3.775	2.453	5.68	29.13	-2.887	2.086
Primary	-0.8293	1.412	-3.453	3.112	1.31	1.744
Secondary	-1.93	2.569	-10.28	345.8	-0.4232	2.914
Polygamy	1.688	1.112	1.793	1.011	1.03	1.128
Age of marriage	0.2263	0.242	-0.0397	0.296	0.3614	0.225
Year of birth	0.2532	0.154	0.125	0.197	0.2245	0.151
$N(\underline{t}) \times \underline{t}$	-0.0011	0.011	-2E-05	0.019	-0.0036	0.010

Table 12: Predicted Selected Characteristics by Unobserved Type

		Type 0	Type 1	Type 2	Type 3
Region	**Balaka**	0.27	0.46	0.02	0.24
	Mchinji	0.10	0.42	0.01	0.46
	Rumphi	0.63	0.11	0.97	0.30
Schooling	**None**	0.18	0.29	0.07	0.05
	Primary	0.62	0.62	0.93	0.85
	Secondary	0.20	0.09	0.001	0.10
Polygamous		0.25	0.32	0.6	0.25
Year of Birth		1971	1981	1975	1979
Age of Marriage		17.3	17.1	16.9	17.9
Prob. assigned to being HIV-positive :					
	Age 20	0	0.16	0.59	0.01
	Age 30	0	0.29	0.65	0.03
	Age 40	0	0.39	0.69	0.13
Share of sample		0.28	0.23	0.02	0.47

Figure 4: Model Fit: Actual and Predicted Reported Beliefs

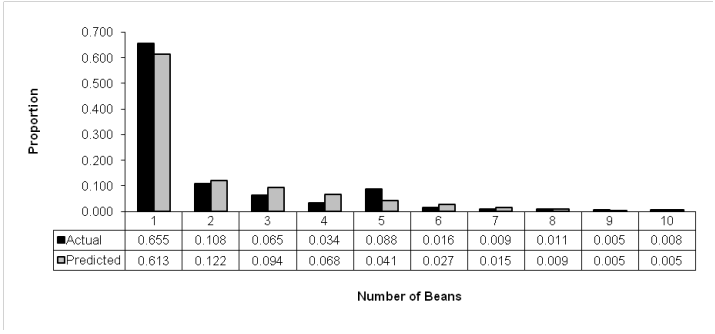

Figure 5: Model Fit: Pregnancy Probabilities, by Age Groups

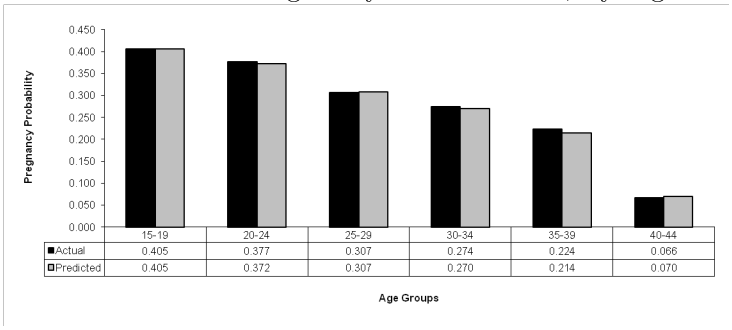

Table 13: Model Fit: Actual and Predicted Pregnancy Probabilities

			Age Group					
			15-19	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**
Region	**Balaka**	**A**	0.414	0.398	0.322	0.312	0.305	0.086
		P	0.408	0.381	0.336	0.305	0.250	0.117
		N	162	226	155	138	105	58
	Mchinji	**A**	0.369	0.357	0.32	0.268	0.192	0.104
		P	0.366	0.350	0.308	0.282	0.231	0.082
		N	130	319	291	168	99	67
	Rumphi	**A**	0.433	0.382	0.282	0.256	0.183	0.043
		P	0.444	0.393	0.290	0.238	0.1743	0.0468
		N	120	233	252	223	131	161
Schooling	**None**	**A**	0.375	0.492	0.27	0.264	0.343	0.085
		P	0.474	0.388	0.308	0.307	0.254	0.107
		N	24	61	89	91	70	59
	Primary	**A**	0.395	0.366	0.315	0.283	0.195	0.063
		P	0.397	0.367	0.307	0.268	0.207	0.062
		N	332	596	520	375	241	208
	Secondary	**A**	0.482	0.372	0.292	0.238	0.167	0.045
		P	0.428	0.388	0.309	0.228	0.177	0.036
		N	56	121	89	63	24	22
Polygamy	**Mono**	**A**	0.412	0.388	0.309	0.286	0.243	0.083
		P	0.414	0.378	0.323	0.280	0.225	0.083
		N	354	605	511	392	214	156
	Poly	**A**	0.362	0.335	0.299	0.241	0.19	0.046
		P	0.355	0.350	0.264	0.241	0.196	0.053
		N	58	173	187	137	121	130

A = actual, P = predicted, N = number of observations

Table 14: Selected Characteristics of Subsample Used for Counterfactual Simulations

		All	By Type			
Variable			0	1	2	3
Type	**0**	0.11	-	-	-	-
	1	0.32	-	-	-	-
	2	0.02	-	-	-	-
	3	0.55	-	-	-	-
Region	**Mchinji**	0.38	0.05	0.44	0.01	0.43
	Balaka	0.31	0.21	0.43	0.01	0.27
	Rumphi	0.31	0.73	0.13	0.98	0.3
Schooling	**None**	0.08	0.05	0.2	0.01	0.02
	Primary	0.76	0.48	0.69	0.99	0.85
	Secondary	0.15	0.47	0.10	0	0.12
Polygamous		0.2	0.15	0.25	0.53	0.18
Number		50,900	6,459	16,472	850	27,119

Table 15: Counterfactual Simulations

		By Type			
	All	**0**	**1**	**2**	**3**
Panel A: Baseline (HIV, without any HIV tests)					
Probability assigned to being HIV-positive:					
Age 17	0.04	0	0.11	0.50	0.003
Age 25	0.09	0	0.23	0.64	0.02
Age 35	0.16	0	0.35	0.65	0.07
HIV prevalence:					
Age 17	0.01	0.004	0.03	0	0.006
Age 25	0.04	0.01	0.10	0	0.02
Age 35	0.06	0.02	0.13	0	0.03
Number of life-cycle births	7.07	6.13	6.28	6.47	7.72
Child Mortality	1.24	1.03	1.22	1.06	1.29
Panel B: No HIV					
Number of life-cycle births	7.22	6.16	6.44	6.61	7.90
Child Mortality	1.14	0.99	1.02	1.06	1.24
Panel C: No mother-to-child transmission					
Number of life-cycle births	7.05	6.13	6.24	6.46	7.72
Child Mortality	1.13	1.00	1.00	1.06	1.23

	All	Type 0	Type 1	Type 2	Type 3
Panel D: HIV test at period of marriage					
Probability assigned to being HIV-positive:					
Age 17	0.04	0	0.11	0.44	0.004
Age 25	0.09	0	0.23	0.59	0.02
Age 35	0.16	0	0.35	0.63	0.07
Number of life-cycle births	7.07	6.13	6.28	6.47	7.72
Child Mortality	1.24	1.03	1.22	1.05	1.29
Panel E: HIV test at age 25					
Probability assigned to being HIV-positive:					
Age 25	0.09	0	0.23	0.51	0.02
Age 35	0.15	0	0.34	0.58	0.07
Number of life-cycle births	7.07	6.13	6.28	6.47	7.72
Child Mortality	1.24	1.03	1.22	1.06	1.29
Panel F: HIV test at age 35					
Probability assigned to being HIV-positive:					
Age 35	0.13	0	0.34	0.52	0.04
Number of life-cycle births	7.07	6.13	6.28	6.46	7.73
Child Mortality	1.24	1.03	1.22	1.06	1.29
Panel G: HIV test at age 25, assigning full accuracy to test results					
Probability assigned to being HIV-positive:					
Age 25	0.04	0.01	0.10	0	0.02
Age 35	0.14	0.01	0.29	0.42	0.07
Number of life-cycle births	7.06	6.13	6.27	6.50	7.72
Child Mortality	1.23	1.03	1.22	1.05	1.29

B Specification of the infection hazard function (Section 3.4)

The perceived probability of getting infected at period t, conditional on not getting infected before is given by

$$h(t) = \frac{1}{1 + \exp\left(-x(t)'\beta\right)},$$

where

$$x\left(t\right)'\beta = \beta_1 + \beta_{2,\mu}t + \beta_{3,\mu}t^2 + \beta_4 \text{Marriage Duration} + \beta_5 \text{Primary}$$

$$+\beta_6 \text{Secondary} + \beta_7 \text{Land High} + \beta_8 \text{Poly} + \beta_9 \text{Balaka} + \beta_{10} \text{Rumphi}.$$

The parameters related to age and age squared are allowed to vary with unobserved type (μ). As described in section 4.3, the same specification is used for the actual infection hazard process, with h replaced by $\tilde{h}$ and β replaced by $\tilde{\beta}$.

C Updating beliefs after testing (Section 3.5)

I want to solve for $\hat{P}(\tau),\ \tau = 1, ..., t_{test}$ from the system of equations presented in Equation 10:

$$\frac{\hat{P}(\tau) S(\tau,\ t_{test})}{1 - \sum_{k=1}^{t_{test}} \hat{P}(k) (1 - S(k,\ t_{test}))} = G(\tau,\ t_{test}) \hat{B}(t_{test}), \quad \tau = 1, ..., t_{test}$$

Isolating $\hat{P}(\tau)$:

$$\hat{P}(\tau) S(\tau,\ t_{test}) = \left[1 - \sum_{k=1}^{t_{test}} \hat{P}(k) (1 - S(k,\ t_{test}))\right] G(\tau,\ t_{test}) \hat{B}(t_{test}), \quad \tau = 1, ..., t_{test}$$

$$\hat{P}(\tau) = \left[1 - \sum_{k=1, k\neq\tau}^{t_{test}} \hat{P}(k) (1 - S(k,\ t_{test}))\right] \frac{G(\tau,\ t_{test})\hat{B}(t_{test})}{S(\tau,\ t_{test}) + (1 - S(\tau,\ t_{test}))G(\tau,\ t_{test})\hat{B}(t_{test})}, \quad \tau = 1, ..., t_{test}$$

Introducing notation:

1. $\psi(\tau,\ t_{test}) = \frac{G(\tau,\ t_{test})\hat{B}(t_{test})}{S(\tau,\ t_{test}) + (1 - S(\tau,\ t_{test}))G(\tau,\ t_{test})\hat{B}(t_{test})}, \quad \tau = 1, ..., t_{test}$

2. $\zeta(\tau,\ t_{test}) = \psi(\tau,\ t_{test}) (1 - S(\tau,\ t)), \quad \tau = 1, ..., t_{test}$

3. $\text{Num}(\tau) = \psi(\tau,\ t_{test}) \prod_{i=1, i\neq\tau}^{t_{test}-1} (1 - \zeta(i,\ t_{test}))$

4. $\text{Den}(\tau) = 1 - \frac{1}{2} \sum_{k=1}^{t_{test}-1} \zeta(k,\ t_{test}) \left(\sum_{i=1, i\neq k}^{t_{test}-1} \zeta(i,\ t_{test})\right)$
   $+ \frac{2}{3} \sum_{k=1}^{t_{test}-1} \zeta(k,\ t_{test}) \left(\sum_{i=1, i\neq k}^{t_{test}-1} \zeta(i,\ t_{test}) \left(\sum_{m=1, m\neq k, m\neq i}^{t_{test}-1} \zeta(m,\ t_{test})\right)\right)$
   $- \cdots + [-1]^{t_{test}} \frac{t_{test}-2}{t_{test}-1} \prod_{i=1}^{t_{test}-1} \zeta(i,\ t_{test})$

The solution to the system of equations is given by:

$$\hat{P}(\tau) = \frac{\text{Num}(\tau)}{\text{Den}(\tau)}, \quad \tau = 1, ..., t_{test} - 1$$

$$\hat{P}(t_{test}) = \psi(t_{test},\ t_{test}) \left[1 - \sum_{k=1}^{t_{test}-1} \hat{P}(k)\,(1 - S(k,\ t_{test}))\right], \quad \tau \geq t_{test}$$

D Probabilities of Actual HIV and Survival History (Section 4.3)

Let $\tilde{h}(t)$ be the actual HIV infection hazard rate. The functional form is assumed to be similar to that of the perceived infection hazard described in equation (2). The probability of getting infected at t, conditional on being HIV-negative until then, is given by

$$\tilde{h}(t) = \frac{1}{1 + \exp\left(-x(t)'\tilde{\beta}\right)}.$$

The unconditional probability of getting infected at period t is given by

$$\tilde{P}(t) = \tilde{h}(t) \prod_{k=1}^{t-1} \left(1 - \tilde{h}(k)\right).$$

Information about the hazard process is contained in the HIV test results, the age in which the tests were taken, and the ages in which a woman is last observed (regardless of testing histories). Let $\overline{t_i}$ represent the age in which a woman is last observed (and is therefore known to be alive at that age). Let t_i^- be the *oldest* age in which a woman is observed to get a negative test result. Let t_i^+ be the *youngest* age a woman is observed to receive a positive test result. Let $H_i = (t_i^-,\, t_i^+,\, \overline{t_i})$ be the vector of observed "HIV history" of woman i, with t_i^- (t_i^+) equaling zero if a woman is never tested negative (positive). Women's "HIV histories" belong to one of the following 4 categories:

1. Never tested:
   The probability of observing a woman who was never tested alive at $\overline{t_i}$, conditional on her initial conditions and unobserved type, is given by

$$\Pr\left(H_i = (0,\, 0,\, \bar{t}) \mid \underline{\Omega}_i^d,\, \text{type}_i = j\right) = \Pr\left(\text{alive at } \bar{t}\right) = 1 - \sum_{k=1}^{\overline{t_i}} \tilde{P}\left(k \mid \underline{\Omega}_i^d,\, \text{type}_i = j\right)\left(1 - S(k,\, \overline{t_i})\right).$$

2. Tested only negative:
   The probability of observing a woman who was only tested negative and is alive at $\overline{t_i}$,

conditional on her initial conditions and unobserved type, is given by

$$\Pr\left(H_i = \left(t^-,\ 0,\ \bar{t}\right)\ \middle|\ \underline{\Omega}_i^d,\ \text{type}_i = j\right) = \Pr\left(\text{negative at } t^-, \text{alive at } \bar{t}\right)$$

$$= \prod_{k=1}^{t^-}\left(1-\tilde{h}(k)\right)\left[1-\sum_{k=t^-+1}^{\bar{t}}\tilde{h}\left(k \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right)\prod_{j=t^-+1}^{k}\left(1-\tilde{h}\left(j \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right)\right)\left(1-S\left(k,\ \bar{t}\right)\right)\right]$$

$$= \prod_{k=1}^{t^-}\left(1-\tilde{h}(k)\right) - \sum_{k=t^-+1}^{\bar{t}}\tilde{P}\left(k \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right)\left(1-S\left(k,\ \bar{t}\right)\right).$$

3. Tested only positive:
   The probability of observing a woman who if infected by t^+ and alive at $\bar{t}$, conditional on her initial conditions and unobserved type, is given by

$$\Pr\left(H_i = \left(0,\ t^+,\bar{t}\right)\ \middle|\ \underline{\Omega}_i^d,\ \text{type}_i = j\right) = \Pr\left(\text{got infected before } t^+, \text{alive at } \bar{t}\right)$$

$$= \sum_{k=1}^{t^+}\tilde{P}\left(k \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right)S\left(k,\ \bar{t}\right).$$

4. Seroconverter:
   The probability of observing a woman who is known to be HIV-negative until t^-, positive by t^+, and alive at $\bar{t}$ $\left(t^- < t^+ \leq \bar{t}\right)$, conditional on her initial conditions and unobserved type, is given by

$$\Pr\left(H_i = \left(t^-,\ t^+,\bar{t}\right)\ \middle|\ \underline{\Omega}_i^d,\ \text{type}_i = j\right) = \Pr\left(\text{got infected between } t^- \text{ and } t^+,\ \text{alive at } \bar{t}\right)$$

$$= \sum_{k=t^-}^{t^+}\tilde{P}\left(k \mid \underline{\Omega}_i^d,\ \text{type}_i = j\right)S\left(k,\ \bar{t}\right).$$