Online Appendix

A Quantitative Theory of the HIV Epidemic:

Education, Risky Sex and Asymmetric Learning

By Christian Aleman-Pericon, Daniela Iorio and Raül Santaeulàlia-Llopis

A Data appendix

List of countries used

Burkina Faso (2003, 2010), Burundi (2010), Cameroon (2004, 2011), Congo (2007), Côte d'Ivoire (2005, 2011), Democratic Republic of Congo (2007), Ethiopia (2005, 2011), Gabon (2012), Ghana (2003), Guinea (2005, 2012), Kenya (2003, 2008), Lesotho (2004, 2009), Liberia (2007), Malawi (2004, 2010), Mali (2006), Mozambique (2009), Niger (2006), Rwanda (2005, 2010), Senegal (2005, 2010), Sierra Leone (2008), Swaziland (2006), Tanzania (2003, 2007, 2011), Uganda (2011), Zambia (2007) and Zimbabwe (2005, 2010).

Survey questions about knowledge of HIV transmision Specifically, respondents answer two questions: (i) *Can you (the respondent) reduce the chances of getting HIV by having one sex partner who has no other partners?* and (ii) *Can you (the respondent) reduce the chances of getting HIV by always wearing a condom?*

B Normalization Results: Discretization of Stages

The discretization discussed in Section 2.2 implies the following assignemnt of DHS datasets, $\omega(s_i, \widetilde{g}_i(\tau_i))$, over a set of five stages of the epidemic as follows:

- Stage $\leq$0: Cameroon 2004, Guinea 2005, Guinea 2012, Lesotho 2004/05, Lesotho 2009/10, Mozambique 2009, Senegal 2005, Sierra Leone 2008, Swaziland 2006/07.
- Stage 1: Cameroon 2011, Cote d'Ivoire 2005, Democratic Republic Congo 2007, Ethiopia 2005, Ghana 2003, Kenya 2003, Liberia 2006/07, Malawi 2004/05, Mali 2006, Niger 2006, Senegal 2010/11, Tanzania 2003/04, and Zambia 2007.
- Stage 2: Gabon 2012, Malawi 2010, Tanzania 2007/08, Zimbabwe 2005.
- Stage 3: Congo Brazaville 2009, Cote d'Ivoire 2011, Kenya 2008/09, Rwanda 2005, Tanzania 2011/12, Zimbabwe 2006.
- Stage $\geq$4: Burkina Faso 2003, Burkina Faso 2010, Burundi 2010, Cote d'Ivoire 2012, Ethiopia 2011, Niger 2012, Rwanda 2010/11, Uganda 2011 and Zimbabwe 2010/11.

C The HIV-Education Gradient: Further Results and Robustness

This appendix collects tables containing results related to our main specification

Table 6: Additional Inference

(A) *HIV-Education Gradient*	(1)	(2)	(3)	(4)	(5)
γ_0	0.0043***	0.0112***	0.0098***	0.0040***	0.0037***
	(0.0006)	(0.0007)	(0.000)	(0.0003)	(0.0003)
$\gamma_0 + \gamma_1$		0.0053***	0.0052***	0.0029***	0.0029***
		(0.0005)	(0.000)	(0.0003)	(0.0003)
$\gamma_0 + \gamma_2$		-0.0004	-0.0005	0.0013***	0.0013***
		(0.0003)	(0.228)	(0.0001)	(0.0001)
$\gamma_0 + \gamma_3$		0.0019	0.0022***	0.0020*	0.0019*
		(0.0014)	(0.005)	(0.0011)	(0.0011)
$\gamma_0 + \gamma_4$		0.0048***	0.0047***	0.0025***	0.0025***
		(0.0005)	(0.000)	(0.0002)	(0.0001)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes

(B) *Rebound*	(1)	(2)	(3)	(4)	(5)
$\gamma_4 - \gamma_2$		0.0052***	0.0051***	0.0011***	0.0012***
		(0.0006)	(0.0005)	(0.0001)	(0.0001)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes

Notes: The underlying econometric models are as specified in the columns of Table 2. Column (1) reports the tests results for the stationary specification. Columns (2) to (5) report the tests results for the non-stationary specification. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: The HIV-Education Gradient, Sexually Active Sample

HIV Status	(1)	(2)	(3)	(4)	(5)
Education	0.0048***	0.0123***	0.0107***	0.0042***	0.0039***
	(0.0009)	(0.0008)	(0.0013)	(0.0003)	(0.0004)
Education * Stage1		-0.0064***	-0.0047***	-0.0007	-0.0004
		(0.0010)	(0.0013)	(0.0004)	(0.0005)
Education * Stage2		-0.0129***	-0.0115***	-0.0029***	-0.0027***
		(0.0009)	(0.0014)	(0.0003)	(0.0004)
Education * Stage3		-0.0101***	-0.0081***	-0.0018	-0.0016
		(0.0020)	(0.0015)	(0.0015)	(0.0014)
Education * Stage4		-0.0067***	-0.0054***	-0.0019***	-0.0014***
		(0.0013)	(0.0016)	(0.0004)	(0.0004)
Male	-0.0267***	-0.0273***	-0.0272***	-0.0268***	-0.0269***
	(0.0039)	(0.0038)	(0.0025)	(0.0022)	(0.0021)
Age	0.0021***	0.0021***	0.0021***	0.0022***	0.0022***
	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0002)
Urban Area	0.0260***	0.0243***	0.0276***	0.0337***	0.0344***
	(0.0059)	(0.0059)	(0.0053)	(0.0038)	(0.0038)
Stage 1	-0.0018	0.0137***	0.0097	-0.0082***	0.0099***
	(0.0062)	(0.0047)	(0.0060)	(0.0012)	(0.0023)
Stage 2	0.0132	0.0573***	0.0676***	0.0010	0.0289***
	(0.0105)	(0.0122)	(0.0136)	(0.0029)	(0.0045)
Stage 3	-0.0114	0.0198	0.0330**	-0.0129**	0.0138**
	(0.0164)	(0.0149)	(0.0128)	(0.0061)	(0.0060)
Stage 4	-0.0027	0.0143***	0.0436***	-0.0183***	0.0056
	(0.0052)	(0.0040)	(0.0097)	(0.0025)	(0.0036)
Agricultural Share	-0.0034***	-0.0034***	-0.0035***	0.0027***	-0.0006
	(0.0003)	(0.0003)	(0.0004)	(0.0005)	(0.0004)
Output per Capita	-0.0000***	-0.0000***	-0.0000***	0.0001**	0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.1128***	0.0918***	0.0554***	-0.1603***	-0.1992***
	(0.0101)	(0.0075)	(0.0098)	(0.0493)	(0.0282)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	329,205	329,205	329,205	329,205	329,205

Notes: All specifications use the "Sexually Active" subsample. he underlying econometric models are as specified in the columns of Table 2. Column (1) reports the tests results for the stationary specification. Columns (2) to (5) report the tests results for the non-stationary specification. We include the same set of controls and fixed effects as in our benchmark specifications in Table 2. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8: The Risky Sex-Education Gradient: Women and Men Separately

(A) Number of Extramarital Partners

	Women					Men				
Risky Sex	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0177**	0.0372***	0.0348***	0.0294***	0.0296***	0.0285***	0.0304***	0.0269***	0.0290***	0.0295***
	(0.029)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)
Education * Stage1		-0.0163**	-0.0222***	-0.0204***	-0.0199***		0.0046	-0.0071	-0.0167**	-0.0167**
		(0.025)	(0.000)	(0.006)	(0.008)		(0.742)	(0.515)	(0.050)	(0.029)
Education * Stage2		-0.0254***	-0.0267***	-0.0243***	-0.0249***		-0.0200	-0.0238***	-0.0290***	-0.0294***
		(0.006)	(0.000)	(0.003)	(0.003)		(0.125)	(0.007)	(0.000)	(0.000)
Education * Stage3		-0.0308***	-0.0288***	-0.0201**	-0.0213**		-0.0069	-0.0047	0.0010	0.0003
		(0.002)	(0.000)	(0.019)	(0.018)		(0.550)	(0.588)	(0.932)	(0.980)
Education * Stage4		-0.0152	-0.0134	-0.0134*	-0.0138*		-0.0010	0.0019	0.0021	0.0018
		(0.107)	(0.185)	(0.081)	(0.087)		(0.928)	(0.845)	(0.840)	(0.866)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	227,935	227,935	227,9358	227,935	227,935	174,831	174,831	174,831	174,831	174,831

(B) Condom Use in Last Intercourse

	Women					Men				
Risky Sex	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0079***	0.0192***	0.0193***	0.0125***	0.0126***	0.0066***	0.0129***	0.0141***	0.0162***	0.0166***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education * Stage1		-0.0102***	-0.0114***	-0.0054**	-0.0055***		-0.0038	-0.0054	-0.0064*	-0.0066*
		(0.001)	(0.142)	(0.011)	(0.002)		(0.0406)	(0.122)	(0.084)	(0.065)
Education * Stage2		-0.0135***	-0.0132***	-0.0075***	-0.0075***		-0.0109***	-0.0111**	-0.0149***	-0.0156***
		(0.000)	(0.000)	(0.000)	(0.000)		(0.010)	(0.028)	(0.000)	(0.000)
Education * Stage3		-0.0138	-0.0133	-0.0068	-0.0068		-0.0101	-0.0108	-0.0131	-0.0136
		(0.164)	(0.000)	(0.482)	(0.501)		(0.611)	(0.640)	(0.592)	(0.585)
Education * Stage4		-0.0134**	-0.0133***	-0.0073**	-0.0080**		-0.0078**	-0.0088*	-0.0066***	-0.0074***
		(0.000)	(0.000)	(0.024)	(0.007)		(0.017)	(0.062)	(0.004)	(0.001)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	163,883	163,883	163,883	163,883	163,883	120,840	120,840	120,840	120,840	120,840

Notes: In panel (A) we report the marginal effects of the associated Tobit model where the endogenous variable is the number of extramarital partners in the past 12 months. In panel (B) we report the coefficients of a linear model where the endogenous variable is binary and refers to use of condom in last sexual intercourse. In both panels we include the same set of controls and fixed effects as in our benchmark specifications in Table **??**. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis.* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 9: The Knowledge-Education Gradient: Women and Men Separately

(A) One Sexual Partner without Other Partners

	Women					Men				
HIV Knowledge	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0135	0.0033	0.0005	0.0074***	0.0076***	0.0076**	0.0010	0.0013	0.0078***	0.0087***
	(0.111)	(0.110)	(0.866)	(0.000)	(0.000)	(0.001)	(0.326)	(0.580)	(0.000)	(0.000)
Education * Stage1		0.0126	0.0120**	0.0047	0.0046		0.0078***	0.0059**	0.0005	-0.0001
		(0.156)	(0.043)	(0.309)	(0.292)		(0.000)	(0.039)	(0.840)	(0.952)
Education * Stage2		0.0072**	0.0093**	0.0038**	0.0033		0.0091***	0.0086***	0.0028*	0.0017
		(0.046)	(0.049)	(0.045)	(0.210)		(0.000)	(0.001)	(0.088)	(0.285)
Education * Stage3		0.0094	0.0120	0.0018	0.0020		0.0057	0.0053	-0.0034	-0.0042
		(0.612)	(0.412)	(0.912)	(0.911)		(0.386)	(0.337)	(0.444)	(0.304)
Education * Stage4		0.0104	0.0134	0.0071	0.0062		0.0069**	0.0075*	0.0019	0.0007
		(0.732)	(0.648)	(0.812)	(0.830)		(0.044)	(0.066)	(0.424)	(0.759)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	213,907	213,907	213,907	213,907	213,907	167,894	167,894	167,894	167,894	167,894

(B) Always Use Condom During Sex

	Women					Men				
HIV Knowledge	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0221	0.0194***	0.0169***	0.0196***	0.0196***	0.0126***	0.0132***	0.0141***	0.0187***	0.0181***
	(0.133)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Education * Stage1		0.0044	0.0073	0.0034	0.0039		-0.0006	-0.0007	-0.0036	-0.0029
		(0.801)	(0.593)	(0.809)	(0.754)		(0.935)	(0.929)	(0.638)	(0.668)
Education * Stage2		-0.0026	0.0003	-0.0017	-0.0025		0.0015	0.0017	-0.0036	-0.0035
		(0.616)	(0.953)	(0.648)	(0.572)		(0.639)	(0.637)	(0.236)	(0.246)
Education * Stage3		0.0009	0.0039	0.0008	0.0014		-0.0028	-0.0024	-0.0073	-0.0064
		(0.956)	(0.804)	(0.952)	(0.920)		(0.770)	(0.789)	(0.415)	(0.496)
Education * Stage4		0.0067	0.0089	0.0047	0.0033		0.0006	-0.0004	-0.0046*	-0.0047**
		(0.946)	(0.929)	(0.963)	(0.974)		(0.828)	(0.910)	(0.065)	(0.029)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	213,763	213,763	213,763	213,763	213,763	167,800	167,800	167,800	167,800	167,800

Notes: In panel (A) we report the coefficients of a linear model where the endogenous variable is binary for *"Can you (the respondent) reduce the chances of getting HIV by having one sex partner who has no other partners?"*. In panel (B) we report the coefficients of a linear model where the endogenous variable is binary for *"Can you (the respondent) reduce the chances of getting HIV by always wearing a condom?"*. In both panels we include the same set of controls and fixed effects as in our benchmark specifications in Table **??**. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis.* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 10: The HIV-Education Gradient: Specification by Stage

(A) *HIV Status*	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Education	0.0044***	0.0022***	0.0014***	0.0020*	0.0025***
	(0.000)	(0.000)	(0.000)	(0.081)	(0.000)
Year-Country Dum.	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes
Sample Size	58,560	112,024	48,615	50,535	118,425

(B) *HIV Status with ART*	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Education	0.0043***	0.0029***	0.0016***	0.0011	0.0015***
	(0.000)	(0.0003)	(0.0002)	(0.0021)	(0.0002)
ART Coverage	0.0003*	-0.0025***	0.0015***	-0.0012***	-0.0003***
	(0.0002)	(0.0005)	(0.0000)	(0.0001)	(0.0000)
Education * ART Coverage	-0.0000	-0.0000	-0.0000***	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Year-Country Dum.	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes
Sample Size	58,560	112,024	48,615	50,535	118,425

Notes: We apply our stationary specification of the HIV-Education gradient separately for each stage of the epidemic. We include year and country fixed effects in all columns. In both panels we include the same set of controls as in our benchmark specifications in Table 2. We exclude Senegal (in Stage 0) and Niger (in Stage 1) since WPP does not provide information about ART coverage for these countries. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

D Solution Algorithm

Computing the recursive stationary equilibrium for Stage 0 (Pre-epidemic Stage)

***Algorithm No.1*: Computation of the recursive stationary equilibrium of the Pre-Epidemic stage:**

Step 1: Make initial guesses of price p.

Step 2: Compute the agents decision rules.

Step 3: Compute the stationary distribution of the population across states (follow Algorithm No.2).

Step 4: Compute aggregate sex demand and aggregate sex supply. Check the aggregate consistency conditions.

Step 5: If conditions are not met, update p and return to Step 2.

Keep in mind that *Algorithm No.1* is general enough so that it can also be applied to compute the stationary equilibrium of any stage of the HIV epidemic.
In the absence of continuous state variables, the decision rules are reduced to single values conditional on the different states

***Algorithm No.2*: Computation of the invariant distribution of the population:**

Step 1: Make an initial guess for the (discrete) density function $\phi_0(e, i, \varepsilon)$ over the state space[46]

Step 2: For all $e \in E$, $i \in \mathcal{I}$ and $\epsilon \in \mathcal{E}$ compute the following expression:

$$\phi_{t+1}(e, i, \varepsilon') = \sum_{e,i,s} \sum_{\varepsilon'|\varepsilon} \gamma\pi(\varepsilon'|\varepsilon)\phi_t(e, i, \varepsilon) + \psi\phi_t(e, i, \varepsilon') \tag{34}$$

Step 3: If $|\phi_{t+1} - \phi_t|$ is close to zero stop, otherwise set $\phi_t = \phi_{t+1}$ and return to Step 2.

Computing the Non-stationary equilibrium for Stage 1 (Myopic Stage)

***Algorithm No.3*: Computation of the solution of the Myopic stage:**
We are after a sequence of $\{\Phi\}_{T_0}^{T_1}(s = 1)$ where at each period $\tau \in \{T_0 + 1, ..., T_1\}$, agents get a permanent unexpected shock to $\widetilde{\gamma}$, $\widetilde{z}$ and $\widetilde{\chi}$ following 13 and 14. T_1 being the period in which Stage 2 quick's in. To get each of the elements of the sequence we need to solve for an entire transition.

Step 0: Set $\tau = T_0 + 1$,

Step 1: Following (13) and (14) compute new values for $\widetilde{\gamma}_\tau$, $\widetilde{z}_\tau$ and $\widetilde{\chi}_\tau$ (Remember agents believe these values will be permanent).

[46]We choose the uniform distribution as the initial values of the distribution. The algorithm should converge regardless of the choice of the initial distribution.

Step 2: Compute the recursive stationary equilibrium of the Myopic stage associated with the new $\widetilde{\gamma}_\tau$, $\widetilde{z}_\tau$ and $\widetilde{\chi}_\tau$.

Step 3: Choose a very large number of transition periods $(\mathscr{T} - \tau)$.

Step 4: Compute the equilibrium transition dynamics from the stationary equilibrium from Step 2 to the equilibrium in period τ. We compute this going backwards.

Step 5: Stop if $\tau = T_1$

Step 6: Replace $\tau = \tau + 1$ and go back to Step 1.

Computing the Non-stationary equilibrium for Stage 2 (Learning Stage)

***Algorithm No.4*: Computation of the solution of the Maturity Stage:**
We are after a sequence of $\{\Phi\}_{T_1}^{\infty}(s = 2)$ that goes from the last period in Stage 1 (Myopic Stage) $(\tau = T_1)$ to the recursive stationary equilibrium of Stage 2 of the epidemic (Learning Stage).

Step 1: Choose a large number of transition periods $(\mathscr{T} - T_1)$

Step 2: Simulate a sequence of $\widetilde{\rho}_e$ by education group, following (21).

Step 3: Compute the recursive stationary equilibrium of Stage 2 of the epidemic. This stationary equilibrium is associated with $\lim_{t\to\infty} \widetilde{\rho}_{e,t} = \rho$. That is, in the stationary equilibrium both education groups have completed learning of the true probability HIV infection risk as a function of sex.

Step 4: Guess a time path for the prices $\{p_t\}_{T_1}^{\mathscr{T}}$.

Step 5: Compute the equilibrium policy (and value) functions iterating backwards in time, $t = \mathscr{T} - 1, ..., T_1$.

Step 6: Simulate the evolution of the population distribution from $t = T_1$ to $t = T$ with the help of the optimal policy functions and the initial distribution $\Phi_{T_1}(g = 1)$.

Step 7: Compare the simulated distribution at $\mathscr{T}$ with the stationary distribution function from Step 3. If they are not the same try increasing the horizon $\mathscr{T}$ and go back to Step 2.

Step 8: Compute the time path of excess demand for sex. If markets don't clear along the path, then update $\{p_t\}_{T_1}^{\mathscr{T}}$ return to step 5.

Computing the Non-stationary equilibrium for Stage 3 (ARV Stage)

***Algorithm No.5*: Computation of the solution of the ARV Stage:**
After T_2 periods in Stage 2 (Learning Stage), ARV's are introduced unexpectedly. We are after a sequence of $\{\Phi\}_{T_2}^{\infty}(s = 3)$ that go from period T_2 of Stage 2 to the recursive stationary equilibrium of Stage 3 of the epidemic (ARV Stage).

Step 1: Choose a large number of transition periods $(\mathscr{T} - T_2)$

Step 2: Compute the recursive stationary equilibrium of Stage 3 of the epidemic. This stationary equilibrium is associated with $\lim_{t\to\infty} \hat{\eta}_t = \widetilde{\eta} = 1$. That is, there is full coverage for all HIV infected individuals.

Step 3: Guess a time path for the prices $\{p_t\}_{T_2}^{\mathscr{T}}$.

Step 4: Given the sequences for ARV coverage by education group $\{\hat{\eta}_t\}_{T_2}^{\mathscr{T}}$, compute the equilibrium policy (and value) functions iterating backwards in time, $t = \mathscr{T} - 1, ..., T_2$.

Step 5: Simulate the evolution of the population distribution from $t = T_2$ to $t = \mathscr{T}$ with the help of the optimal policy functions and the initial distribution $\Phi_{T_2}(g = 2)$.

Step 6: Compare the simulated distribution at $\mathscr{T}$ with the stationary distribution function from Step 2. If they are not the same try increasing the horizon $\mathscr{T}$ and go back to Step 3.

Step 7: Compute the time path of excess demand for sex. If markets don't clear along the path, then update $\{p_t\}_{T_2}^{\mathscr{T}}$ return to Step 4.