Appendices

Online Appendix

OA.1 Data Construction and Variable Definitions

This appendix defines the samples and the variables used in the paper.

OA.1.1 Data source and cohort coverage

The NLSY79 follows a nationally representative cohort born 1957–1964, aged 14–22 at the first interview in 1979. Interviews are annual from 1979 to 1994 and biennial thereafter.

OA.1.2 Panel structure and alignment to model time

The data form an annual panel indexed by woman i and calendar year t, with age at interview

$$\text{Age}_{it} = t - \text{BirthYear}_i,$$

where birth year is inferred from age at baseline when age is missing. The model runs on four-year periods from age 14, so annual observations map into

$$\text{AgeBin}_{it} = 14 + 4\left\lfloor \frac{\text{Age}_{it} - 14}{4} \right\rfloor,$$

the periods 14–17, 18–21, ..., 38–41. Period-level objects (employment, experience, the fertility hazard) aggregate annual measures within the period by the rules below.

OA.1.3 Estimation samples and the model's age grid

Estimation sample. Every moment the model targets, and every AFQT quartile in the paper, is defined on one sample of women (Table OA.1). Of the 6,283 NLSY79 women, 5,939 have a valid AFQT score; 119 of these hold a college degree without having attended college at ages 18–22, a route the model does not allow, and are dropped, leaving 5,820. AFQT quartiles are assigned on this sample, one observation per woman, with cut-points at 17.2, 38.4 and 65.2 percentile points; a woman in a reduced-form sample who is outside it is assigned a quartile by the same cut-points. Education is the highest grade ever completed.

Conditioning sets. Each block of moments is computed over the women, or woman–years, for whom its object is defined (Table OA.2); differences in the number of observations across

Table OA.1. Estimation Sample

	Women
NLSY79 women	6,283
with a valid AFQT score	5,939
with a valid date of first birth, or no birth	5,939
a college degree implies college attendance at 18–22	5,820

Notes: AFQT quartiles are drawn on the last row, one observation per woman.

blocks reflect these conditioning sets and item non-response, not different samples.

Table OA.2. Targeted Moments: Unit of Observation and Conditioning Set

Block	Unit	Conditioning set
Schooling and early fertility	woman	all women; college graduation among women who attended college, by whether they had a first birth by 21
First births by age	woman	all women; cumulative share with a first birth by the end of each period
Marriage by age	woman	all women, by education; cumulative share married by the end of each period
Employment	woman–year	all women, by education and period
Contraception use	woman–year	childless women aged 22–41 who have had intercourse and are not pregnant
Unintended first birth by 21	woman	all women, except those whose first birth by 21 has no wantedness answer
Unintended share by age at first birth	woman	women with a first birth at 14–41 that has a wantedness answer
Method failure (and pill share)	woman–wave	childless women aged 18–33 at the seven contraception waves, using a reversible method, not trying to conceive

Notes: Every block is restricted to the women of Table OA.1 except the method-failure block, which is built on the within-method sample of Section 2 and uses the same quartile cut-points. Employment follows the definitions of Appendix OA.10.

Standard errors. A cluster bootstrap over women (1,000 replications) resamples the schooling, first-birth, marriage, employment, contraception-use and unintended-first-birth moments jointly, so their covariances across blocks are those of the data. The method-failure moments are bootstrapped on their own sample, the unintended shares by age at first birth use binomial variances, and both blocks enter the covariance matrix as independent of the rest.

Reduced-form samples. The samples of Section 2 are woman–wave panels of the seven waves that field the contraception and wantedness modules (1982, 1984, 1985, 1986, 1988, 1990 and 1992). The unconditional sample keeps women aged 18–35 at the wave with a valid AFQT score and drops woman–waves at which the woman reports a pregnancy she had wanted: 32,091 woman–waves (5,884 women). The within-method sample keeps those woman–waves, removes intended births from the outcome instead, and contains the 16,087 woman–waves at which a pill, barrier or behavioral method is reported. Education in these samples is the highest grade completed at the wave.

Age grid. Every target is measured on the model's four-year periods, 14–17 through 38–41. First birth and marriage are absorbing states, so the data moment for a period is the share of women who have had a first birth (married) by the last age in that bin, not an average of a person-year indicator over the bin; employment and contraception use are within-period flows, averaged over the four ages of the bin. The adherence parameters $c_{\theta g}$ are indexed by the coarser groups 14–17, 18–25 and 26–33, with ages 34–41 assigned the 26–33 value; the reduced-form tables of Section 2 and Online Appendix OA.5 use the same groups where they split by age. The contraception block starts at age 22: the NLSY79 question is asked of women who have ever had intercourse and are not pregnant, a universe that differs sharply by education at 18 (Section 4).

Exposure and the allowance scale. Two ability-indexed inputs are measured on the estimation sample and held fixed. The exposure share $\pi_{\theta t}$ (Section 3.4) is the share of women in quartile θ who are childless at the start of period t and report, in the 1983–85 rounds, a first intercourse by its end. The allowance scale s_θ (Section 3) is mean total net family income at interviews at ages 14–17 (1979–82, deflated by the interview-year mean), by quartile, relative to the overall mean and renormalized to average one: 0.63, 0.86, 1.07, 1.45. Both are unweighted, like every target.

OA.1.4 Global cleaning conventions and special codes

NLSY79 variables encode nonresponse and survey routing (refusal, don't know, valid skip, non-interview) as negative values. Three conventions apply throughout:

1. **Nonresponse codes:** values < 0 are missing unless they have a structural meaning, such as "no spouse" for spouse income.

2. **Structural zeros:** economically meaningful zeros, such as spouse income with no partner present, are set to 0.

3. **Deflation:** dollar amounts are in 2016 dollars, deflated by the CPI by calendar year.

OA.1.5 Cognitive ability

Ability is the AFQT score of the NLSY79 created score files. Its quartiles, $q \in \{1, 2, 3, 4\}$, drawn once per woman on the estimation sample of Online Appendix OA.1.3, are the ability measure throughout.

OA.1.6 Education

Education is highest grade completed (HGC), taken as the woman's maximum over the panel to remove year-to-year reporting variation, and grouped as

$$\text{HSD}: \; \text{HGC} < 12, \quad \text{HSG}: \; 12 \leq \text{HGC} < 16, \quad \text{COL}: \; \text{HGC} \geq 16.$$

College attendance at ages 18–22 is the indicator

$$\mathbf{1}\left\{\exists\, t \text{ s.t. } 18 \leq \text{Age}_{it} \leq 22 \text{ and } \text{HGC}_{it} > 12\right\},$$

equal to one if she reports more than 12 completed years at any interview at those ages.

OA.1.7 Fertility and pregnancy histories

First birth timing. Age at first birth is the year of the first child's birth, from the created child birth-date variables, minus the woman's birth year,

$$\text{AgeAtFirstBirth}_i = \text{BirthYearChild1}_i - \text{BirthYear}_i,$$

set to 99 for women with no recorded birth in the observation window.

Wantedness and contraception. Pregnancy-level indicators from the fertility and contraception modules are aggregated to the model's four-year periods.

(i) Wantedness. For each pregnancy the fertility module records whether the respondent was using a method when she became pregnant, whether she had stopped it or was not using one in order to become pregnant, and whether she wanted to become pregnant at that time (Q9-81: yes; didn't matter; not at that time; not at all). The birth is *unintended* if the answer is "not at that time" or "not at all", or "didn't matter" from a woman who was using a method, and intended if it is "yes", "didn't matter" from a woman using none, or if she had stopped, or was not using, a method in order to conceive, in which case the question is not asked. The targeted moment uses the record of the pregnancy that produced the woman's *first* birth: for a first birth in or before 1982, the first pregnancy of the 1982 retrospective history; for a later birth, the first pregnancy reported at the first interview in the year of the birth or the two years after it, since after 1982 the roster starts at the previous interview.[70]

[70] For the older cohorts a first birth before 1982 is therefore reported retrospectively, and retrospective reports of wantedness can drift toward rationalizing a child who has since arrived (Joyce et al., 2002; Rosenzweig and Wolpin, 1993). The gradient the model targets is not a recall artefact. Among women born in 1962–64, 81 percent of whose first births by 21 were reported within two years of the birth, the top-quartile difference in the probability of an unintended first birth by 21 is -26.3 percentage points (s.e. 2.2), against

Of the 4,580 women in the estimation sample with a first birth at ages 14–41, 4,136 have a matched record and 4,128 of them answer the wantedness question. The moment is, for each quartile, the share of its women whose first birth came at ages 14–21 and was unintended; a woman whose first birth by 21 has no wantedness answer is dropped from the denominator as missing at random.[71]

(ii) Contraception at conception. For each pregnancy p,

$$\text{NoContraception}_{ip} \equiv \mathbf{1}\{\text{no contraceptive method reported at the time of conception}\},$$

where a method is any reported method, including rhythm and withdrawal.

(iii) Contraceptive method (Q9-65 battery). The within-method analysis of Section 2.3 uses the methods a woman reports currently using at each wave, in three tiers ordered by typical-use efficacy: the pill; barrier methods (condom, foam, jelly or cream, suppository, diaphragm, sponge, cervical cap); and behavioral methods (rhythm, withdrawal, natural family planning). IUD and sterilization are excluded, for the reasons given in Section 2. A woman reporting several methods is assigned to the most effective tier. Women whose *only* method is abstinence are excluded, since they have no exposure; abstinence combined with rhythm or withdrawal remains, as a fallible backup.

Classifying a window birth. The within-method outcome is built from child birth-date records rather than from the self-reported contraceptive-failure item (Q9-79), which classifies most imperfect-use failures as deliberate cessation and yields annual failure rates an order of magnitude below typical-use benchmarks. A birth is flagged for a woman–wave if any child's birth date falls in the 12-month window from July of year t through June of $t+1$, which starts after the fielding period, so that the method reported at wave t precedes the outcome. The pregnancy record at the next interview classifies the birth. It is *intended* if she had stopped using, or was not using, a method in order to become pregnant (Q9-80), or if she wanted the pregnancy; it is *unintended* if it came too soon or was not wanted at all, and also if it "didn't matter" while she was using a method: the mistimed-or-unwanted

-21.1 (2.8) among women born in 1957–59, whose reports are almost all retrospective (joint test of equal gradients across three cohort groups, $p = 0.123$; 0.110 with race effects). Within age at first birth, the long-lag reports make low-AFQT mothers look, if anything, less often unintended than high-AFQT mothers (Online Appendix Table OA.8), which would bias the gradient toward zero.

[71] A person-level flag, the maximum over waves of the answer for the first pregnancy *listed at that wave*, is a different object: after 1982 the first listed pregnancy is the first since the previous interview, and such a flag disagrees with the first birth's own report for 14.7 percent of women (Online Appendix OA.3). The reduced-form outcome "unintended birth" of Section 2 is built the same way for every birth: a live birth in the twelve-month window whose pregnancy record, at the first of the next two interviews that has one, says "not at that time" or "not at all". Counting any new pregnancy reported at the next interview instead would mix in pregnancies that did not end in a birth in the window and would lose a quarter of them to the same indexing problem.

versus intended dichotomy has no place for ambivalence (Santelli et al., 2003), and the failure literature reads use itself as the intention to avoid a birth (Kost et al., 2008; Sundaram et al., 2017). Only unintended births count, so the outcome measures the model's birth probability $p_n(\theta, g, e)$; the same classification defines the unintended first birth of Section 3.4 and Online Appendix OA.3. A birth whose pregnancy has no record is missing rather than a failure: the literature counts a conception during use (Kost et al., 2008; Sundaram et al., 2017), and a birth that cannot be tied to a record cannot be shown to be one. Of the 911 window births in the within-method sample, 144 have no record; 409 are planned, wanted, or "didn't matter" to a woman using nothing, and are coded zero; 358 are failures. Treating unrecorded births as missing assumes they are missing at random within cells; counting them as failures would put 18 percent of the lowest quartile's window births and 13 percent of the highest quartile's into the numerator by default, and the two conventions bound the gradient. The within-method sample conditions only on the method reported at t and keeps the woman–wave itself; the sample unconditional on method drops woman–waves at which the interview reports a pregnancy the woman had wanted, since she is not at risk.

Mapping to the model. Childless women exposed to the risk of conception choose a method each period. The targeted contraception moments are use rates,

$$\Pr\Big(\text{AnyMethod}_{it} = 1 \;\Big|\; \text{childless, ever had intercourse, not pregnant,} \\ \text{period } b,\ \text{Educ}_{it},\ \mathbf{1}\{\text{AFQT}_i > \text{median}\}\Big),$$

for periods b covering ages 22–38, with the high-school and college groups split at the median AFQT and dropouts pooled (Section 4). In the model the same object is the probability of choosing a method among childless women at risk, $1 - P(0)$, averaged over the corresponding cells; exposure enters the birth probabilities, not this conditional share.

OA.1.8 Marriage and partner outcomes

Marital status. Marital status is defined annually from marriage start and end dates,

$$\text{Married}_{it} = \mathbf{1}\{t \in [\text{MarriageStart}_i, \text{MarriageEnd}_i)\},$$

with a marriage that has a valid start date and no end date treated as ongoing.

Partner earnings and work. Partner wage-and-salary income comes from the spouse/partner earnings modules, with "no spouse" codes set to 0. Partner weeks and hours worked define partner employment in the earnings-process estimation.

OA.1.9 Labor market outcomes: hours, earnings, employment, experience

Annual hours. Total annual hours and annual weeks worked come from the Work History and Weekly Status files.

Annual earnings. Earnings are annual wage-and-salary income of the respondent and of the spouse or partner, in 2016 dollars.

Interpolation and internal consistency checks. Two checks precede estimation: (1) earnings are set to 0 in years with zero hours; (2) earnings recorded as zero in years with low but positive hours are set to missing and linearly interpolated within woman across years with valid neighbouring values. The checks remove measurement-error spikes while preserving low earnings that low hours corroborate.

Employment and experience. A woman-year is employed if it has (i) at least 26 weeks worked, (ii) average weekly hours above 20 and (iii) real annual wage-and-salary income of at least $10,500 (2016 dollars). Annual experience is $\text{ExpYear}_{it} = \mathbf{1}\{\text{employed}\}$ and cumulative experience is $\text{CumExp}_{it} = \sum_{\tau \leq t} \text{ExpYear}_{i\tau}$.

OA.2 Additional Descriptive Statistics

This appendix reports descriptive tables on birth timing, marriage and labor-market outcomes that supplement Section 2.4. Variables and samples are defined in Appendix OA.1.

OA.2.1 Fertility Timing and Outcomes by Ability Quartile

Table OA.3 reports the life-cycle fertility patterns that are the model's estimation targets, on the model's four-year age grid: first-birth probabilities among women childless at the start of each period (Panel A), the cumulative shares targeted in estimation (Panel A′), and age at first birth, marriage at first birth and completed fertility (Panel B).

OA.2.2 Birth Timing and Education

Table OA.4 reports the share of women whose first birth falls in each model period, by completed education.[72] Early motherhood is concentrated among the least educated: at

[72]Entries are shares of all women in the education group. Only the displayed age bins are reported, so columns need not sum to one; the residual is first births after the last bin or no observed first birth.

Table OA.3. Fertility Timing and Outcomes by Ability Quartile

	Ability Quartile			
	Q1 (Low)	Q2	Q3	Q4 (High)
Panel A: Conditional First-Birth Probability by Model Period (%)				
14–17	28	16	9	3
18–21	48	38	25	16
22–25	38	36	28	22
26–29	25	25	26	29
30–33	18	17	16	26
34–37	10	11	10	12
38–41	2	4	4	5
Panel A′: Share with a First Birth by End of Period (%)				
by 17	28	16	9	3
by 21	63	48	31	18
by 25	77	67	50	36
by 29	82	75	63	55
by 33	85	79	69	66
by 37	87	81	72	71
by 41	87	82	73	72
Panel B: Age at First Birth and Completed Fertility				
Age at first child	20.17	21.67	23.51	25.63
Married at first birth	0.48	0.66	0.80	0.88
At least one child by age 40	0.87	0.82	0.73	0.72
Women	1,455	1,455	1,455	1,455

Notes: NLSY79 estimation sample; age bins are the model's four-year periods. Panel A: probability of a first birth within the period among women childless at its start. Panel A′: cumulative share with a first birth by the end of the period, the moment targeted in estimation. Panel B: mean age at first birth, share married at first birth, and share with a child by age 40. The sample is the 5,820 women of Online Appendix Table OA.1, about 1,455 per quartile; no single cell count applies to Panel A, whose denominator is the women still childless at the start of each period, and Online Appendix Table OA.2 reports the counts block by block.

Table OA.4. Timing of the First Birth by Completed Education (share of all women in the group, %)

	HS Dropout	HS Graduate	College Graduate
14–17	41	14	2
18–21	32	31	6
22–25	12	19	15
26–29	3	9	23
30–33	1	5	14
34–37	1	2	6
38–41	0	1	2
Women	598	4,013	1,209

Notes: Share (%) of women in each completed-education group whose first birth occurred in the indicated model period (four-year age bins); columns do not sum to 100 because some women have no first birth by 41. Estimation sample of 5,820 women.

ages 14–17 the first-birth share is 41 percent for high-school dropouts, 14 percent for high-school graduates and 2 percent for college graduates.

OA.2.3 Early Births and Marriage

Childbearing outside marriage can lower later marriage prospects through economic circumstances and the returns to partner search (Becker, 1991).[73] Table OA.5 reports the probability of ever marrying for women whose first birth precedes first marriage (Panel A) and for women with no premarital birth (Panel B). In Panel A the rate rises with completed education and falls with the age at the premarital birth; in Panel B, which includes the women who remain childless, it is 85–94 percent. The association of premarital fertility with lower marriage rates, strongest for the least educated, is consistent with selection and with marriage-market penalties to premarital childbearing.

OA.2.4 Education, Experience, and Labor-Market Outcomes

Table OA.6 reports wage levels, wage growth and experience by ability quartile and age, with wage statistics computed among employed women as defined in Appendix OA.1. The wage gap between the top and bottom quartiles widens from $3,354 at age 20 to more than $35,000 at age 40 (Panel A); returns to experience are steeper at higher ability (Panel B); and higher-ability women accumulate more experience by a given age (Panel C) and see faster annualized wage growth (Panel D). Log wage changes after a non-work spell are weak or negative for lower-ability women and modestly positive for higher-ability women, and

[73] Grogger and Bronars (2001) report that among initially unwed mothers the median time from first birth to first marriage is eleven quarters, and that within twelve years only 62–64 percent of white and 32 percent of Black mothers had married.

Table OA.5. Probability of Ever Marriage: Premarital Birth vs. No Premarital Birth

	Education Outcome		
Age at first birth	HS Dropout	HS Graduate	College Graduate
Panel A: Premarital First Birth			
14–17	67 (132)	81 (343)	83 (18)
18–21	58 (79)	75 (441)	72 (32)
22–29	46 (28)	59 (182)	69 (26)
Panel B: No Premarital Birth			
All ages	94 (325)	89 (2,848)	85 (1,088)

Notes: Share (%) ever married during the survey window; number of women in parentheses. Panel A: women whose first birth precedes their first marriage, by age at first birth. Panel B: women with no birth before their first marriage—those whose first marriage precedes their first birth and those who remain childless. Women whose marriage–birth order is not recorded are excluded. Education is completed education.

time out of the labor force after the first birth rises with ability (Panel E): high-ability women face both steeper returns to continuous experience and larger career costs from early childbearing.

Table OA.6. Descriptive Statistics by Ability: Labor Market Outcomes

	Ability Quartile			
Outcome	Q1 (Low)	Q2	Q3	Q4 (High)
Panel A: Wage at Given Age (2016 \$, workers only)				
Age 20	19,705	21,546	22,821	23,014
Age 25	23,920	27,856	32,093	38,487
Age 30	27,801	33,639	38,757	49,210
Age 40	30,329	39,877	46,281	65,610
Panel B: Wage Growth with Experience (log points ×100, from the first year worked)				
5 years worked	26	35	44	58
10 years worked	40	52	64	79
15 years worked	48	62	74	90
Panel C: Cumulative Work Experience (years)				
At age 25	1.80	3.12	3.77	3.89
At age 30	3.51	5.95	7.10	7.38
At age 40	7.69	12.38	14.42	14.19
Panel D: Annualized Log Wage Growth Rate (%)				
Average	2.70	3.13	3.50	4.33
Panel E: Wage Changes Around Non-Work Spells				
Δ log wage after 1-year gap	0.02	−0.01	0.00	0.04
Δ log wage after 3-year gap	−0.02	−0.03	0.05	0.05
Δ log wage after 5-year gap	−0.01	−0.12	0.02	0.07
Time out after first child (years)	0.31	0.50	0.56	0.60
Women	1,473	1,483	1,493	1,490

Notes: The 5,939 NLSY79 women with a valid AFQT; "Women" gives the number per quartile. A year counts as worked with more than 26 weeks, more than 20 hours a week and annual earnings of at least \$10,500 (2016 dollars). Panel A: mean annual earnings of women working at each age (246–873 woman-years per cell). Panel B: $\ln w$ in the year the woman reaches 5, 10 or 15 years worked minus $\ln w$ in her first year worked (547–1,155 women per cell). Panel C: cumulative years worked, all women observed at that age (436–1,273 per cell). Panel D: mean change in log earnings between consecutive years worked. Panel E: change in log earnings from the last year worked before a spell out of work to the first year worked after it, by the spell's length in survey years (370–650 spells per cell); time out after the first child is weeks not worked in the five years after the first birth, divided by 52 (4,681 mothers).

OA.3 What the Unintended-Birth Moment Measures

The two wantedness blocks—the four per-woman shares and the eight composition cells of panel C of Table OA.36—carry 8 and −3 percent of the nested-test statistic, 5 percent together (Table 4). This appendix describes what the survey's wantedness question records, how the moment is built on both sides, and the per-woman counts behind Figure 1.

The survey question. Each woman's first birth is matched to its own pregnancy record (Online Appendix OA.1): of the 4,580 women in the structural sample with a first birth at ages 14–41, 4,136 are matched and 4,124 have a classifiable status at conception (the twelve without one are dropped here; ten of them answer the wantedness question and are kept in the targeted moment, which in turn drops the six women with a status but no answer, so it uses 4,128). That status (Q9-78 to Q9-80) falls into five groups: the method in use failed; she had stopped using a method, not in order to conceive; she had used none since the previous interview and was not trying to conceive; she had stopped in order to conceive; she was using nothing because she wanted a birth. The wantedness question (Q9-81) is asked of the first three groups only, so every mistimed or unwanted birth belongs to one of them.

Selection into the matched sample is not what produces the gradient. The match rate is 90.6, 90.5, 90.1 and 90.0 percent from the lowest to the highest quartile, and the hypothesis that it is equal across quartiles is not rejected ($p = 0.950$). Nor do the unmatched births have room to overturn the gradient: the top-quartile difference in the unintended share is -16.1 percentage points among matched births, and assigning every unmatched birth to unintended in one quartile and to intended in the other—the worst case in each direction—leaves it between -24.3 and -4.5 points. The bounds exclude zero.

Three facts follow. *First, unintended first births are mostly births to non-users.* 73.2 percent of mistimed or unwanted first births were conceived with no method in use, 85.1 percent in the lowest quartile and 53.0 percent in the highest, and 54 percent by women who had used no method since the previous interview; method failures account for 27 percent. *Second, the ability gradient is in non-use, not in failure.* The unintended share of first births falls by 16.0 percentage points from the lowest to the highest quartile. Births to women who had never used a method contribute 21.6 points of that fall, births after stopping 1.8, and failures -7.4: failures are a larger share of first births in the top quartile (16.6 percent) than in the bottom (8.5), because top-quartile women use methods more. *Third, the gradient is a timing composition.* Among teenage first births the unintended share lies between 63 and 70 percent in every quartile, and adding single-year indicators for age at first birth moves the top-quartile difference in the unintended share from -0.160 (s.e. 0.021) to $+0.0385$ (s.e. 0.0218). Low-ability women report more of their first births unintended because they have them earlier, when first births in every quartile are more often unintended.

The moment on both sides. A birth while a method is in use is a failure, the smallest of the three components and the only one whose gradient runs the other way, so a moment built on failures would ask the model to reproduce through failures a gradient the data place in non-use. The data moment is instead built from the first birth's own report (a person-level flag from the pregnancy listed first at each wave disagrees with it for 14.7 percent of women) and counts women rather than births. In the model a first birth is unintended if, at the start of its period, $\Delta V + \sigma_\omega \varepsilon^\omega < 0$ at the realized value of the transitory component, whatever method the woman chose (Section 3.4); a woman with a negative realized value who uses nothing because the cost of a method and her taste shock outweigh the expected cost of a birth is the survey's "never used, not wanting" respondent, unintended on both sides. The sign of ΔV, with no band of indifference around zero, is a fair counterpart of the survey question because the answer "didn't matter" is rare, 3.7 percent of first births.

The estimated scale of the transitory component is $\sigma_\omega = 0.212$, in the units of ΔV. At age 18, among women childless and at risk, the systematic part of ΔV has a standard deviation of 0.069 across all women and between 0.044 and 0.056 within a quartile, so the shock is about 3 times the dispersion the state produces and 4 times the within-quartile dispersion. The ability gradient is in the means, -0.046 in the lowest quartile against -0.144 in the highest at 18, a spread of 0.115, with shares below indifference of 0.58 and 0.75: for an individual woman the classification is largely the draw, for a quartile it is the mean. The restricted model needs a larger shock, $\sigma_\omega = 0.287$, and a wider spread of means (-0.120 to -0.343) to reach a gradient four fifths the size of the data's (Section 5.3).

Per woman. The unintended share of births divides by births, which move with ability too; the object the paper needs counts women, and Table OA.7 does so on the same women throughout. Unintended births are 116 percent of the lifetime gap in first births, because intended births run slightly the other way; they are 62 percent of the gap by 17 and 51 percent by 21. The gradient holds within race and ethnicity: the top-to-bottom-quartile difference in unintended first births by 21 is -15.9 percentage points among non-Black non-Hispanic women (s.e. 2.4), -28.9 among Black women (s.e. 4.0) and -10.5 among Hispanic women (s.e. 3.6), while intended first births by 41 do not differ significantly by ability within any of the three groups. The targeted moment is the per-woman share at the age where the gap is: the share of each quartile's women whose first birth came by 21 and was unintended, 32.5 percent in the lowest quartile and 10.2 percent in the highest. The eight composition cells of panel C of Table OA.36 are the second target; Section 4.2 describes what each block identifies and Section 5.3 how the two models fit them.

OA.3.1 The Use Margin

The decomposition of Section 2.3 needs the use margin measured on women at risk rather than on births. Table OA.9 estimates a linear probability model with the indicator for using any method as the outcome, the controls of the birth-hazard regressions of Section 2.3, and the woman–wave panel of the seven contraception waves. The item is asked of women who have had intercourse and are not pregnant, and it begins in 1984, so the sample is 25,009 woman–waves.

The raw share reporting a method rises from 62.1 percent in the lowest AFQT quartile to 75.9 percent in the highest. With controls the top-quartile difference is 0.147 (s.e. 0.013), and unlike the within-method gradient it survives race: 0.132 (s.e. 0.014) with race fixed effects, 0.138 (s.e. 0.020) among non-Black non-Hispanic women, 0.139 (s.e. 0.034) among

Table OA.7. First births per woman, intended and unintended, by AFQT quartile

	Q1	Q2	Q3	Q4	All	Q4 minus Q1
First birth by age 17 (% of women)						
Any	26.7	15.6	8.3	3.2	13.4	-23.6 (1.3)
unintended	16.7	10.8	5.8	2.2	8.8	-14.5 (1.1)
intended	10.0	4.9	2.5	1.0	4.5	-9.0 (0.8)
First birth by age 21 (% of women)						
Any	61.0	45.8	29.8	17.3	38.2	-43.7 (1.7)
unintended	32.5	25.3	17.0	10.2	21.1	-22.3 (1.5)
intended	28.5	20.5	12.9	7.1	17.1	-21.4 (1.4)
First birth by age 25 (% of women)						
Any	75.1	64.6	48.1	33.9	55.2	-41.3 (1.7)
unintended	37.0	31.5	23.2	15.8	26.8	-21.2 (1.6)
intended	38.1	33.1	24.9	18.1	28.4	-20.0 (1.7)
First birth by age 41 (% of women)						
Any	86.3	80.9	71.3	70.0	77.1	-16.3 (1.6)
unintended	40.1	35.6	27.8	21.3	31.2	-18.8 (1.7)
intended	46.2	45.3	43.5	48.7	45.9	2.5 (1.9)
Women	1,326	1,335	1,348	1,347	5,356	

Notes: Percent of women with a first birth by the age shown, split by the wantedness the woman reported for that pregnancy (unintended: "not at that time" or "not at all", or "didn't matter" while using a method; intended: "yes", "didn't matter" while using none, or stopped contraception to conceive). Sample: every woman in the structural sample; a woman whose first birth by the age shown has no wantedness record is dropped for that age, so "any" is the sum of the two rows on a common sample of 5,356 women and differs slightly from the first-birth moments of Section 4, which use all 5,820 (by 21: 61.0 against 63.1 percent in the lowest quartile, 17.3 against 18.5 in the highest). Last column: Q4 minus Q1, robust standard error in parentheses.

Table OA.8. The within-age wantedness gradient by the lag between a first birth and its report

	Years between the first birth and its report		
	0–2	3–5	6 or more
Q4 minus Q1	-0.004	0.145	0.209
	(0.025)	(0.060)	(0.095)
Share unintended	0.337	0.563	0.624
First births	3,003	739	386
Quartile by lag interactions, joint p-value (age at first birth and cohort effects): 0.092; with race effects: 0.058			

Notes: First births at 14–41 with a matched pregnancy record and a wantedness answer. Columns: years between the birth and the interview that reports it; births before 1980 are reported in the 1982 retrospective history, later ones within two years. Q4 minus Q1: top-quartile coefficient in a linear probability model of an unintended first birth on AFQT quartile and single-year age-at-first-birth indicators, robust standard error in parentheses. Last row: joint test that the quartile coefficients are equal across the three lag groups, with age-at-first-birth and birth-cohort indicators, and with race and ethnicity added. The per-woman comparison by birth cohort is in the text.

Table OA.9. Use of Any Contraceptive Method by AFQT Quartile

	(1) Pooled	(2) + race FE	(3) Non-Black non-Hisp.	(4) Black	(5) Hispanic
AFQT Q2	0.065***	0.059***	0.076***	0.049**	0.060**
	(0.012)	(0.012)	(0.020)	(0.019)	(0.026)
AFQT Q3	0.107***	0.096***	0.119***	0.081***	0.070**
	(0.012)	(0.013)	(0.020)	(0.023)	(0.032)
AFQT Q4	0.147***	0.132***	0.138***	0.139***	0.194***
	(0.013)	(0.014)	(0.020)	(0.034)	(0.035)
Observations	25,009	25,009	14,084	6,966	3,959

Notes: LPM. Unit: woman × wave, 1984–1992, women aged 18–35 not reporting a pregnancy they had wanted at that interview. Outcome: reports using any contraceptive method. Controls: education group, parity, age at the wave, marital status at the previous interview, window length and wave fixed effects. Columns (3)–(5) estimate the same regression within each NLSY79 screener group. Standard errors clustered by respondent. Reference: AFQT Q1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Black women and 0.194 (s.e. 0.035) among Hispanic women. Two limits remain. The item records whether a method was used, not how consistently, so it cannot separate a woman who chose not to contracept from one who meant to and did not manage it. And it conditions on reported intercourse, which itself rises with ability at these ages (Online Appendix OA.6.4).

OA.4 Education and the Ability Gradient

This appendix defines the controls of Table 1 and reports the two further pieces of evidence summarized in Section 2.3: a prospective design that uses only what was measured in 1979, before the births, and the within-education regressions on adult women.

Definitions for Table 1. The unit is a childless woman × annual interview t, 1979–1993, and the outcome is a first child born in calendar year $t + 1$; women whose first birth has no pregnancy record are left out of the unintended and intended columns. All columns include age and year fixed effects. Enrollment status at t takes four values: in high school, in college, out of school without a diploma, out of school with one; in columns 3–5 every girl is enrolled in high school, so only the highest grade completed at t enters. The 1979 expectations are the schooling the respondent expected to complete (SCHOOL-31) and the number of children she expected (FER-3), entered with the Rotter locus-of-control score and parents' highest grade, each with a missing indicator.

Table OA.10. First Births by 18 and 21 Among Girls Childless at 14–16 in 1979

	% of women		Q4 minus Q1, percentage points			
	Q1	Q4	(1)	(2)	(3)	(4)
First birth by 18	26.4	4.8	-21.9 (2.4)	-17.8 (2.8)	-15.0 (2.9)	-13.6 (3.1)
unintended	17.6	2.7	-14.9 (2.0)	-14.2 (2.5)	-13.3 (2.5)	-10.2 (2.6)
First birth by 21	50.7	14.5	-36.4 (3.0)	-29.9 (3.6)	-25.4 (3.8)	-23.1 (4.1)
unintended	29.5	7.2	-22.5 (2.6)	-21.7 (3.1)	-20.1 (3.3)	-14.5 (3.4)
intended	21.3	7.4	-14.0 (2.4)	-8.3 (2.7)	-5.3 (2.9)	-8.7 (3.3)
Schooling in 1979: enrollment, grade for age, expected				✓	✓	✓
Expected children, Rotter, parents' schooling					✓	✓
Race and ethnicity						✓
Women				1,803		

Notes: LPM, one observation per woman. Sample: NLSY79 women aged 14–16 and childless at the 1979 interview. Outcomes: first birth by the age shown, and first birth reported unintended or intended on its own pregnancy record, as in Figure 1; a woman whose first birth by that age has no record is left out for that outcome. Entries in columns (1)–(4): top minus bottom AFQT quartile, percentage points, robust standard errors in parentheses. All columns control for age at the 1979 interview. Column (2) adds enrollment status, grade for age (on track, one year behind, two or more) and expected schooling (fewer than 12 years, 12, 13–15, 16 or more), all from 1979; column (3) adds the number of children expected (FER-3), the Rotter score and parents' highest grade; column (4) adds race and ethnicity.

OA.4.1 Girls followed from 1979

Table OA.10 follows the 1,803 women who were 14–16 and childless at the 1979 interview, with every control measured at that interview. Column (2) holds fixed enrollment status, whether the girl was on track for her age or behind, and the schooling she expected to complete (SCHOOL-31: "as things now stand, what is the highest grade or year you think you will actually complete?"). The gradient in first births by 18 falls from -21.9 to -17.8 percentage points, while the gradient in unintended first births by 18 barely moves, from -14.9 to -14.2. With expected fertility, the Rotter score and parents' schooling added (column 3), the gradient in unintended first births by 21 is -20.1 (s.e. 3.3), against -5.3 (s.e. 2.9) for intended ones. The same holds within expected schooling: among girls who expected to finish high school and go no further, 30 percent in the lowest quartile and 13 percent in the highest had an unintended first birth by 21; among girls who expected a college degree, 32 and 5 percent. Race and ethnicity (column 4) reduce the unintended gradient by 21 to -14.5 (s.e. 3.4).

OA.4.2 Adults within completed education

Among adult women, whose schooling is largely complete, I estimate a linear probability model for the probability of a birth separately within each education group:

$$B_{it} = \alpha_e + \sum_{q=2}^{4} \beta_{eq} \, \mathbf{1}[\mathrm{AFQT}_i = q] + \mathbf{x}_{it}' \boldsymbol{\gamma}_e + \delta_t + \varepsilon_{it}, \qquad \text{estimated within education } e, \quad (\mathrm{OA.1})$$

where $B_{it} \in \{0, 1\}$ equals one if woman i had a live birth in the 12-month window following snapshot wave t, $\mathbf{1}[\text{AFQT}_i = q]$ are AFQT quartile indicators (quartile 1 omitted), $\mathbf{x}_{it}$ contains parity at the snapshot year (0/1/2+ prior live births), age-at-wave indicators and marital status at the previous annual interview (never married, married, separated, divorced or widowed), and δ_t are snapshot-year fixed effects; these are the controls of the within-method regressions (Section 2.3). Standard errors are clustered by respondent. The sample is all NLSY79 women aged 18–35 not trying to conceive, pooled across the seven survey waves of Section 2.1. If AFQT matters only through education, $\beta_{eq} = 0$ within each group. Figure OA.1 plots the predicted birth probability by quartile within each education group and Table OA.11 reports the coefficients.

The gradient holds within both non-college groups and is absent among college graduates: the top-quartile coefficient is -4.5 percentage points among high-school dropouts ($p < 0.05$) and -2.9 among high-school graduates ($p < 0.01$), while among college graduates the Q3 and Q4 coefficients are positive with standard errors of about 2.5 pp. Panel B of Table OA.11 shows the same pattern for unintended births, those the woman reported as having come too soon or not wanted at all. Pooling the education groups, the top quartile has 1.1 fewer unintended births per hundred women a year (s.e. 0.3), about half the raw rate (1.6 against 3.3); marital status accounts for most of the difference from the gap in births of any kind, because married women have more births but fewer unintended ones. These estimates are unchanged by attrition re-weighting and strengthen when women whose AFQT was measured after a first birth are dropped (Additional Results, Section S.3). Family background does not account for them: adding parents' highest grade, log 1979 family income and 1979 poverty status leaves the top-quartile coefficient at -0.047 (s.e. 0.019) among dropouts and -0.024 (s.e. 0.006) among high-school graduates. A race control cuts both by more than half: with race added in its place the two coefficients are -0.019 (s.e. 0.018) and -0.012 (s.e. 0.006), and with the two blocks together -0.026 and -0.010. Under the fuller specification of Table OA.13, which also holds enrollment and the age-adjusted score, both coefficients are -0.004 (s.e. 0.021 and 0.007). Online Appendix OA.5.2 separates the two blocks across every sample and outcome and decomposes the movement variable by variable. The first-birth timing gradient on which the model is estimated is present and monotone within every race and ethnicity group (Online Appendix OA.5.6), though not of the same size in each; the residual 18–35 hazard gradient is concentrated among Hispanic women and, less precisely, Black women. The adult gradient within education is thus smaller than the teen gradient of Table 1 and, unlike it, does not survive a race control: most of the ability gap in first births is realized before schooling is, and it is there that the gradient survives family background and race together (columns 4–5 of Table 1).

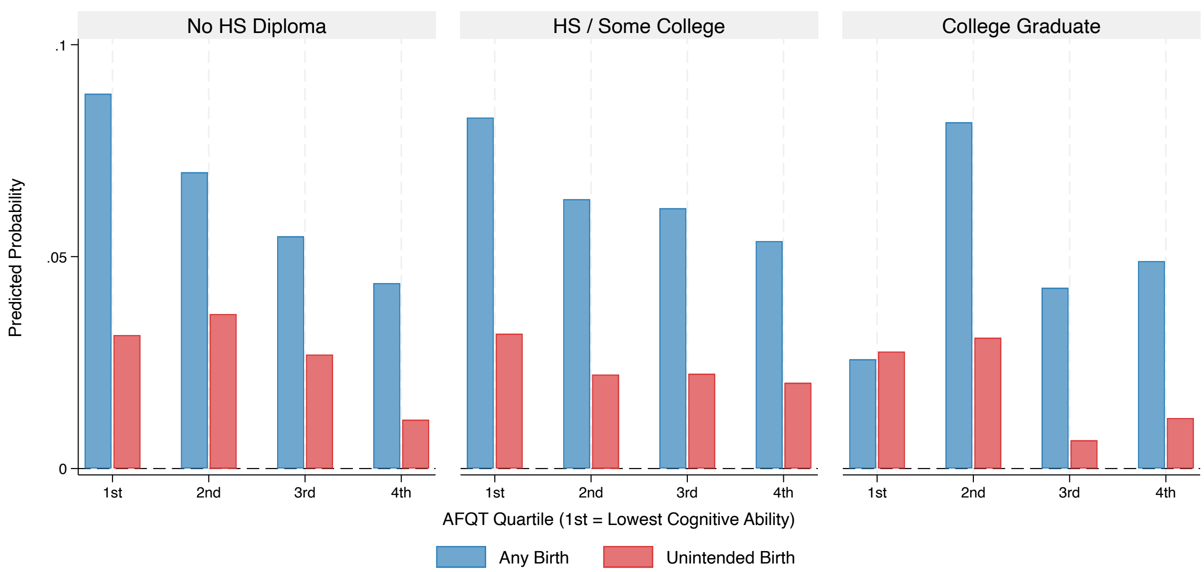


Figure OA.1. Birth probability by AFQT quartile and education group, unconditional on contraceptive method. Denominator: NLSY79 women aged 18–35 not actively trying to conceive, pooled across seven survey waves. Blue bars: any live birth in the 12-month window. Red bars: a live birth in the same window that the woman reported, on its pregnancy record at the next interview that has one, as having come too soon or not wanted at all. Births with no pregnancy record at either of the next two interviews (21 percent; 26 in the lowest quartile, 15 in the highest) are left out of the red bars rather than counted as intended. LPM-predicted probabilities from Equation (OA.1), evaluated at AFQT quartile means; controls set at sample averages within each education group.

Table OA.11. Within-Education Birth Probability by AFQT Quartile (LPM, 1982–1992)

	(1) No HS Diploma	(2) HS / Some College	(3) College
Panel A: Any birth			
AFQT Q2	-0.019*	-0.019***	0.056**
	(0.010)	(0.005)	(0.028)
AFQT Q3	-0.034**	-0.021***	0.017
	(0.015)	(0.005)	(0.025)
AFQT Q4	-0.045**	-0.029***	0.023
	(0.018)	(0.006)	(0.024)
Observations	5,251	22,342	4,498
Panel B: Unintended birth			
AFQT Q2	0.005	-0.010***	0.003
	(0.007)	(0.003)	(0.024)
AFQT Q3	-0.005	-0.009***	-0.021
	(0.011)	(0.003)	(0.021)
AFQT Q4	-0.020*	-0.012***	-0.016
	(0.011)	(0.004)	(0.021)
Observations	5,149	22,023	4,476

Notes: LPM estimated separately within each education group. Unit: woman $\times$ snapshot wave (1982–1992). Panel A outcome: any live birth within a fixed 12-month window (July of t through June of $t+1$). Panel B outcome: any birth reported as unwanted or mistimed. Sample: women not trying to conceive, unconditional on contraceptive method. Controls: parity, age-at-wave dummies, marital status at the previous interview, and snapshot-year FE. SE clustered by respondent. Reference: AFQT Q1.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

OA.5 Robustness of the Within-Method Gradient

This appendix checks the within-method gradient of Section 2.3 against additional controls and sampling weights in equations (OA.1) and (2) (Appendix OA.5.1), separates the family-background controls from the race controls (Appendix OA.5.2), checks the timing of the AFQT relative to first births and attrition by ability (Appendix OA.5.3), adds the non-cognitive scales (Appendix OA.5.4), checks the classification of the failure outcome (Appendix OA.5.5), and estimates the gradient within race and ethnicity (Appendix OA.5.6) and without conditioning on method (Appendix OA.5.7). The Additional Results (not for publication, Section S.3) add the descriptive relationship between ability and family background.

OA.5.1 Additional Controls and Sampling Weights

Table OA.12 reports the baseline within-method regressions of Section 2.3, whose coefficients the main text quotes.

Table OA.12. Within-Method Birth Probability by AFQT Quartile (LPM, 1982–1992): Baseline Estimates

	(1) Pooled	(2) Pill	(3) Barrier	(4) Behavioral
AFQT Q2	-0.013***	-0.011**	-0.024*	-0.006
	(0.004)	(0.005)	(0.013)	(0.019)
AFQT Q3	-0.011**	-0.011**	-0.024*	0.015
	(0.004)	(0.005)	(0.013)	(0.021)
AFQT Q4	-0.014***	-0.015***	-0.025**	0.017
	(0.004)	(0.005)	(0.013)	(0.023)
Observations	16,087	11,299	3,828	960

Notes: Estimates of equation (2). Linear probability of a birth in the twelve months after the wave among women using contraception, 1982–1992 waves, by AFQT quartile (Q1 omitted). A birth counts only if classified unintended (Online Appendix OA.1.7). Controls: education, parity, age, marital status at the previous interview and wave fixed effects; the pooled column adds method fixed effects. Standard errors clustered by woman.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA.13 varies the controls of the baseline regressions cumulatively, from a specification without age and marital status (column 1) through enrollment status and age-adjusted AFQT quartiles (columns 3–4), unweighted (Panel A) and with the NLSY79 cross-sectional weights (Panel B). The last three columns add the 1979 family-background variables alone (column 5), race and ethnicity alone (column 6), and both (column 7); Appendix OA.5.2 takes up what each block does. Table OA.14 reports the baseline and the fully controlled specification by method. The unconditional counterpart is in Appendix OA.5.7; the within-education and unintended-birth counterparts are in the Additional Results, Section S.3.

The demographic controls leave the within-method gradient where it is: the top-quartile coefficient is −0.016 without age and marital status, −0.014 at the baseline and −0.011 with enrollment and the age-adjusted score. So does family background: adding parents' highest grade, log 1979 family income and 1979 poverty status leaves it at −0.011 (s.e. 0.004) in column (5). The race and ethnicity indicators, added in their place, take it to 0.000 (s.e. 0.004) in column (6), and the two blocks together to −0.001 (s.e. 0.005) in column (7). They are not interchangeable, and it is race that does the work; Appendix OA.5.2 separates them across every sample and outcome. By method (Table OA.14), the pill gradient goes from −0.015 to −0.003 (s.e. 0.005) under the full specification; the barrier and behavioral samples are too small to say. The within-education and unintended-birth gradients behave the same way. These are statements about the annual hazard at ages 18–35 conditional on parity and marital status, which leaves out the teenage years where most of the gap in unintended first births arises. Appendix OA.5.6 takes up what a race control does in these regressions and shows that the first-birth timing gradient the model is estimated on holds within every race and ethnicity group, as does the gradient in unintended first births by 21 counted per woman (Online Appendix OA.3).

I report unweighted estimates and use the weights as a specification test. The NLSY79 design is stratified on race, ethnicity and 1978 family income. Where the stratifying variables enter the regression, unweighted least squares is consistent for a correctly specified conditional mean and, under a conditional information matrix equality, more efficient than weighted least squares (DuMouchel and Duncan, 1983; Winship and Radbill, 1994; Wooldridge, 2001); the baseline omits race on the grounds of footnote 15, so the test rather than the theorem is what licenses the unweighted estimates. A gap between the two estimates signals heterogeneity across strata, not a reason to prefer the weighted estimate (Solon et al., 2015). The moments of the structural model are computed unweighted. The last row of each table reports the DuMouchel and Duncan (1983) test that the weighted and unweighted AFQT coefficients are equal.

Within method, weighting leaves the baseline gradient at −0.012 (s.e. 0.005), and the test does not reject in any column ($p = 0.67$ at the baseline). Births unconditional on method (Appendix OA.5.7) are the exception: the weighted baseline coefficient is −0.019 (s.e. 0.006) against −0.030 unweighted, and the test rejects in every column until race is added. The heterogeneity is the one Appendix OA.5.6 documents, a gradient concentrated among Hispanic and, less precisely, Black women, whom the weights down-weight; Solon et al. (2015) recommend reporting it rather than averaging it away. The object the structural model targets is not affected: the ratio of first births by age 21 between the lowest and highest

Table OA.13. Within-Method Birth Probability by AFQT Quartile: Additional Controls and Sampling Weights (LPM, Pooled)

	(1) Previous	(2) Baseline	(3) + Enrol.	(4) Age-adj.	(5) + Family	(6) + Race	(7) + Both
Panel A: Unweighted							
AFQT Q2	-0.014***	-0.013***	-0.012***	-0.008**	-0.008**	-0.004	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
AFQT Q3	-0.013***	-0.011**	-0.010**	-0.009**	-0.010**	-0.001	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
AFQT Q4	-0.016***	-0.014***	-0.013***	-0.011***	-0.011**	0.000	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Panel B: NLSY79 sampling weights							
AFQT Q2	-0.014***	-0.013**	-0.013**	-0.007	-0.008*	-0.003	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
AFQT Q3	-0.011**	-0.010*	-0.009*	-0.007	-0.007	-0.001	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
AFQT Q4	-0.013**	-0.012**	-0.011**	-0.008*	-0.007	-0.000	-0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Test weighted = unweighted, p-value	0.500	0.665	0.637	0.803	0.623	0.998	0.992
Observations	16,087	16,087	16,087	16,087	16,087	16,087	16,087

Notes: Sample and outcome as in Table OA.12, column (1); all columns include method-group FE. Column (1): education group at the wave, parity, window length and snapshot-year FE. Column (2), the baseline: (1) plus age-at-wave dummies and marital status at the previous interview. (3) adds enrollment status at the wave; (4) replaces the AFQT quartiles by quartiles of AFQT residualized on age-at-test dummies; (5) adds mother's and father's highest grade, log 1979 family income and 1979 poverty status (with missing indicators); (6) adds the race and ethnicity indicators instead; (7) adds both blocks. Panel B weights every column by the NLSY79 cross-sectional weight for the wave. Test row: p-value of the DuMouchel and Duncan (1983) test that the weighted and unweighted AFQT coefficients are equal (joint test on the normalized weight $\times$ AFQT interactions in the unweighted regression augmented with the weight and its interaction with every regressor). SE clustered by respondent. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

quartiles is 4.8 with the 1979 weights against 3.7 unweighted.

OA.5.2 Family Background or Race?

The specification that removes the adult gradient adds two blocks of controls at once: the woman's race and ethnicity, and her family's circumstances in 1979—her parents' highest grade completed, the family's income and whether it was below the poverty line. Reported together, as they are in column (7) of Table OA.13, they invite the reading that socioeconomic status rather than ability accounts for the gradient. The two blocks do different work. Table OA.15 separates them and collects the top-quartile coefficient across the samples and outcomes the paper uses: column (2) adds family background alone, column (3) race and ethnicity alone, and column (4) both.

Family background is not what absorbs the gradient. Within method the top-quartile

Table OA.14. Within-Method Birth Probability by AFQT Quartile and Method: Baseline vs. Full Controls

	Pill		Barrier		Behavioral	
	(1) Baseline	(2) Full	(3) Baseline	(4) Full	(5) Baseline	(6) Full
AFQT Q2	-0.011**	-0.004	-0.024*	-0.012	-0.006	0.013
	(0.005)	(0.004)	(0.013)	(0.011)	(0.019)	(0.019)
AFQT Q3	-0.011**	-0.002	-0.024*	-0.014	0.015	0.035*
	(0.005)	(0.005)	(0.013)	(0.011)	(0.021)	(0.021)
AFQT Q4	-0.015***	-0.003	-0.025**	-0.007	0.017	0.031
	(0.005)	(0.005)	(0.013)	(0.012)	(0.023)	(0.024)
Observations	11,299	11,299	3,828	3,828	960	960

Notes: Sample and outcome as in Table OA.12, columns (2)–(4). "Baseline" is Table OA.12. "Full" is column (7) of Table OA.13, Panel A: enrollment status, age-adjusted AFQT quartiles, race and family background added; unweighted. SE clustered by respondent. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

coefficient is -0.014 (s.e. 0.004) at the baseline and -0.015 (s.e. 0.005) with family background added; among pill users -0.015 and -0.013; for first births of any kind -0.030 and -0.026; for unintended first births -0.011 and -0.009; among high-school dropouts -0.045 and -0.047; among high-school graduates -0.029 and -0.024. In two of the six rows the coefficient is larger in absolute value with family background than without it, and in none of them does it lose significance. The race and ethnicity indicators do what the two blocks do together: column (3) is within a thousandth of column (4) in four of the six rows, and neither column leaves a coefficient significant at the five percent level, apart from the high-school-graduate row in column (3).

The Gelbach (2016) decomposition reaches the same conclusion from the other direction, by assigning the movement between the baseline and the fully controlled specification to the individual covariates (Table OA.16). In the unconditional sample the Black and Hispanic indicators account for 110 percent of the movement in the top-quartile coefficient, and within method for 116 percent—more than all of it in both, because the four family-background variables contribute with the opposite sign. Parents' schooling contributes next to nothing—mother's and father's high-school completion carry coefficients of 0.004 and -0.003, neither distinguishable from zero—and 1979 family income contributes with the wrong sign: conditional on ability, women from higher-income families have *more* births in a given wave, so controlling for income widens the ability gap rather than narrowing it.

What a race control does here is mechanical. Forty-seven percent of lowest-quartile women are Black and seven percent of top-quartile women are, so a race indicator removes

Table OA.15. Family Background and Race Separately: Top-Quartile Coefficient by Sample and Outcome (LPM)

	(1) Baseline	(2) + Family background	(3) + Race	(4) + Both
Within method, pooled	-0.014***	-0.015***	-0.004	-0.005
	(0.004)	(0.005)	(0.005)	(0.005)
Oral contraceptive	-0.015***	-0.013***	-0.004	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)
Any first birth	-0.030***	-0.026***	-0.010*	-0.010*
	(0.005)	(0.005)	(0.005)	(0.006)
Unintended first birth	-0.011***	-0.009***	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.004)
No HS diploma	-0.045**	-0.047**	-0.019	-0.026
	(0.018)	(0.019)	(0.018)	(0.019)
HS / some college	-0.029***	-0.024***	-0.012**	-0.010
	(0.006)	(0.006)	(0.006)	(0.006)
Parents' schooling, 1979 income, poverty		✓		✓
Race and ethnicity			✓	✓

Notes: Each cell is the AFQT top-quartile coefficient from a separate linear probability model, with the standard error below. Rows 1–2 are the within-method failure hazard of equation (2); rows 3–4 the unconditional first-birth and unintended-first-birth hazards; rows 5–6 the within-education first-birth hazard of equation (OA.1) inside the two non-college groups. Every column holds the baseline controls of the row's own equation, so that column (1) reproduces the estimate reported in the main text or in Appendix OA.4. Column (2) adds mother's and father's highest grade, log 1979 family income and 1979 poverty status, with missing indicators; column (3) adds the Black and Hispanic indicators instead; column (4) adds both. SE clustered by respondent. Reference: AFQT Q1; non-Black non-Hispanic.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the between-group component of AFQT by construction. The AFQT is measured before labor-market entry, and the premarket-skill literature reads that between-group component as skill rather than as a confounder (Carneiro and Heckman, 2003; Neal and Johnson, 1996), though Carneiro et al. (2005) caution that scores adjusted for schooling at the date of the test are not cleanly premarket; footnote 15 states why the baseline excludes race on those grounds. Whether the ability gradient is a race gradient in disguise is a separate question, and it is answered the way the paper answers the same question for education—within groups. Appendix OA.5.6 does so: the first-birth timing gradient on which the model is estimated is present and monotone within every race and ethnicity group, as is the gradient in unintended first births by age 21 counted per woman (Online Appendix OA.3). The within-method failure gradient is the exception, and Appendix OA.5.6 reports it as one.

Table OA.16. Gelbach Decomposition of the Change in the Top-Quartile Coefficient (Unconditional Sample)

	Contribution to Q4 gap	Share of gap
Mother completed HS	0.0010	-5.7%
Father completed HS	-0.0017	9.2%
Log family income 1979	0.0038	-20.8%
In poverty 1979	-0.0019	10.6%
Black	-0.0153	83.9%
Hispanic	-0.0048	26.2%
Family background (four rows above)	0.0012	-6.5%
Race and ethnicity	-0.0201	110.1%
Total (baseline minus controlled)	-0.0189	100%

Notes: Each row is $\hat{\gamma}_k\hat{\beta}_k$, where $\hat{\gamma}_k$ is the top-quartile coefficient in a regression of background variable k on the AFQT quartile dummies and the baseline controls, and $\hat{\beta}_k$ its coefficient in the fully controlled specification (Gelbach, 2016). Rows sum to the difference between the top-quartile coefficients with and without the background block, up to the missing-indicator terms. The within-method version is in the Additional Results, Section S.3.

OA.5.3 AFQT Timing and Attrition

The ASVAB was administered in summer 1980. Table OA.17 reports how many women had a first birth in or before 1980, by quartile: 27 percent overall, 45 percent of the lowest quartile and 12 percent of the highest, so the concern that a birth could have depressed a subsequently measured test score applies disproportionately to low-ability women. Excluding them *strengthens* every gradient—the unconditional top-quartile coefficient goes from -0.030 to -0.035, the within-method one from -0.014 to -0.015—which is what one expects if the excluded women are those whose early births are least informative about fertility control. Table OA.18 reports the share of women with a valid AFQT score interviewed in each analysis wave and later, by quartile, among the subsamples the NLSY79 retained (the military subsample was dropped in 1985 and the supplemental poor-white subsample in 1991; both have atypical ability distributions and are excluded here). Retention differs across quartiles in one wave and in the long run, and in opposite directions. In 1982 the lowest quartile is retained 2.1 percentage points less often than the highest ($p < 0.01$); from 1984 to 1992 the gap is under one point and never distinguishable from zero; and from 1994 onward it turns positive, reaching 3.9 points in 2010 ($p < 0.05$) and 3.1 in 2018, so that late attrition falls more heavily on *high*-ability women. Neither pattern moves the estimates. Table OA.19 re-weights the baseline regressions by the inverse probability of being interviewed in the wave: the within-method top-quartile coefficient is -0.014 weighted and unweighted alike, and no coefficient in any of the three specifications moves by more than 0.002.

Table OA.17. AFQT Measured After the First Birth

Panel A: Women with a first birth in or before 1980

	Q1 (lowest)	Q2	Q3	Q4 (highest)	All
Women with valid AFQT	1,473	1,483	1,493	1,490	5,939
First birth in or before 1980	659	446	289	185	1,579
share	0.447	0.301	0.194	0.124	0.266

Panel B: Baseline gradients with (All) and without (Excl.) those women

	Within method (eq. 2)		Any birth (eq. 1)		Unintended birth	
	(1) All	(2) Excl.	(3) All	(4) Excl.	(5) All	(6) Excl.
AFQT Q2	-0.013***	-0.013***	-0.019***	-0.022***	-0.006**	-0.010**
	(0.004)	(0.005)	(0.005)	(0.006)	(0.003)	(0.004)
AFQT Q3	-0.011**	-0.012**	-0.024***	-0.032***	-0.009***	-0.014***
	(0.004)	(0.005)	(0.005)	(0.006)	(0.003)	(0.004)
AFQT Q4	-0.014***	-0.015***	-0.030***	-0.035***	-0.011***	-0.015***
	(0.004)	(0.005)	(0.005)	(0.006)	(0.003)	(0.004)
Observations	16,087	12,768	32,091	23,755	31,648	23,449

Notes: Panel B columns: pooled within-method LPM (Table OA.12, column 1); unconditional any-birth and unintended-birth LPM (Table OA.24). Baseline controls; SE clustered by respondent. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA.18. Attrition by AFQT Quartile

Interviewed in	Q1 (lowest)	Q2	Q3	Q4 (highest)	Q1 minus Q4 (SE)
1982	0.970	0.984	0.984	0.992	-0.021*** (0.005)
1984	0.976	0.971	0.979	0.975	0.001 (0.006)
1985	0.961	0.959	0.968	0.964	-0.003 (0.008)
1986	0.952	0.939	0.951	0.945	0.007 (0.009)
1988	0.928	0.918	0.930	0.929	-0.002 (0.011)
1990	0.933	0.914	0.931	0.934	-0.001 (0.010)
1992	0.931	0.928	0.930	0.926	0.006 (0.011)
1994	0.928	0.914	0.918	0.910	0.018 (0.011)
2000	0.857	0.825	0.857	0.853	0.005 (0.014)
2010	0.825	0.805	0.797	0.786	0.039** (0.016)
2018	0.761	0.723	0.727	0.730	0.031* (0.018)

Notes: Share of NLSY79 women with a valid AFQT score interviewed in the year shown. Last column: Q1 minus Q4 difference from a regression on quartile dummies, robust SE in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA.19. Baseline Gradients Re-weighted for Attrition

	Within method (eq. 2)		Any birth (eq. 1)		Unintended birth	
	(1) Unweighted	(2) IPW	(3) Unweighted	(4) IPW	(5) Unweighted	(6) IPW
AFQT Q2	-0.013***	-0.012***	-0.019***	-0.019***	-0.006**	-0.006*
	(0.004)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)
AFQT Q3	-0.011**	-0.011***	-0.024***	-0.024***	-0.009***	-0.009***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)
AFQT Q4	-0.014***	-0.014***	-0.030***	-0.029***	-0.011***	-0.011***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)
Observations	16,087	16,087	32,091	32,091	31,648	31,648

Notes: IPW columns weight each woman-wave by the inverse of the predicted probability of being interviewed in that wave, from a logit on AFQT quartile × wave, race, birth year, mother's schooling and 1979 family income estimated on all women with a valid AFQT score. Otherwise as Table OA.17, Panel B.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

OA.5.4 Cognitive or Non-Cognitive Skill?

Heckman et al. (2006) find that non-cognitive skill matters at least as much as cognitive skill for teenage pregnancy, which raises the question of whether the AFQT quartile stands in for traits the ASVAB does not measure. The NLSY79 fields two non-cognitive scales, and it fields both before the outcome windows of this paper open: the Rotter (1966) locus-of-control score in 1979 and the Rosenberg (1965) self-esteem score in 1980, against contraception and wantedness modules that begin in 1982. Neither scale can therefore be an outcome of the fertility behavior it is used to explain.

Table OA.20 adds them to the baseline specification of each row, one block at a time, and repeats the race column of Table OA.15 as a scale. The non-cognitive blocks do almost nothing. Within method the top-quartile coefficient is −0.014 at the baseline and −0.013 with both scales; for first births of any kind −0.030 and −0.029; for unintended first births −0.011 and −0.010; among high-school dropouts −0.045 and −0.047. Across the six rows the two scales together leave between 89 and 105 percent of the baseline coefficient standing, and in two rows the coefficient is larger in absolute value with them than without. The race indicators, in the same table and on the same samples, leave 26 percent of the within-method coefficient and −7 percent of the unintended-birth coefficient.

Two qualifications keep this from being more than it is. The scales are coarse—four item pairs and ten items—so classical measurement error attenuates how much of the gradient they can absorb, and the comparison is between controls rather than between the latent factors a factor model would recover. What the table rules out is that the ability gradient

Table OA.20. Non-Cognitive Skill: Top-Quartile Coefficient by Sample and Outcome (LPM)

	(1) Baseline	(2) + Rotter	(3) + Rosenberg	(4) + Both	(5) + Race
Within method, pooled	-0.014***	-0.013***	-0.014***	-0.013***	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Oral contraceptive	-0.015***	-0.013***	-0.015***	-0.014***	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Any first birth	-0.030***	-0.030***	-0.029***	-0.029***	-0.010*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Unintended first birth	-0.011***	-0.010***	-0.010***	-0.010***	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
No HS diploma	-0.045**	-0.047**	-0.045**	-0.047**	-0.019
	(0.018)	(0.018)	(0.019)	(0.019)	(0.018)
HS / some college	-0.029***	-0.028***	-0.028***	-0.028***	-0.012**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Rotter locus of control, 1979		✓		✓	
Rosenberg self-esteem, 1980			✓	✓	
Race and ethnicity					✓

Notes: Each cell is the AFQT top-quartile coefficient from a separate linear probability model, with the standard error below; samples, outcomes and baseline controls are those of Table OA.15, so column (1) is identical in the two tables. Column (2) adds the Rotter locus-of-control score fielded in 1979, column (3) the Rosenberg self-esteem score fielded in 1980, column (4) both, and column (5) the Black and Hispanic indicators in their place. Both scores are standardized over the estimation sample, so their coefficients are per standard deviation; a missing score is set to zero and flagged with an indicator, the convention used for parents' schooling in Table OA.15. SE clustered by respondent. Reference: AFQT Q1; non-Black non-Hispanic.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

documented here is an artifact of the non-cognitive traits the NLSY79 happens to measure. It does not rule out a non-cognitive component in the adherence parameter itself, and the paper does not claim otherwise: Section 3.4 reads $c_{\theta g}$ as *fertility-control capacity* rather than cognition alone, which is why the counterfactuals of Section 6 remove the need for adherence instead of raising skill.

OA.5.5 Measuring the Failure Outcome

The failure outcome rests on the classification of each window birth by the pregnancy record at the next interview (Online Appendix OA.1.7), and two of its conventions are choices rather than facts. Table OA.21 reports the within-method gradient under four conventions, from the most inclusive, column (a), to the one the paper uses, column (d).

The conventions differ mainly in the treatment of births whose pregnancy has no record: 16 percent of window births, 18 percent in the lowest AFQT quartile against 13 percent

Table OA.21. The Within-Method Gradient Under Four Conventions for the Failure Outcome

	(1) (a) no record = failure	(2) (b) no record missing	(3) (c) + wanted not a failure	(4) (d) + ambivalence rule
AFQT Q2	-0.018***	-0.015***	-0.014***	-0.013***
	(0.005)	(0.004)	(0.004)	(0.004)
AFQT Q3	-0.019***	-0.014***	-0.012***	-0.011**
	(0.005)	(0.005)	(0.004)	(0.004)
AFQT Q4	-0.022***	-0.018***	-0.015***	-0.014***
	(0.005)	(0.005)	(0.004)	(0.004)
Failures	530	391	368	347
Mean failure rate	0.0327	0.0243	0.0229	0.0216
Observations	16,226	16,087	16,087	16,087

Notes: LPM, equation (2), pooled across method groups with method fixed effects and the controls of Table OA.12. Column (a): a window birth is a failure unless the pregnancy record at the next interview reports it intended, and a birth with no record is a failure. (b): births with no record are missing. (c): births the woman reports she wanted are not failures. (d): "didn't matter" is a failure when a method was in use and not when none was; the classification used in the paper and in the estimation. Failures and the mean failure rate are computed over the estimation sample, so they fall short of the counts in Appendix OA.1.7 by the woman–waves dropped for missing covariates or singleton cells. Standard errors clustered by respondent. Reference: AFQT Q1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in the highest, because the women who miss an interview or leave the block blank are disproportionately of low measured ability. Since the outcome is rare, assigning those births to the numerator moves the estimate: the top-quartile coefficient is -0.022 (s.e. 0.005) in column (a), which counts them as failures, and -0.018 (s.e. 0.005) in column (b), which treats them as missing. Column (c) also stops counting the births the woman reports she wanted, and column (d) adds the rule for "didn't matter": -0.014 (s.e. 0.004).

Neither end of the ladder is free of assumption. Column (a) requires that an unrecorded pregnancy was avoided, and column (d) that unrecorded births are missing at random within the cells of the regression, which the quartile difference above makes unlikely in the direction of the gradient. The two therefore bracket the gradient, between -0.014 and -0.022, significant at either end. The same classification defines the eight within-method failure moments, so the model is estimated on column (d).

OA.5.6 The Ability Gradient Within Race and Ethnicity

Table OA.22 reports the share of women with a first birth by ages 17, 19, 21 and 25 by AFQT quartile separately for non-Black non-Hispanic, Black and Hispanic women, using the pooled quartile cut-points of the main text, and for the largest group also with quartiles drawn within the group. The timing gradient is present, large and monotone in every group: the ratio of first births by 21 between the lowest and highest quartiles is 3.7 among non-Black non-Hispanic women, 3.2 among Hispanic women and 2.3 among Black women. Among non-

Black non-Hispanic women, 18.8% of the lowest quartile have a first birth by 17 against 2.1% of the highest, 15.1% and 1.5% with quartiles drawn within the group. The first-birth timing moments the structural model is estimated on are not a race gradient in disguise.

Table OA.23 does the same for the reduced-form hazards of Section 2, where the picture differs. The 12-month birth hazard at ages 18–35, conditional on education, parity, age, marital status and wave, has a top-quartile coefficient of -0.010 with race fixed effects and 0.002 (s.e. 0.007) among non-Black non-Hispanic women, against -0.011 and -0.032 among Black and Hispanic women; within contraceptive method the coefficients are -0.004, 0.003 (s.e. 0.006), -0.008 and -0.003. Two things reconcile this with Table OA.22. First, the hazard regressions condition on education, parity and marital status at ages 18–35, by which time, for the majority group, the ability gradient has already played out through schooling, marriage and the first birth itself; the timing gradient at 14–21 is what those controls absorb. Second, a race indicator removes the between-group part of the variation in ability along with the racial difference in timing, and AFQT accounts for most of that difference: 74 percent of the Black–white gap in first births by 21 (from 20.3 to 5.2 percentage points) and all of the Hispanic–white gap (from 10.9 to -1.5), the pattern Neal and Johnson (1996) documents for wages, qualified by Carneiro et al. (2005).[74] Restricting the sample to white women, as Francesconi (2002) does, or to the cross-sectional sample (Heckman et al., 2006), would leave the timing and unintended-birth moments unchanged and remove the within-method gradient.

OA.5.7 Unconditional Birth Hazard by Ability

The within-method test conditions on the method to isolate the per-use effectiveness margin. A complementary question is whether the *unconditional* ability gradient in births—combining both the method-selection and per-use margins—is quantitatively important. I estimate a version of equation (2) on the full population of NLSY79 women aged 18–35 who were not actively trying to conceive, pooled across the same survey waves, without conditioning on the method used. The outcome is again a live birth in the fixed 12-month window. In a separate specification, I restrict the outcome to births that the woman reports as unwanted or mistimed, isolating conceptions that represent failures of fertility control.

Table OA.24 reports the results. The unconditional birth gradient is declining in AFQT:

[74] Within-method failure (Panel C) is the exception. Black and Hispanic women using the same method have annual failure rates of 3.9 and 2.3 percent against 1.4 for non-Black non-Hispanic women, as Sundaram et al. (2017) find in the National Survey of Family Growth, and AFQT accounts for little of that gap (from 2.1 to 2.1 percentage points for Black women, 0.5 to 0.5 for Hispanic women). The within-method evidence of Section 2.3 is therefore a pooled fact, as are the eight within-method failure moments of the estimation, which carry 14 percent of the nested test (Section 5.3); the timing and unintended-birth moments hold within every group (Table OA.22 and Online Appendix OA.3).

Table OA.22. First-Birth Timing by AFQT Quartile Within Race and Ethnicity

	Q1 (lowest)	Q2	Q3	Q4 (highest)	Q1/Q4	Women
Non-Black non-Hispanic (pooled quartiles)						
First birth by 17	0.188	0.110	0.046	0.021	8.9	405/ 747/1,074/1,276
First birth by 19	0.398	0.265	0.148	0.066	6.0	
First birth by 21	0.560	0.426	0.270	0.150	3.7	
First birth by 25	0.726	0.637	0.472	0.331	2.2	
Black (pooled quartiles)						
First birth by 17	0.224	0.144	0.074	0.049	4.6	693/ 466/ 242/ 103
First birth by 19	0.433	0.294	0.178	0.155	2.8	
First birth by 21	0.623	0.451	0.302	0.272	2.3	
First birth by 25	0.765	0.642	0.521	0.417	1.8	
Hispanic (pooled quartiles)						
First birth by 17	0.157	0.070	0.045	0.000	–	375/ 270/ 177/ 111
First birth by 19	0.365	0.200	0.209	0.108	3.4	
First birth by 21	0.549	0.363	0.266	0.171	3.2	
First birth by 25	0.757	0.656	0.418	0.315	2.4	
Non-Black non-Hispanic (quartiles drawn within the group)						
First birth by 17	0.151	0.062	0.040	0.015	10.1	876/ 875/ 876/ 875
First birth by 19	0.332	0.195	0.112	0.048	6.9	
First birth by 21	0.490	0.343	0.216	0.123	4.0	
First birth by 25	0.688	0.535	0.425	0.293	2.4	

Notes: NLSY79 women with a valid AFQT score. Share with a first birth by the age shown, by AFQT quartile and NLSY79 race/ethnicity screener group. The first three panels use the pooled quartile cut-points of the main text; the last panel draws quartiles within the non-Black non-Hispanic group. Q1/Q4: ratio of the two shares. Women: number of women in Q1/Q2/Q3/Q4.

Q4 women are 3.0 percentage points less likely to have a birth than Q1 women ($p < 0.01$). The unintended-birth column shows a gradient of the same shape (-1.1 pp for Q4, $p < 0.01$), which is what a difference in how reliably methods are used would produce and is harder to reconcile with a difference in how many births were intended. The unconditional gradient is larger than the within-method one (-1.4 pp), but the two are estimated on different populations: the unconditional sample includes women using no method. The comparison that isolates method selection holds the sample fixed. Among method users, adding method fixed effects moves the Q4 coefficient from -1.2 to -1.4 pp, so method selection compresses rather than generates the ability gradient. The results are consistent with a genuine per-use effectiveness channel: higher-ability women implement a chosen method more consistently and correctly. The structural model developed in Section 3 formalizes this channel as ability-dependent adherence with a method of externally fixed efficacy.

Table OA.23. Reduced-Form Birth Hazards by AFQT Quartile Within Race and Ethnicity (LPM)

	(1) Pooled	(2) Pooled + race FE	(3) Non-Black non-Hispanic	(4) Black	(5) Hispanic
Panel A: Any birth, unconditional on method (eq. 1)					
AFQT Q2	-0.019***	-0.011**	0.001	-0.011	-0.019*
	(0.005)	(0.005)	(0.007)	(0.007)	(0.011)
AFQT Q3	-0.024***	-0.010**	0.003	-0.010	-0.028**
	(0.005)	(0.005)	(0.007)	(0.009)	(0.012)
AFQT Q4	-0.030***	-0.010*	0.002	-0.011	-0.032**
	(0.005)	(0.005)	(0.007)	(0.014)	(0.015)
Observations	32,091	32,091	18,187	8,764	5,140
Panel B: Unintended birth, unconditional on method (eq. 1)					
AFQT Q2	-0.006**	-0.001	0.007*	-0.004	-0.009
	(0.003)	(0.003)	(0.004)	(0.005)	(0.007)
AFQT Q3	-0.009***	-0.001	0.006	-0.001	-0.006
	(0.003)	(0.003)	(0.004)	(0.007)	(0.009)
AFQT Q4	-0.011***	0.001	0.006*	0.006	-0.007
	(0.003)	(0.003)	(0.004)	(0.011)	(0.009)
Observations	31,648	31,648	18,020	8,580	5,048
Panel C: Birth within contraceptive method (eq. 2)					
AFQT Q2	-0.013***	-0.008*	0.001	-0.014**	-0.012
	(0.004)	(0.004)	(0.006)	(0.007)	(0.009)
AFQT Q3	-0.011**	-0.003	0.003	-0.002	-0.005
	(0.004)	(0.005)	(0.006)	(0.010)	(0.011)
AFQT Q4	-0.014***	-0.004	0.003	-0.008	-0.003
	(0.004)	(0.005)	(0.006)	(0.012)	(0.012)
Observations	16,087	16,087	9,708	4,076	2,303

Notes: Panels A and B: sample and controls as in the unconditional hazard regressions of Appendix OA.5.7, i.e. all NLSY79 women aged 18–35 not trying to conceive; Panel C: as in Table OA.12, column (1). Column (1) pooled; column (2) adds race/ethnicity fixed effects; columns (3)–(5) estimate separately within group, keeping the pooled AFQT quartile cut-points. SE clustered by respondent. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA.24. Birth Probability Unconditional on Contraceptive Method (LPM, 1982–1992)

	(1) Any birth	(2) Unintended birth
AFQT Q2	-0.019***	-0.006**
	(0.005)	(0.003)
AFQT Q3	-0.024***	-0.009***
	(0.005)	(0.003)
AFQT Q4	-0.030***	-0.011***
	(0.005)	(0.003)
Observations	32,091	31,648

Notes: LPM. Denominator: all NLSY79 women aged 18–35 not actively trying to conceive, pooled across seven snapshot cohorts ($t \in \{1982, 1984, 1985, 1986, 1988, 1990, 1992\}$). Unconditional on method choice; includes users and non-users. Column 1: any live birth in a fixed 12-month window (July of t through June of $t+1$). Column 2: any birth reported as unwanted or mistimed. Controls: education group, parity, age-at-wave dummies, marital status at the previous interview, and snapshot-year FE. SE clustered by respondent. Reference: AFQT Q1.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

OA.6 Alternative Channels

This appendix takes up the channels other than adherence that could produce an ability gradient in births among contraceptive users: the failure-rate gradient that the method mix alone would predict at published typical-use rates (Appendix OA.6.1); the first-birth rate among women using no method (Appendix OA.6.3) and the ability gradient in coital frequency (Appendix OA.6.4), which test the identifying restriction on baseline birth risk; and a bound on differential abortion (Appendix OA.6.5). The method-selection multinomial logit of Section 2.3 is in Appendix OA.6.2; the ability gradient in men's first parenthood and marital sorting are in the Additional Results (not for publication, Section S.3).

OA.6.1 Method Effectiveness and the Method Mix

Table OA.25 lists every method in the NLSY79 contraception battery with its first-year typical-use and perfect-use failure rate (Trussell, 2011) and the share of women in the bottom and top AFQT quartiles reporting it. Among the reversible methods with appreciable use in this cohort the pill has the lowest typical-use failure rate, and it is used more by low-ability women. Assigning each woman the typical-use failure rate of her most effective reported method and averaging by quartile, the method mix alone predicts a Q1/Q4 failure ratio of 1.06, against an observed birth-rate ratio of 2.72: method choice accounts for a few percent of the gap, and the rest is realized within method. Among all women asked the method question, pill use is nearly the same across quartiles (45 vs. 47 percent) because bottom-quartile women are far more often sterilized (22 vs. 6 percent); among users of reversible methods, the population of Section 2.3, the pill share is higher at the bottom.

OA.6.2 Method-Selection Multinomial Logit

Table OA.26 reports predicted method-choice probabilities from the multinomial logit of Section 2.3. Predicted probabilities are averaged over the sample distribution of education, parity, and wave. The pill-use probability declines from 76.5 percent in the lowest AFQT quartile to 62.5 percent in the highest, while the barrier-method probability rises from 15.5 percent to 29.4 percent. Behavioral method use is roughly constant across quartiles. Figure OA.2 plots these probabilities.

OA.6.3 Baseline Birth Risk Among Non-Users

The model attributes the within-method ability gradient to adherence because baseline birth risk p^0_{be} is common within an education–age cell: in equation (3) an ability gradient in p^0 and one in $c_{\theta g}$ are observationally equivalent given the method. The moment that separates them

Table OA.25. Contraceptive Methods in the NLSY79: Failure Rates and Use by AFQT Quartile

		First-year failure (%)		Share reporting method (%)		
NLSY79 method	Tier	Typical use	Perfect use	Q1 (lowest)	Q4 (highest)	All
Female sterilization	Excluded	0.5	0.5	22.4	5.8	13.5
Male sterilization (partner)	Excluded	0.15	0.10	1.7	4.4	3.4
IUD	Excluded	0.8	0.6	5.2	2.0	3.3
Norplant (implant)	Excluded	0.05	0.05	0.0	0.0	0.0
Depo-Provera (injectable)	Excluded	6	0.2	0.0	0.0	0.0
Oral contraceptive pill	Pill	9	0.3	44.9	46.5	47.0
Diaphragm	Barrier	12	6	1.6	9.5	4.8
Cervical cap	Barrier	16	9	0.1	0.1	0.1
Sponge	Barrier	12	9	0.4	1.9	1.1
Condom	Barrier	18	2	7.0	14.4	10.8
Foam	Barrier	28	18	1.8	2.1	2.0
Jelly / cream	Barrier	28	18	0.8	0.5	0.6
Suppository	Barrier	28	18	1.0	1.3	1.2
Natural family planning	Behavioral	24	0.4	0.3	1.2	0.7
Rhythm / calendar	Behavioral	24	5	1.5	3.7	2.7
Withdrawal	Behavioral	22	4	1.5	3.4	2.8
Douching	Behavioral	85	85	2.1	0.2	0.9
Abstinence	Behavioral	0	0	1.3	3.8	2.3
Any method		85 (no method)		87.9	88.4	88.6
Woman-waves				5,787	6,216	24,331

Notes: Failure rates are the percentage of women experiencing an unintended pregnancy within the first year of use, from Trussell (2011), Table 1 (cervical cap from Trussell 2004; douching is not a method and is assigned the no-method rate). Shares: NLSY79 women aged 18–35 asked the method question in 1982–1992, woman-wave observations; multiple methods may be reported. "Excluded" methods are LARC and sterilization, which the within-method analysis omits (Section 2.1).

is the first-birth rate among childless women using *no* method, who face p^0_{be} whatever their adherence, so that any unmodelled correlate of ability in fecundity or exposure must show up there. The sample is the universe of the contraception screening question (FFER-139), childless women aged 18–33 who were sexually active and not pregnant, which is also the universe behind the contraception moments of Section 4; the method battery behind the within-method regressions is asked only of users. The outcome is a first birth before the next observed wave.

Table OA.27 reports the result. Among non-users the first-birth rate over the interval to the next wave (1.8 years on average) is 4.2 percent in the lowest quartile and 4.7 in the highest, 2.4 and 2.5 percent a year annualized; panel B of Table 2 reports the same rate for the lowest quartile and the regression-adjusted differences of column (2). The top-quartile gradient within education, age and wave is +0.7 pp (s.e. 1.2 pp), a zero with the sign opposite to the one a confound would need; among users over the same waves and controls

Table OA.26. Predicted Method-Choice Probabilities by AFQT Quartile (1982–1992)

AFQT Quartile	P(Pill)	P(Barrier)	P(Behavioral)
Q1 (Lowest)	0.765	0.155	0.080
Q2	0.718	0.193	0.089
Q3	0.687	0.227	0.086
Q4 (Highest)	0.625	0.294	0.081
Observations		15,095	

Notes: Predicted probabilities from multinomial logit estimated on NLSY79 women reporting reversible contraceptive use, pooled across seven snapshot waves (1982–1992). Base outcome: pill. Probabilities averaged over the sample distribution of education group, parity, and wave. LARC and sterilization excluded. The estimation sample is 15,095 woman-waves, against 16,087 in the within-method linear probability model of Table OA.12. The difference is mainly that the logit leaves out woman-waves at which the respondent reported a pregnancy she had wanted, as the unconditional sample does, since her method report then says little about avoiding a birth; the within-method regressions keep them and count a planned birth in the window as no failure.

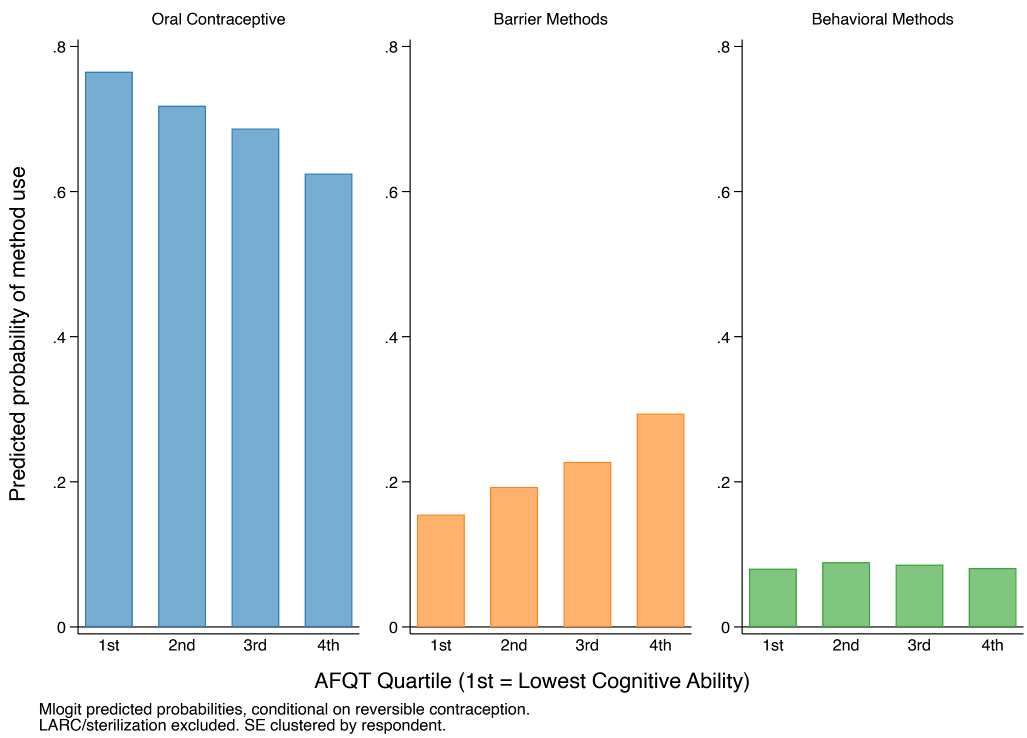


Figure OA.2. Predicted method-choice probability by AFQT quartile. Multinomial logit estimated on NLSY79 women reporting reversible contraceptive use across seven survey waves (1982–1992); base outcome = pill. Predicted probabilities averaged over the sample distribution of education, parity, and wave. LARC/sterilization excluded.

it is −1.9 pp. The gradient appears where a method is in use and not where none is.

The bound is not tight. The non-user cells hold 745 to 938 person-years, so the ninety-

Table OA.27. First-birth rate by AFQT quartile, among non-users and users

	(1) No method, raw	(2) No method, controls	(3) Users, controls
AFQT Q2	-0.0002	-0.0008	-0.0192
	(0.0105)	(0.0108)	(0.0120)
AFQT Q3	0.0021	0.0018	-0.0126
	(0.0101)	(0.0107)	(0.0117)
AFQT Q4	0.0053	0.0071	-0.0187
	(0.0103)	(0.0122)	(0.0118)
Observations	3,372	3,372	5,819
R-squared	0.000	0.008	0.010

Notes: Linear probability models. Unit: woman × screening wave. Outcome: a first birth between the interview and the next observed wave. Sample: childless women aged 18–33 who were sexually active and not currently pregnant, the universe of the contraception screening question (FFER-139 in 1984–1990, Q9-64G/H from 1992). Columns (1)–(2) are women reporting no method, column (3) women reporting a method. Controls in (2) and (3): education group, age at wave, wave, and exposure length. SE clustered by respondent. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

five percent interval around the non-user gradient, from about -1.7 to $+3.1$ pp, contains a baseline-risk gradient as large as the within-method one, -1.4 pp. What the check establishes is that the one moment that can contradict the restriction does not, and that its point estimate has the wrong sign for the alternative, as does the ability gradient in coital frequency (Online Appendix OA.6.4). The baseline-risk model, in which ability moves baseline risk instead of adherence and everything else is re-estimated on the same moments, and the model in which it moves both, are reported in Section 5.4.

OA.6.4 Coital Frequency and the Exposure Margin

Adherence $c_{\theta g}$ is the share of *exposures* on which a method is used as intended, while the moment that identifies it is an *annual* failure rate; the two differ by the number of exposures per year, which the model does not contain. If low-ability women have intercourse more often, they fail more often per year at identical per-exposure adherence, and the estimator books the difference as adherence.

The NLSY79 asks women how many times they had intercourse in the past month in 1984 and 1985 (FFER-137), two of the snapshot waves behind equation (2): 4,764 person-waves and 140 failures, ages 19 to 29, mean frequency 8.3 times a month. Table OA.28 shows that the premise does not hold. Frequency rises with ability: the top-to-bottom gap is 1.92 acts a month (s.e. 0.44, $t = 4.41$) with the baseline controls of Table OA.12, 1.59 (s.e. 0.41) with marital status at the wave added, and it is present within each marital state.

Table OA.28. Coital frequency by AFQT quartile

	Times per month		log(1 + times)	
	(1) Baseline	(2) + Marital at wave	(3) Baseline	(4) + Marital at wave
AFQT Q2	0.9522**	0.7297*	0.1320***	0.1015***
	(0.3927)	(0.3886)	(0.0394)	(0.0386)
AFQT Q3	1.4903***	1.1397***	0.1883***	0.1409***
	(0.4150)	(0.4115)	(0.0405)	(0.0396)
AFQT Q4	1.9197***	1.5900***	0.2685***	0.2242***
	(0.4356)	(0.4300)	(0.0433)	(0.0423)
Observations	4,764	4,764	4,764	4,764

Notes: Dependent variable: times of intercourse in the past month (FFER-137), levels in columns (1)–(2) and $\log(1+\text{times})$ in (3)–(4). Unit: woman $\times$ snapshot wave. Sample: the within-method sample of Table OA.12 restricted to the 1984 and 1985 waves. Controls in (1) and (3): the baseline of Table OA.12 (education group at the wave, parity, age-at-wave dummies, marital status at the previous interview, method group and wave); columns (2) and (4) add marital status at the wave. Exposure length is constant at one year on these waves and is omitted. SE clustered by respondent. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA.29, on identical rows in every column, shows that frequency barely moves failures: one additional act per month adds 0.00028 to the annual failure probability (s.e. 0.00035), and controlling for it moves the top-quartile gradient slightly away from zero, from -0.0224 to -0.0239 with frequency-quintile fixed effects. No coefficient on these rows is individually significant; the columns establish the direction of the movement, not its precision. Attributing the whole top-quartile gradient to exposure would require top-quartile women to have about 81 fewer acts per month than bottom-quartile women, ten times the sample mean, rather than the 1.92 more they report.

The same two objects give a sharper statement. With ν exposures per year and p the annual within-method failure rate, the per-exposure rate is $q = 1 - (1-p)^{1/\nu}$, the object the technology in equation (3) parameterizes. At quartile means the annual rate falls from 5.0 to 1.9 percent between the bottom and top quartiles, a ratio of 0.391, while the per-exposure rate falls from 6.2 to 1.8 per ten thousand acts, a ratio of 0.294: the targeted moment understates the ability gradient in the parameter it identifies, because high-ability women reach a lower annual failure rate on roughly thirty percent more exposures.

Three limits remain. The measure exists only for 1984 and 1985, so the exercise speaks to model periods $t = 3$ and $t = 4$ and not to the teens, where the counterfactual of Section 6 does its work; a one-month recall window is a proxy for the rate of exposure; and frequency is self-reported, although overturning the result would require high-ability women to *over*-report.

Table OA.29. Within-method birth probability, with and without coital frequency

	(1) No control	(2) Linear	(3) Quadratic	(4) Quintile FE	(5) + Marital at wave
AFQT Q2	-0.0217**	-0.0219**	-0.0219**	-0.0223**	-0.0224***
	(0.0087)	(0.0087)	(0.0087)	(0.0087)	(0.0087)
AFQT Q3	-0.0158*	-0.0162*	-0.0162*	-0.0165*	-0.0167*
	(0.0096)	(0.0096)	(0.0096)	(0.0096)	(0.0096)
AFQT Q4	-0.0224**	-0.0229**	-0.0229**	-0.0239**	-0.0240**
	(0.0094)	(0.0094)	(0.0094)	(0.0095)	(0.0095)
Times per month		0.0003	0.0003		
		(0.0003)	(0.0005)		
Observations	4,764	4,764	4,764	4,764	4,764

Notes: Linear probability models for a first birth in the year following the snapshot wave, the outcome of Table OA.12. Every column is estimated on the same 4,764 rows, the within-method sample restricted to 1984 and 1985. Column (1) is the baseline specification of Table OA.12: education group, parity, age-at-wave dummies, marital status at the previous interview, method group and wave. Column (2) adds times per month linearly, (3) adds its square, (4) replaces it by frequency-quintile fixed effects, and (5) adds marital status at the wave to (4). SE clustered by respondent. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

OA.6.5 A Bound on Differential Abortion

Let b_q be the share of quartile-q women with a first birth by age A and A_q their abortions per woman by A. Abortion can generate the birth gradient only if high-ability women abort more: the conception gradient is zero when $A_4 - A_1 = b_1 - b_4$. The NLSY79 records roughly 40 percent of abortions (Jones and Forrest, 1992, Table 5), so reported counts $\tilde{A}_q$ are scaled by $s = 2.5$. Table OA.30 reports, by age 20, 22 and 25, the levels, the gaps and the ratio A_4/A_1 that abortion would need to explain the entire birth gap. By age 25 the birth gap is 0.417 and the reported abortion gap -0.028; explaining the birth gap would require top-quartile women to abort 2.04 times as many pregnancies as bottom-quartile women, against an observed ratio of 1.18. Even assigning every unreported abortion to the top quartile leaves a conception gap of 0.104. Table OA.31 reports the regression version.

Table OA.30. Births, Abortions, and the Bound on the Abortion Channel, by AFQT Quartile

	By age 20	By age 22	By age 25
Panel A: Levels by AFQT quartile			
First birth (share), Q1 (lowest)	0.493	0.633	0.752
First birth (share), Q2	0.348	0.488	0.642
First birth (share), Q3	0.216	0.336	0.474
First birth (share), Q4 (highest)	0.120	0.203	0.336
Reported abortions per woman, Q1 (lowest)	0.082	0.116	0.161
Reported abortions per woman, Q2	0.106	0.143	0.187
Reported abortions per woman, Q3	0.103	0.143	0.179
Reported abortions per woman, Q4 (highest)	0.102	0.144	0.189
Any reported abortion (share), Q1 (lowest)	0.069	0.088	0.115
Any reported abortion (share), Q2	0.091	0.112	0.137
Any reported abortion (share), Q3	0.091	0.120	0.138
Any reported abortion (share), Q4 (highest)	0.088	0.116	0.144
Any pregnancy: birth or abortion (share), Q1 (lowest)	0.520	0.655	0.766
Any pregnancy: birth or abortion (share), Q2	0.398	0.531	0.684
Any pregnancy: birth or abortion (share), Q3	0.283	0.406	0.537
Any pregnancy: birth or abortion (share), Q4 (highest)	0.192	0.286	0.421
Panel B: Q1 minus Q4 gaps			
Birth gap, $b_1 - b_4$	0.373	0.430	0.417
Reported abortion gap, $a_1 - a_4$	-0.020	-0.028	-0.028
Any-pregnancy gap	0.328	0.369	0.345
Conception gap, reported abortions ($s = 1$)	0.353	0.403	0.388
Conception gap, uniform under-reporting ($s = 2.5$)	0.323	0.361	0.346
Conception gap, under-reporting in Q4 only ($s_1 = 1,\ s_4 = 2.5$)	0.200	0.187	0.104
Panel C: Could abortion alone generate the birth gap?			
Required A_4/A_1 for a zero conception gap, uniform s	2.82	2.48	2.04
Required A_4/A_1 if Q1 fully reported ($s_1 = 1$)	5.54	4.70	3.59
Observed a_4/a_1	1.24	1.24	1.18
Women		5,939	

Notes: The 5,939 NLSY79 women with a valid AFQT score, quartiles at the estimation-sample cut-points (Online Appendix OA.1.3). Abortions are dated reports from the 1984–2000 fertility modules (the 1984 module asks about lifetime abortions; later modules since the last interview), de-duplicated by month and year. s is the under-reporting factor applied to reported counts. "Required A_4/A_1" is $(s\,\tilde{A}_1 + b_1 - b_4)/(s\,\tilde{A}_1)$, the ratio of true top- to bottom-quartile abortions at which conceptions would be equal across the two quartiles. The adversarial row assigns all unreported abortions, $(s-1)$ times the sample mean per woman, to the top quartile.

Table OA.31. Any Abortion by Age 25 on AFQT Quartile (LPM)

	(1) Any abortion by 25	(2) + race, cohort	(3) + family background	(4) Abortions by 25 (count)
AFQT Q2	0.022*	0.026**	0.021	0.028
	(0.012)	(0.012)	(0.013)	(0.019)
AFQT Q3	0.023*	0.029**	0.017	0.016
	(0.012)	(0.013)	(0.013)	(0.020)
AFQT Q4	0.030**	0.037***	0.023	0.030
	(0.012)	(0.014)	(0.015)	(0.022)
Women	5,939	5,939	5,939	5,939

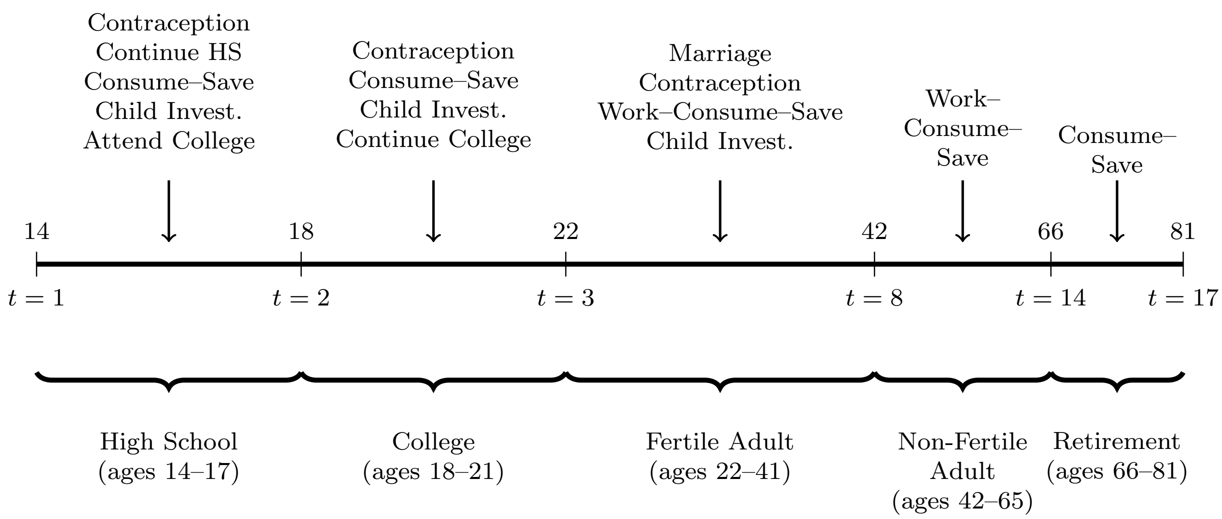
Notes: One observation per woman. Column (4) outcome is the number of reported abortions by age 25. Robust SE. Reference: AFQT Q1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

OA.7 The Model

OA.7.1 Environment, timing, and state space

Section 3 of the paper gives the state, the choices and the young-adult recursion; this appendix states the schooling stages, the post-fertile stages and the transitory wantedness shock. Figure OA.3 shows the life cycle of a woman who attends college; Figures OA.4 and OA.5 show the within-period timing in young adulthood and in the college period (the teen period is analogous).

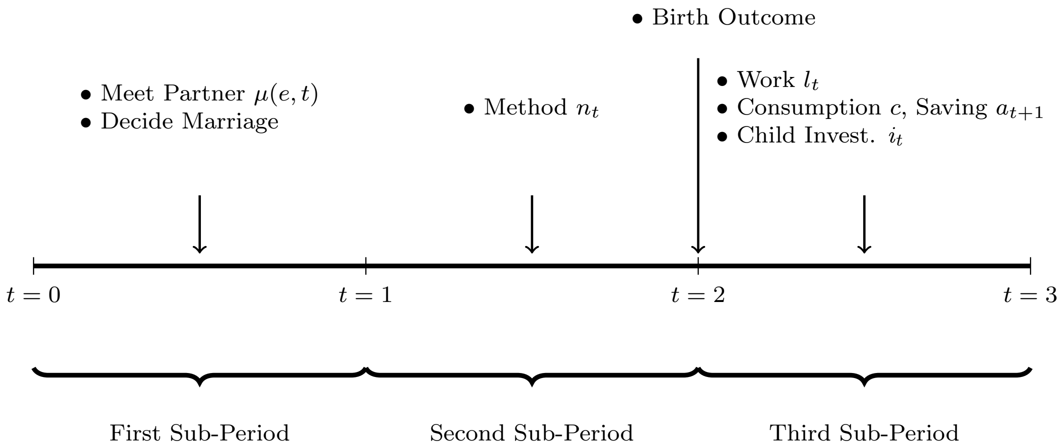
Figure OA.3. Women Attending College Life Cycle



Notes: The figure shows the life cycle of a woman who attends college. A model period is four years and age $= 10 + 4t$, so $t = 1$ covers ages 14–17 and $t = 17$ ages 78–81. The life cycle is divided into five stages: (i) high school ($t = 1$), (ii) college age ($t = 2$), (iii) fertile adult ($t = 3$–7, ending at 41), (iv) non-fertile working adult ($t = 8$–13), and (v) retirement ($t = 14$–17). Ages are shown above the timeline and period indices below it; the horizontal axis is not drawn to scale, since the two schooling periods carry most of the decisions. Above the timeline the figure lists the decisions taken in each stage.

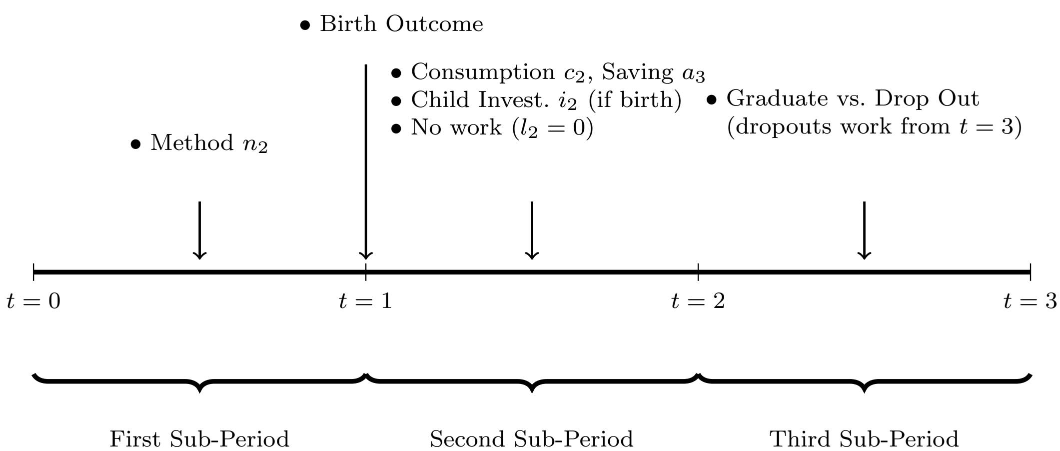
The two schooling stages differ from young adulthood in two respects. There is no marriage sub-stage while a woman is enrolled: women who leave school after $t = 1$ meet partners from $t = 2$ (ages 18–21), college women from $t = 3$. And the schooling continuation decision—continue high school or drop out at $t = 1$, graduate from college or drop out, at the end of $t = 2$—is taken *after* the fertility outcome, so that a birth can raise the cost of continuing; high-school graduates also face the college-entry decision at the end of $t = 1$. Women not enrolled in college at $t = 2$ follow the young-adult timing of Figure OA.4. Sections OA.7.4–OA.7.7 give the Bellman equations stage by stage in the notation fixed below (Section OA.7.1).

Figure OA.4. Childless Women Between Ages 22–41: Within-Period Timing



Notes: Each period is divided into three sub-periods: (i) marriage (if single), (ii) contraception (if childless and fertile), and (iii) labor supply, consumption–saving, and (if a birth occurs) child investment.

Figure OA.5. College-Track Women (Ages 18–21): Within-Period Timing



Notes: Within-period timing at $t = 2$ (ages 18–21) for women enrolled in college. Sub-period (i): choose a contraceptive method n_2, then a stochastic birth may occur. Sub-period (ii): consumption–saving and child investment (if birth); college students do not work ($l_2 = 0$) and receive the student allowance $w_C\, s_\theta$ net of the monetary cost of college TC. Sub-period (iii): at the end of the period, choose whether to graduate or drop out; a birth while enrolled raises the psychic cost of graduating by $|\phi_{k,D}|$, and a dropout enters the labor market at $t = 3$.

Time is discrete in four-year periods, $t \in \{1, \ldots, T\}$, with decisions made for $t = 1, \ldots, T-1$ and terminal period $T = 17$ (age 78), in which agents consume all remaining resources and die. Fertility is feasible through ages 14–41, i.e. for $t \leq T_F = 7$; women can work through age 65 and retire from age 66 onward. Each woman can have at most one child, and the child resides with the household for one period only, so child investment is a one-time choice made in the birth period. The state $\Omega_{it} = \{a_t, \theta_i, e_t, x_t, m_t, k_t, m_k\}$ is defined in Section 3.1; V_t^ℓ denotes the value function in period t and within-period sub-stage $\ell \in \{1, 2, 3\}$, and in non-fertile and retirement periods, which have a single stage, ℓ is suppressed. Choice variables are next-period assets $a_{t+1} \in [0, a_{\max}]$, consumption $c_t \geq 0$, female labor supply $l_t \in \{0, 1\}$,[75] child investment $i_t \geq 0$ (only if $k_t = 2$), and the contraceptive method $n_t \in \{0, 1, 2\}$ (only in fertile periods when $k_t = 1$).

OA.7.2 Income, taxes/transfers, equivalence scales, and experience

Disposable resources. Gross annual household non-asset income, $\tilde{y}_t^0(\Omega_{it}, l_t)$, includes female earnings when working and spousal earnings when married. Disposable annual income follows the approximation to the U.S. tax-and-transfer system in Daruich and Fernández (2024),

$$y_t^a(\Omega_{it}, l_t) = \lambda_\tau \big(\tilde{y}_t^0(\Omega_{it}, l_t)\big)^{1-\tau} + \underline{y}(m_t),$$

and model-period resources aggregate annual resources:

$$y_t(\Omega_{it}, l_t) = 4\, y_t^a(\Omega_{it}, l_t).$$

Budget constraint. The within-period budget constraint is

$$\phi_c\big(m_t, \mathbf{1}_{\{k_t=2\}}\big)\, c_t + a_{t+1} = (1+r)a_t + y_t(\Omega_{it}, l_t) - \mathbf{1}\{k_t = 2\}\, i_t,$$

where $\phi_c(\cdot)$ is the equivalence scale, which depends on marital status and the presence of a newborn.

Experience accumulation. Experience evolves according to

$$x_{t+1} = x_t + \mathbf{1}\{l_t = 1\}.$$

[75] I follow Attanasio et al. (2008) in modeling female labor supply.

OA.7.3 Discrete choices and taste shocks

Discrete alternatives (labor supply, schooling continuation, college entry) carry i.i.d. Type-I extreme value taste shocks. For a generic discrete choice $d \in \mathcal{D}$ with shocks $\varepsilon_t(d)$, scale $\sigma_{\mathcal{D}}$ and choice-specific value net of shocks $v_t(\Omega_{it}, d)$, the ex-ante value is

$$V_t(\Omega_{it}) = \sigma_{\mathcal{D}} \log \sum_{d \in \mathcal{D}} \exp\left(\tfrac{v_t(\Omega_{it}, d)}{\sigma_{\mathcal{D}}}\right),$$

the log-sum inclusive value. It differs from $\mathbb{E}_\varepsilon\big[\max_{d\in\mathcal{D}}\{v_t(\Omega_{it}, d) + \sigma_{\mathcal{D}}\varepsilon_t(d)\}\big]$ by the additive constant $\gamma_E \sigma_{\mathcal{D}}$, where $\gamma_E \approx 0.5772$ is the Euler–Mascheroni constant: the standard Type-I extreme value distribution has mean γ_E rather than zero. That constant is omitted at every discrete-choice node, here and in the code. Because it does not vary with d, it leaves the choice probabilities at a node unchanged; because the scales $\sigma_{\mathcal{D}}$ differ across nodes, it shifts the level of continuation values, and the shift is absorbed into the estimated taste-shock scales and non-pecuniary costs. The same convention applies to the method-choice log-sum of Appendix OA.8.

OA.7.4 Working, non-fertile and retired households (ages 42–81; $t = 8$–17)

After age 41 ($t \geq 8$) fertility risk is absent and no child is present. For $t = 8, \ldots, 13$ (ages 42–65) the household chooses whether the woman works, $l_t \in \{0, 1\}$, consumption and next-period assets. It draws taste shocks $\{\varepsilon_t(0), \varepsilon_t(1)\}$ for labor supply, so that with $v_t(\Omega_{it}, l)$ the choice-specific value net of shocks, $V_t(\Omega_{it}) = \mathbb{E}_\varepsilon\big[\max_{l\in\{0,1\}}\{v_t(\Omega_{it}, l) + \sigma_l \varepsilon_t(l)\}\big]$.

Conditional on l, the choice-specific problem is

$$\begin{aligned} v_t(\Omega_{it}, l) = \max_{a_{t+1}\geq 0,\, c_t \geq 0} & \left\{u(c_t) + \psi_l \, \mathbf{1}_{\{l=1\}} + \beta \, \mathbb{E}\big[V_{t+1}(\Omega_{i,t+1}) \mid \Omega_{it}, l, a_{t+1}\big]\right\} \\ \text{s.t.} \qquad & \phi_c(m_t, 0)\, c_t + a_{t+1} = (1+r)a_t + y_t(\Omega_{it}, l), \\ & x_{t+1} = x_t + \mathbf{1}_{\{l=1\}}, \end{aligned}$$

where gross annual household income is

$$\tilde{y}_t^0(\Omega_{it}, l) = \mathbf{1}_{\{l=1\}}\, w(\Omega_{it}) + \mathbf{1}_{\{m_t=1\}}\, w^h(\Omega_{it}),$$

and disposable model-period resources are

$$y_t(\Omega_{it}, l) = 4\left[\lambda_\tau \big(\tilde{y}_t^0(\Omega_{it}, l)\big)^{1-\tau} + \underline{y}(m_t)\right].$$

Here $w(\Omega_{it})$ is the woman's wage as a function of education, experience and the other state variables, and spousal labor income $w^h(\Omega_{it})$ is received only when married.[76]

From age 66 onward ($t = 14, \ldots, 16$) the household is retired: labor supply is zero, the only intertemporal choice is savings, and gross annual non-asset income is the woman's own Social Security benefit plus the spousal benefit when married, $\tilde{y}_t^0 = ss_t(e_t) + \mathbf{1}_{\{m_t=1\}}\, ss_t^h(e_t)$, both depending on education only. The problem is the one above with $l_t = 0$ and ψ_l absent; in the terminal period $t = T = 17$ agents consume all remaining resources and die.

OA.7.5 Young adulthood (ages 22–41; $t = 3$–7)

In young adulthood schooling is complete (e_t is fixed) and marriage-market and fertility risk are active until $t = T_F = 7$. The three sub-stages—marriage if single, contraception if childless, then labor supply, consumption–saving and child investment—are stated in Section 3.2, with $j \in \{k, nk, ok\}$ indexing the child-status outcome ($k_t = 2, 1, 3$).

The transitory wantedness shock. Writing $\Delta V_t = V_t^{3,j=2}(\Omega_{it}) - V_t^{3,j=1}(\Omega_{it})$ and $\tilde{\Delta V}_t = \Delta V_t + \sigma_\omega \varepsilon_{it}^\omega$ for its realized value, the contraception sub-stage value is $V_t^{3,j=1}(\Omega_{it}) + \mathbb{E}_{\varepsilon^\omega}\left[\sigma_n \log \sum_{n=0}^2 \exp\{(p_n(\theta_i, g(t), e_t)\, \tilde{\Delta V}_t - \kappa_n)/\sigma_n\}\right]$, where the outer expectation is computed by seven-point Gauss–Hermite quadrature and the choice probabilities are the logit shares of Section 3.4 evaluated at $\tilde{\Delta V}_t$. In simulation each woman draws her own ε_{it}^ω, chooses a method at the realized value, and a first birth in the period is recorded as unintended if $\tilde{\Delta V}_t < 0$. With $\sigma_\omega = 0$ the expectation is the integrand. With the exposure shares of Section 4, the woman is at risk with probability $\pi_{\theta t}$ and the value is $(1 - \pi_{\theta t}) V_t^{3,j=1} + \pi_{\theta t}\, V_t^2$.

Never having a child. In the last fertile period $t = T_F = 7$, a reduced-form utility shifter $\phi_{nk}(e)$ for remaining childless, by education, matches the observed mass of women who never have children; it enters the no-birth continuation value inside the contraception sub-stage:

$$V_7^{3,j=1}(\Omega_{i7}) + \mathbf{1}_{\{k_7=1\}}\, \phi_{nk}(e_7).$$

OA.7.6 College age (ages 18–21; $t = 2$)

Women who do not enroll in college at $t = 2$ follow the young-adult problem of Section OA.7.5. This subsection states the college track: enrolled women do not work, receive the student allowance $w_C\, s_\theta$ and pay direct schooling costs TC throughout $t = 2$, and at the

[76]The husband's earnings are a reduced-form function of the woman's observed characteristics, as in Adda et al. (2017); Sheran (2007); van der Klaauw (1996).

end of the period choose between graduating and dropping out, in which case they enter the labor market at $t = 3$ as high-school graduates; a birth while enrolled raises the psychic cost of graduating by $|\phi_{k,D}|$.

Sub-stage 2: consumption–saving and (if a birth occurs) child investment. Let $j \in \{1, 2\}$ denote the fertility outcome in $t = 2$ ($j = 2$ if a first birth occurs, $j = 1$ otherwise, as in Section 3.2). Enrolled women do not work; the household solves

$$
\begin{aligned}
V_2^{2,j}(\Omega_{i2}) &= \max_{a_3\geq 0,\, c_2 \geq 0,\, i_2 \geq 0} \Big\{ u(c_2) + \mathbf{1}_{\{j=2\}}\, u_k(i_2) + \beta\, W_2^j(a_3) \Big\} \\
\text{s.t.} \qquad & \phi_c(m_2, \mathbf{1}_{\{j=2\}})\, c_2 + a_3 = (1+r)a_2 + 4(w_C\, s_\theta - \mathrm{TC}) - \mathbf{1}_{\{j=2\}}\, i_2,
\end{aligned}
$$

where disposable resources are the ability-scaled student allowance net of direct schooling costs over the four-year period, $4(w_C\, s_\theta - \mathrm{TC})$, with w_C and TC annual amounts like every other resource term,[77] and W_2^j is the value of the graduation decision that closes the period.

Sub-stage 3: graduate vs. drop out. At the end of $t = 2$, college women choose $d \in \{G, CD\}$ with Type-I extreme value shocks. A graduation cost ϕ_{grad}, incurred only upon completion, matches observed college dropout. Here and throughout this appendix, ϕ_{grad}, $\phi_{k,D}$, $\phi_{k,HS}$ and $\phi_{k,BAC}$ enter utility additively and are restricted to be non-positive, so each cost is the magnitude of its coefficient. Then

$$
\begin{aligned}
W_2^j(a_3) = \mathbb{E} \max \Big\{ & V_3^1(\Omega_{i3}; e_3 = \mathrm{COL}) + \phi_{grad} + \mathbf{1}_{\{j=2\}}\, \phi_{k,D} + \sigma_{CD}\varepsilon_2(G), \\
& V_3^1(\Omega_{i3}; e_3 = \mathrm{HSG}) + \sigma_{CD}\varepsilon_2(CD) \Big\},
\end{aligned}
$$

where Ω_{i3} carries a_3, experience $x_3 = x_2$, marital status $m_3 = m_2$ and, after a birth, $k_3 = 3$: the child born while she is enrolled has left the household by $t = 3$, as after every birth. A dropout enters the labor market at $t = 3$ as a high-school graduate.

Sub-stage 1: contraception. At the start of $t = 2$, childless women in the college track choose a method as in Section OA.7.5, with $V_2^{2,j=2}$ and $V_2^{2,j=1}$ as the continuation values: writing $\Delta V_2 = V_2^{2,j=2}(\Omega_{i2}) - V_2^{2,j=1}(\Omega_{i2})$, which enters as the realized $\Delta\tilde{V}_2$ of Section OA.7.5 because the transitory wantedness shock is drawn before the method choice in every fertile

[77] In school periods the student allowance net of schooling costs is a non-taxable, transfer-like resource and enters the budget constraint directly; the tax/transfer mapping applies to labor-market income when working.

period,

$$V_2^1(\Omega_{i2}) = (1-\pi_{\theta 2})\, V_2^{2,j=1}(\Omega_{i2})$$
$$+\, \pi_{\theta 2}\Big[V_2^{2,j=1}(\Omega_{i2}) + \mathbb{E}_{\varepsilon^\omega}\Big(\sigma_n \log \sum_{n=0}^{2} \exp\Big\{\frac{p_n(\theta_i, g(2), e_2)\, \Delta\tilde{V}_2 - \kappa_n}{\sigma_n}\Big\}\Big)\Big].$$

OA.7.7 Teenage (ages 14–17; $t = 1$)

At $t = 1$ all women are in high school. The within-period timing is: (i) contraception, (ii) after observing the fertility outcome, continue high school vs. drop out, and (iii) consumption–saving and, if a birth occurs, child investment. Teens who remain in school receive an allowance w_{HS}; dropouts enter the labor market and choose whether to work.

Sub-stage 3: consumption–saving, child investment, and college entry at $t = 2$. Let $j \in \{1, 2\}$ denote the fertility outcome in $t = 1$. Conditional on the schooling decision $d \in \{\text{HSG}, \text{HSD}\}$ (complete high school vs. drop out) from sub-stage 2, teens solve

$$\begin{aligned}
v_1^{3,j}(\Omega_{i1}; d) = \max_{a_2 \geq 0,\, c_1 \geq 0,\, i_1 \geq 0} \Big\{ & u(c_1) + \mathbf{1}_{\{j=2\}} u_k(i_1) + \mathbf{1}_{\{d=\text{HSG}\}} \mathbf{1}_{\{j=2\}}\, \phi_{k,HS} \\
& + \beta\Big[\mathbf{1}_{\{d=\text{HSG}\}} V_2^{E,j}(\Omega_{i2}) + \mathbf{1}_{\{d=\text{HSD}\}} V_2^1(\Omega_{i2})\Big]\Big\} \\
\text{s.t.} \qquad \phi_c(m_1, \mathbf{1}_{\{j=2\}})\, c_1 + a_2 &= (1+r)a_1 + y_1(\Omega_{i1}; d) - \mathbf{1}_{\{j=2\}}\, i_1, \\
x_2 &= x_1 + \mathbf{1}_{\{d=\text{HSD}\}} \mathbf{1}_{\{l_1=1\}}.
\end{aligned}$$

Resources satisfy $y_1(\Omega_{i1}; d) = 4w_{HS}$ if $d = \text{HSG}$, with w_{HS} the annual high-school allowance, while if $d = \text{HSD}$ the teen chooses whether to work, $l_1 \in \{0, 1\}$, with the labor-supply taste shocks and ψ_l of the working stage, so that $v_1^{3,j}(\Omega_{i1}; \text{HSD})$ is the expected maximum over l_1 of the problem above plus $\psi_l \mathbf{1}_{\{l_1=1\}}$, and

$$y_1(\Omega_{i1}; \text{HSD}) = 4\Big[\lambda_\tau\big(\tilde{y}_1^0(\Omega_{i1})\big)^{1-\tau} + \underline{y}(m_1)\Big], \qquad \tilde{y}_1^0(\Omega_{i1}) = \mathbf{1}_{\{l_1=1\}}\, w(\Omega_{i1}).$$

At the end of $t = 1$, teens who complete high school ($d = \text{HSG}$) draw a Type-I extreme value shock and choose whether to enroll in college at $t = 2$, $d_C \in \{C, NC\}$. With $v_2^1(\cdot)$ the beginning-of-period value at $t = 2$ given the education choice, the value at the enrollment node (E) is

$$V_2^{E,j}(\Omega_{i2}) = \max_{d_C \in \{C, NC\}} \big\{v_2^1(\Omega_{i2}; d_C) + \mathbf{1}_{\{d_C=C\}}\, \kappa_C(\theta, k_1) + \sigma_C \varepsilon_2(d_C)\big\}.$$

The non-pecuniary entry term is

$$\kappa_C(\theta, k_1) = \frac{\xi_C}{\theta^{\omega_C}} + \xi_\theta\,(\theta - 1) + \mathbf{1}_{\{k_1=2\}}\,\phi_{k,BAC},$$

so a woman who had a first birth at 14–17 bears the additional attendance cost $|\phi_{k,BAC}|$; her child, born in $t = 1$, has left the household by $t = 2$ ($k_2 = 3$), as after every birth. Unlike the four ϕ wedges, κ_C is not restricted in sign: ξ_C is bounded below zero and $\phi_{k,BAC} \le 0$, but ξ_θ is estimated positive, so the non-pecuniary term falls in magnitude across quartiles and, at the estimates of Table OA.39, turns slightly positive in the top quartile—enrollment is a small psychic benefit, rather than a cost, for the highest-ability women without a child.

Sub-stage 2: continue high school vs. drop out. After observing j, teens choose $d \in \{\text{HSG}, \text{HSD}\}$ with Type-I extreme value shocks:

$$V_1^{2,j}(\Omega_{i1}) = \max_{d\in\{\text{HSG},\text{HSD}\}} \{v_1^{3,j}(\Omega_{i1}; d) + \sigma_{HS}\varepsilon_1(d)\}.$$

Sub-stage 1: contraception. At the start of $t = 1$, teens choose a method in the same way, with $\Delta V_1 = V_1^{2,j=2}(\Omega_{i1}) - V_1^{2,j=1}(\Omega_{i1})$, again entering as its realized $\tilde{\Delta V}_1$, and the teen exposure share $\pi_{\theta 1}$:

$$V_1^1(\Omega_{i1}) = (1 - \pi_{\theta 1})\,V_1^{2,j=1}(\Omega_{i1}) + \pi_{\theta 1}\Big[V_1^{2,j=1}(\Omega_{i1}) + \mathbb{E}_{\varepsilon^\omega}\Big(\sigma_n \log \sum_{n=0}^{2} \exp\Big\{\frac{p_n(\theta_i, 1, e_1)\,\tilde{\Delta V}_1 - \kappa_n}{\sigma_n}\Big\}\Big)\Big].$$

OA.8 Method Choice: Choice Probabilities and Expected Conception Risk

This appendix states the fertility-control sub-problem of Section 3.4 in closed form. The stage-2 objective for a childless woman is

$$V_t^{(2)} = V_t^{\text{no child}} + \mathbb{E}_\varepsilon \max_{n\in\{0,1,2\}} \big\{p_n\,\Delta V - \kappa_n + \sigma_n\varepsilon_n\big\}, \qquad \Delta V \equiv V_t^{\text{child}} - V_t^{\text{no child}},$$

with p_n from equation (3), $\kappa_0 = 0$, and ε_n i.i.d. Type-I extreme value. Writing $v_n = p_n\Delta V - \kappa_n$,

$$P(n) = \frac{e^{v_n/\sigma_n}}{\sum_{n'} e^{v_{n'}/\sigma_n}}, \qquad \mathbb{E}\max_n\{v_n + \sigma_n\varepsilon_n\} = \sigma_n \log \sum_{n'} e^{v_{n'}/\sigma_n} + \sigma_n\gamma_E,$$

and the solution carries the inclusive value $\sigma_n \log \sum_{n'} e^{v_{n'}/\sigma_n}$, omitting $\sigma_n \gamma_E$ under the convention of Section OA.7.3. The expected probability of a first birth is $\bar{p} = \sum_n P(n) p_n$, the share using any method is $1 - P(0)$, and the pill share among users is $P(2)/(1 - P(0))$.

Comparative statics. For $\Delta V < 0$, v_n is increasing in the efficacy of n and decreasing in κ_n; $P(0)$ is increasing in ΔV, so participation falls continuously as the value of a birth rises and reaches the no-method corner only as $\Delta V \to +\infty$. A higher $c_{\theta g}$ lowers p_1 and p_2 in proportion to $(\rho_n^0 - \rho_n)$, which lowers the failure rate within each method and, by raising $|v_n - v_0|$, raises the probability of using one. The within-method failure moments respond to c and ρ_n^0 but not to κ; the any-method use moments respond to both.

Units and normalizations. ΔV, κ_n and σ_n are in utility units. Because the efficacy ratios ρ_n are fixed externally and the non-adherent rates ρ_n^0 are common across women, $c_{\theta g}$ has a scale: $c_{\theta g} = 1$ is perfect use and $c_{\theta g} = 0$ the failure rate of a user who never adheres. The counterfactuals of Section 6 use this scale. Equalizing adherence to the top quartile changes p_n at unchanged κ_n; a uniform cut in the cost of use scales κ_1 and κ_2 by $1 - \delta$ for everyone (the pill's additive education shifter is unchanged, so every woman's cost of method n falls by the same $\delta \kappa_n$; the reported experiment sets $\delta = 1$, no cost of use, and a partial cut delivers a ten percent fall in first births by 21); and a long-acting reversible method is added to the menu with $\rho_{\text{LARC}} = 0.005/0.85$ (IUD, 0.2–0.8% per year, perfect use of the levonorgestrel device to typical use of the copper device; Trussell, 2011), a failure rate that does not depend on c, at the utility cost that reproduces Colorado's take-up (and, in a variant, at the pill's cost κ_2).

Smoothness. $P(n)$, $\bar{p}$ and the log-sum are differentiable in every parameter, and the use and pill-share moments average these probabilities across simulated women rather than the realized method draws (Appendix OA.11.6). The within-method-failure moments average the conception probability each woman faces among the woman-periods whose simulated draw is that method, so they are smooth in the parameters given the draws, up to the selection of which woman-periods enter each cell.

OA.9 Interpretation and Scope of the Fertility-Control Channel

The non-adherence anchors. The anchor ρ_n^0 of equation (3) is estimated rather than imported from Trussell (2011) because the NLSY79 failure measure of Section 2.3 is a first-birth rate per year of *reported* use, which includes misreported use, discontinuation between interviews and births after stopping: 3.3 percent a year in the lowest quartile against the 12.6 percent the observed method mix implies at published typical-use rates (Online Appendix OA.6.1). Estimating the anchors also leaves the ratio of pill to low-tier failure among imperfect users to the data. With a non-adherent user fully unprotected ($\rho_n^0 = 1$) that ratio would lie between 0.82 and 0.99 over the estimated range of adherence, against about 0.5 in the data in every quartile. At $\widehat{\rho}_1^0 = 0.950$ and $\widehat{\rho}_2^0 = 0.433$ the implied ratio runs from 0.45 at the lowest estimated adherence to 0.38 at the highest, in the range of the data. The two anchors are not ordered: $\rho_2^0 \leq \rho_1^0$ (a non-adherent pill user no worse protected than a non-adherent low-tier user) is not imposed, because the anchors absorb misreported use and discontinuation, which need not be ranked like the methods themselves. Imposing it would bind only on the restricted model, whose estimated anchors violate it, while the estimated model's satisfy it, and so could only raise D.

What $c_{\theta g}$ absorbs. Adherence is identified from within-method birth rates (Section 3.4), so it carries whatever produces an ability gradient in births among users of a given method other than the method itself. Three sources enter. The first is pregnancy resolution: the model maps method use into live births, not into conceptions, so $c_{\theta g}$ absorbs any ability-correlated difference in abortion. Because abortion is more common at the top of the ability distribution in this cohort, ignoring it inflates the estimated adherence gradient by roughly a sixth; it cannot generate the gradient, which would take a top-to-bottom abortion ratio of 2.04 against 1.18 observed (Appendix OA.6.5), so the estimates carry a small upward correction rather than being a lower bound. The second is behavior around the method. The acts adherence requires—taking a pill daily, using a barrier method at every occasion, keeping appointments, negotiating with a partner—are the ones whose lapses the adherence literature ties to the absence of a routine and to not understanding the method's instructions (Rosenberg et al., 1995), to school performance (Wu et al., 2016), and to executive function and working memory (Insel et al., 2006; Stilley et al., 2010); they include the partner's conduct (Ashraf et al., 2014), and the reported-use measure includes lapses and discontinuation between interviews. Their common level goes to ρ_n^0, which is the same for every woman; any ability gradient in them goes to $c_{\theta g}$, which is the reliability of use as practised, not of the method as prescribed. The third is the trait: these acts are the ones non-cognitive skills govern, and Heckman et al. (2006) show that cognitive and non-cognitive skills both predict teen pregnancy, and Almlund et al. (2011) that measured cognitive ability partly reflects non-cognitive traits, so $c_{\theta g}$ is the fertility-control capacity of the skill bundle the AFQT proxies. The estimates condition on measured AFQT and do not depend on which trait does the work; the interpretation of the policy lever does, which is why Section 6 studies interventions that remove the need for adherence rather than interventions that raise cognitive skill.

The exception is the youngest age group. No method report exists before 1982, when the youngest respondents were already 18, so $c_{\theta,14\text{–}17}$ is identified residually, from the teen first-birth hazards and the wantedness of teen first births given the channels pinned down at later ages; since the teen gradient is the largest, it is the parameter on which the central claim leans hardest.

Scope. The model tracks only the first birth, and a child imposes its cost for one four-year period; the moments that identify fertility control are first-birth hazards, so later births add states without identifying variation. Both restrictions bias the welfare results in a known direction. A child cost lasting several periods raises the lifetime cost of an early birth most for women with the steepest wage profiles, so the one-period cost understates the gains from improving fertility control. Higher-parity births work the same way for the lowest-ability women, whose completed fertility is highest, and the opposite way for high-ability women, who delay and then compress childbearing; on net they widen the ability gap in the cost of early motherhood. Neither restriction affects the identification of adherence, which rests on first-birth hazards at ages where second births are rare.

What an adherence-proof technology removes. The NLSY79 cohort, born 1957–1964, was fertile from the mid-1970s to the late 1990s, when effectiveness depended overwhelmingly on adherence; long-acting methods diffused in the United States after the cohort had largely completed its fertility. That environment is why the channel is visible, and it bounds what the estimates license. An adherence-proof method removes the adherence margin, so a lowest-quartile woman who can choose one today bears a smaller burden than the model attributes to her, and the level of the gradient is a statement about this cohort rather than a constant of the population. The counterfactual of Section 6 is the model's answer to what removing that margin would have done for this cohort under universal take-up, an upper bound in both senses, not a forecast for a current cohort. What transfers is qualitative: where effectiveness depends on user behavior, differences in fertility-control capacity generate differences in birth timing that the cost of contraception cannot undo, so the policy implication concerns which kind of instrument reaches which women, not a magnitude.

OA.10 Wage Process Estimation

This appendix describes the estimation of the earnings profiles that parameterize the model's earnings opportunities: conditional mean annual earnings by age, education, cognitive ability and experience for women, and by age and the wife's education for husbands. The estimated coefficients and the retirement-income process are in the Additional Results, Section S.4.

OA.10.1 Wage measures and estimation samples

Women. Let w_{it} denote real annual wage-and-salary income (2016 dollars). The sample is the woman-years that meet the employment definition of Appendix OA.1 (minimum weeks, hours and annual earnings). The dependent variable is in levels, as the model is parameterized. Income, weeks and hours refer to the same calendar year: the survey's past-calendar-year income is assigned to the year it describes, and weeks and hours come from the weekly work-history arrays.

Husbands/partners. Let w^m_{it} denote the partner's real annual wage-and-salary income. The age profile is in the woman's own age, Age_{it}, as in her wage equation: the partner's age is not used, and the model evaluates his earnings at her age. The sample is the years in which the woman is married and partner earnings exceed the earnings threshold used for women. The spouse's income, weeks and hours are reported for the calendar year before the interview, so each observation is deflated with that year's CPI and dated by the woman's age in that year. Education in the husband regressions is the woman's, $\text{Educ}^f_i \in \{\text{HSD}, \text{HSG}, \text{COL}\}$, so that the coefficients reflect the partners that women of each education level attract through assortative mating.

OA.10.2 Baseline specification: women

I estimate:

$$\begin{aligned} w_{it} &= \alpha_t + \beta_1 \text{Age}^f_{it} + \beta_2 \left(\text{Age}^f_{it}\right)^2 + \delta_0 \, \text{CumExp}_{it} \\ &\quad + \sum_{e \in \{\text{HSG,COL}\}} \left(\gamma_e + \delta_e \, \text{CumExp}_{it}\right) \mathbf{1}\{\text{Educ}^f_i = e\} \\ &\quad + \sum_{q=2}^{4} \left(\gamma_q + \delta_q \, \text{CumExp}_{it}\right) \mathbf{1}\{\text{Ability}_i = q\} \\ &\quad + \sum_{e \in \{\text{HSG,COL}\}} \sum_{q=2}^{4} \left(\gamma_{eq} + \delta_{eq} \, \text{CumExp}_{it}\right) \mathbf{1}\{\text{Educ}^f_i = e\} \, \mathbf{1}\{\text{Ability}_i = q\} + \varepsilon_{it}. \end{aligned} \tag{OA.2}$$

where α_t are calendar-year fixed effects. The interactions $\text{CumExp} \times \text{Educ}^f \times \text{Ability}$ let returns to experience vary across education and ability quartiles. In the model, experience x_t counts periods worked, and each adds four years to CumExp.

OA.10.3 Baseline specification: husbands/partners

For husbands/partners I estimate:

$$\begin{aligned}
w_{it}^{m} &= \alpha_t^m + \gamma_{\mathrm{HSG}}^m \mathbf{1}\{\mathrm{Educ}_i^f = \mathrm{HSG}\} + \gamma_{\mathrm{COL}}^m \mathbf{1}\{\mathrm{Educ}_i^f = \mathrm{COL}\} + \beta_1^m \mathrm{Age}_{it} + \beta_2^m \left(\mathrm{Age}_{it}\right)^2 \\
&+ \sum_{e \in \{\mathrm{HSG},\mathrm{COL}\}} \left(\beta_{1e}^m \mathrm{Age}_{it} + \beta_{2e}^m \left(\mathrm{Age}_{it}\right)^2\right) \mathbf{1}\{\mathrm{Educ}_i^f = e\} \\
&+ \gamma_{mb}^m \mathbf{1}\{\mathrm{MarryBeforeBirth}_i = 1\} \\
&+ \sum_{e \in \{\mathrm{HSG},\mathrm{COL}\}} \gamma_{e,mb}^m \mathbf{1}\{\mathrm{Educ}_i^f = e\} \mathbf{1}\{\mathrm{MarryBeforeBirth}_i = 1\} + u_{it}.
\end{aligned} \tag{OA.3}$$

with year fixed effects α_t^m; the indicator $\mathbf{1}\{\mathrm{MarryBeforeBirth}_i = 1\}$ captures differences in spouse earnings associated with marrying before the first birth.

OA.10.4 Model inputs and aggregation to four-year periods

The estimated coefficients generate predicted annual earnings by (age, education, ability quartile, cumulative experience). A model period is four years, so the predicted annual earnings at the period's start-of-bin age are the period's annual earnings opportunity, and cumulative experience is updated by the model's accumulation rule. In retirement, labor earnings are replaced by an education-specific deterministic retirement income constructed from the pre-retirement predicted wage arrays.

OA.10.5 Which Moments Identify Adherence

All moments enter one criterion, so no block identifies any parameter on its own. Andrews et al. (2017) make the question answerable anyway: the estimator's sensitivity to the data moments is $\Lambda = -(G'WG)^{-1}G'W$, where G is the Jacobian of the simulated moments already computed for the standard errors (Section 4.2) and W the same fixed diagonal the criterion uses. No additional model solves are needed. Reporting $\Lambda_{kj}\hat{\sigma}_j$—the effect on estimate k of a one-standard-error change in moment j—and summing $|\Lambda_{kj}\hat{\sigma}_j|$ within each block gives the share of the identifying variation each block supplies.

For the 18–25 adherence increments the within-method failure block supplies 21 percent, second to first-birth timing at 30 percent, with contraceptive use at 15 percent. For the 14–17 increments the within-method block supplies only 6 percent: the NLSY79 did not ask about methods before 1982, so that cell is identified from the teen hazards and the wantedness of teen first births, as Section 3.4 states. The contrast is the quantitative version of that limitation, and it is why the paper's claim that adherence is disciplined by measured within-method failure is made for the adult increments and not for the teen cell. Per-parameter shares for all 76 parameters are in the replication package.

OA.10.6 Sensitivity of the fertility-control estimates to the wage process

The exercise. Selection into employment, measurement error in AFQT and the earnings floor all attenuate the ability gradient that least squares recovers in (OA.2), and an understated gradient understates the opportunity cost of an early birth, leaving more of the within-education birth gradient of Section 2 for the adherence channel to explain. The adverse direction is therefore a gradient steeper than the estimate, and the sensitivity is run only in that direction. Let ζ multiply every coefficient in (OA.2) that varies with the ability quartile—the level terms γ_q, the experience interactions δ_q and the education interactions γ_{eq} and δ_{eq}—holding the Q1 profile, the education gradient and the age profile fixed, so that ζ scales the Q2–Q4 premia and $\zeta = 1$ is the estimate; a proxy reliability of 0.9 implies a measurement-error correction of $\zeta \approx 1.11$. At $\zeta \in \{1.11, 1.25, 1.5, 2\}$ I re-estimate the twelve parameters of the fertility-control technology, the three age-group intercepts α_g and the nine ability increments, holding every other parameter at its estimated value. Adherence is then the only margin that can absorb the change in wages, so the design overstates how far it moves relative to a full re-estimation. Table OA.32 reports, at each ζ, the Q4$-$Q1 earnings gap of a high-school graduate at ten years of experience, the adherence gaps, the criterion and the model's ability gradient in unintended first births by age 21.

Results. The adherence gaps barely move up to $\zeta = 1.5$, which raises the earnings gap from \$7.1k to \$10.7k: the 14–17 gap is 0.551 at the estimate and 0.546 at $\zeta = 1.5$, and the 18–25 gap 0.335 and 0.340. At $\zeta = 2$ (\$14.3k) the 14–17 gap is 1 percent below the estimate and the 18–25 gap is unchanged. The criterion rises monotonically in ζ, by 9 percent at $\zeta = 2$: a steeper earnings gradient does not substitute for the adherence gradient but worsens the fit, almost entirely through the labor-supply moments, which are where a change in wages bites. The mechanism is the identification of Section 4.2. The adherence parameters are pinned by the within-method birth rates, which condition on the method in use and depend on adherence and the externally fixed efficacies, not on wages; scaling the earnings gradient shifts the value of avoiding a birth, method choice, schooling and labor supply, but leaves those moments where they were, and at $\zeta = 1.5$ the search ran 114 generations of a 24-member population over the twelve adherence slots and lowered the criterion by 5.06. The last column shows the same separation on the margin the nested test turns on (Section 5.3): the unintended-birth gradient is 21.6 points at every ζ, against 22 in the data. It cannot move, for a reason at each end. The lowest quartile's earnings

Table OA.32. Wage-process sensitivity: adherence gradient as the ability gradient in earnings is scaled

	Earnings gap Q4−Q1		Adherence gap			
ζ	\$000	rel. to $\zeta = 1$	14–17	18–25	Criterion	Unintended-birth gradient (pp)
1.00	7.1	1.00	0.551	0.335	458.5	21.6
1.11	7.9	1.11	0.549	0.334	464.0	21.6
1.25	8.9	1.25	0.549	0.334	466.6	21.6
1.50	10.7	1.50	0.546	0.340	479.5	21.6
2.00	14.3	2.00	0.547	0.335	498.3	21.6

Notes: ζ scales every ability-varying coefficient of the female earnings function (OA.2), holding the Q1 profile, the education gradient and the age profile fixed; $\zeta = 1$ is the estimate. "Earnings gap" is the predicted Q4−Q1 annual earnings of a high-school graduate at ten years of experience and age 30, in thousands of 2016 dollars. "Unintended-birth gradient" is the model's share of women whose first birth came by age 21 and was unintended, lowest quartile minus highest, in percentage points (data: 22); it is invariant to ζ because the lowest quartile's earnings are held fixed and the highest quartile's teen adherence is at its ceiling (see text). At each ζ the twelve fertility-control parameters are re-estimated with every other parameter held at its estimated value.

profile is held fixed by construction, so nothing changes for the women whose births make up the gradient's first term; and the highest quartile's teen adherence is at its ceiling, so its unintended births before 21 are the failures of near-perfect use, which a higher wage does not alter. The second and third quartiles, whose wages do rise, have fewer unintended early births at $\zeta = 2$, but by less than a tenth of a point. What ζ moves is labor supply, and the criterion records it.

OA.11 Estimation Details

This appendix reports the externally set parameters, the weighting matrix, the standard-error calculation, the distribution theory for the nested test of Section 4.2, and the solution, simulation and optimization algorithms. Table OA.33 maps each object of the fertility-control technology to the moments that carry its identifying variation (Section 4.2). The model is overidentified: 138 targeted moments against 76 structural parameters.

OA.11.1 Externally set parameters

Table OA.34 lists the parameters fixed outside SMM. The discount factor and the coefficient of relative risk aversion are Adda et al.'s (2017) estimates; tuition, taxes and transfers are set from published schedules and standard calibrations; and the bounds on conception probabilities rule out both perfect control and deterministic fecundity.

Table OA.33. Fertility-Control Parameters and the Moments That Identify Them

Object	Moments	Variation used
λ_{be} (baseline odds)	first-birth shares by age $\times$ quartile (28), jointly with the schooling block	level and age profile of first births, given the education composition of each quartile
ρ_n (method efficacy)	external: Trussell (2011)	perfect-use failure rates
$c_{\theta g}$ (adherence)	birth rate within pill / within low tier, by quartile (8)	ability gradient in failure holding the method fixed
ρ_1^0, ρ_2^0 (non-adherent failure)	the same 8 moments	level and pill/low-tier ratio of failure, common across quartiles
$\kappa_1, \kappa_2, \sigma_n$ (method costs)	any-method use by education $\times$ age $\times$ ability half (25)	participation given ΔV; the pill share among users (8) is untargeted
$\omega_{0,\theta}$ (child preference)	unintended first birth by 21, share of women, by quartile (4); first-birth timing	first births by 21 to women with $\Delta V < 0$, given adherence and method costs
σ_ω (transitory wantedness)	unintended share of first births by age at first birth, Q1 and Q4 (8)	dispersion of wantedness within a quartile at a given age at birth
ΔV (opportunity cost)	wages (external); employment, marriage, schooling blocks	ability gradient in wages and experience returns

Over-identifying: the ability *gradient* in the 28 first-birth shares—their variation across quartiles at each age, as opposed to their level and age profile—must be reproduced by the three channels jointly.

Notes: Each row lists an object of the fertility-control technology, the moment block that carries most of its identifying variation, and the variation that moves it; counts are numbers of moments (full list in Table OA.35), and "External" marks an object fixed outside the estimation. All moments enter one criterion, so the mapping says which variation is informative in the sense of Andrews et al. (2017), not that each block identifies its object alone.

Table OA.34. Externally Set Parameters

Parameter	Value	Source / Interpretation
Discount factor β_a	0.959 (annual)	Estimate, Adda et al. (2017, Table 4)
Risk aversion γ	1.98	CRRA; estimate, Adda et al. (2017, Table 4)
Risk-free rate r_a	0.04 (annual)	Real return; assumed, following Adda et al. (2017)
Monetary cost of college TC	\$9,964	Tuition, fees and books, 2016 \$
Tax progressivity τ	0.18	Heathcote et al. (2017)
Tax scale λ_τ	0.85	Normalization; income in thousands of 2016 \$
Transfer floor $\underline{y}_S$ (single)	\$8,634	SSI federal benefit rate, 2011
Transfer floor $\underline{y}_C$ (couple)	\$12,943	
Bounds on baseline risk $(p_{\min}, p_{\max})$	(0.05, 0.80)	Normalization; neither binds
Method efficacy ρ_1, ρ_2	0.03/0.85, 0.003/0.85	Perfect-use failure relative to no method
LARC efficacy ρ_{LARC}	0.005/0.85	Counterfactual only; IUD 0.2–0.8%/yr
Equivalence scale ϕ_{ca}, ϕ_{ck}	0.8, 0.5	Second adult, child

Notes: Monetary values in 2016 dollars per year, deflated by the CPI-U. Annual flows converted to four-year model periods as described in text. τ is the estimate of Heathcote et al. (2017), adopted by Daruich and Fernández (2024); the transfer floors are the SSI federal benefit rate (Social Security Administration, 2026); the method efficacies and the LARC rate follow Trussell (2011). The bounds on baseline risk are a normalization: the estimated risk is interior in every cell, and the criterion and every estimated baseline risk are unchanged for any $p_{\max}$ between 0.8 and 1, the odds λ_{be} rescaling to absorb it (Section 3.4). ρ_1 and ρ_2 are perfect-use rates; ρ_{LARC} is the midpoint of a range running from perfect use of the levonorgestrel IUD to typical use of the copper IUD, and so is conservative. TC is the annual monetary cost of attending college: in-state tuition and required fees at public four-year institutions, \$8,804 in 2016–17 (National Center for Education Statistics, 2018), plus books and supplies, \$1,160 (College Board, 2016). ρ_2 is the pill's perfect-use failure rate in Trussell (2011); ρ_1 is a use-weighted average of the perfect-use rates of the barrier and behavioral methods in the low tier.

OA.11.2 The weighting matrix

W is diagonal and fixed, with $W_{jj} = 1/\hat{\sigma}_j^2$ and $\hat{\sigma}_j$ the bootstrap standard error of data moment j from 1,000 replications resampling women, with each replication resampling women jointly across the core samples, so that the covariances among those blocks are not forced to zero; the within-method-failure and age-at-first-birth blocks come from their own samples and enter as independent of the rest. The two child-investment ratios come from outside the NLSY79 and have no bootstrap standard error; they are weighted by $1/(|m_j| + 0.01)^2$. I do not use the efficient $W = \mathcal{V}^{-1}$: Altonji and Segal (1996) show that it carries severe small-sample bias, because $\widehat{\mathcal{V}}$ is correlated with the sample moments, and a diagonal W is standard in structural life-cycle work (Adda et al., 2017; Keane and Wolpin, 1997); and $\widehat{\mathcal{V}}$ is numerically singular here, with rank 136 of 138.

OA.11.3 Standard errors

With W fixed, $\widehat{\Theta}$ is a minimum-distance estimator with asymptotic variance

$$\widehat{\mathrm{Var}}(\widehat{\Theta}) = \left(1 + \tfrac{1}{S}\right)(G'WG)^{-1}\, G'W\,\mathcal{V}\,WG\,(G'WG)^{-1}, \qquad G = \frac{\partial m^{sim}(\Theta)}{\partial \Theta'}\bigg|_{\widehat{\Theta}},$$

where $\mathcal{V}$ is the bootstrap variance of the data moments and S is the ratio of the simulated to the estimation sample, the simulation-noise inflation of McFadden (1989). The inflation factor refers to the criterion the estimator minimized, evaluated on 10,000 simulated women, so $S = 1.72$ and $1+1/S = 1.58$; the Jacobian is differenced on 160,000 women, because finite-difference noise in each column, of the order of the simulation standard deviation over the step h, enters $G'WG$ as a positive-definite term that overstates the curvature of the criterion and understates the standard errors. G is computed by two-sided finite differences with a step of one percent of each parameter's box width and again at half that step; a parameter whose standard error moves by more than a quarter between the two steps is reported as unstable. Three kinds of parameter carry no standard error: a parameter at a corner of its box, where the estimator is not asymptotically normal (the top-quartile increment at 26–33 in the estimated model; the 14–17 and 26–33 adherence intercepts in the restricted one); a direction along which the weighted Jacobian is numerically flat (in the estimated model the 26–33 intercept and its second- and third-quartile increments, where adherence is essentially complete in every quartile, the saturated top-quartile increment at 14–17 and the dropout increment to the baseline-risk odds at 14–25; ψ_{lk} for dropouts in the restricted model); and a parameter fixed at a nesting point by construction. The condition number of $G'WG$ over the remaining directions is reported as a check that the flat directions do not contaminate their neighbours.

OA.11.4 Distribution theory for the nested test

The restricted model imposes $c_{\theta g} = c_g$ by fixing the nine adherence increments at zero and is otherwise identical to the unrestricted one: same technology, moments, weights and bounds. The difference in minimized criteria, $D = L_R - L_U$, is therefore a test statistic. Two departures from χ^2_9 run in opposite directions. Because $W \neq \mathcal{V}^{-1}$, D is asymptotically a weighted sum of chi-squares; I therefore also report the mean-matched statistic $D \cdot 9/[(1 + 1/S)\,\mathrm{tr}(P_r\Sigma)]$, whose first moment equals that of a χ^2_9 under the null, with P_r the projection the nine restrictions add on the weighted moment space, $\Sigma = W^{1/2}\mathcal{V}W^{1/2}$ computed from the Jacobian and $\mathcal{V}$ at $\widehat{\Theta}$, and the factor $1+1/S$ the simulation noise in the model moments. Because the null fixes the nine increments at the boundary of their parameter space, the true null distribution is stochastically smaller than χ^2_9 and the asymptotic p-value is conservative. A re-estimation bootstrap of the critical value would cost roughly two hundred short global searches; given the size of the statistic in Section 5 relative to the one-percent critical value of 21.7, it cannot change the verdict, and I do not run it.

Overidentification. The criterion also delivers an overidentification statistic, and it is worth saying what it does and does not show. At the estimated vector it is 132.1 on 62 degrees of freedom for the estimated model and 183.5 on 71 for the restricted one. Read against a χ^2 both reject, and I do not claim otherwise: a life-cycle model matching this many moments is not a literal description of the data-generating process, and the paper does not treat it as one. The statistic reported is the minimized criterion rescaled to have the χ^2 mean under the null, $\mathrm{df} \cdot g'Wg/\mathrm{tr}[(1+1/S)W\mathcal{V}]$, so it already allows for the diagonal weighting matrix and for the simulation noise of the 10,000 simulated histories; it matches the first moment only, since with $W \neq \mathcal{V}^{-1}$ the criterion is asymptotically a weighted sum of chi-squares. The moments are also strongly dependent within blocks—the cumulative first-birth shares by quartile and age are mechanically related—so the effective number of independent restrictions is well below the nominal count.

What the statistic does show is comparative, and that is what the paper's claim rests on. Per degree of freedom it is 2.13 for the estimated model against 2.58 for the restricted one. Imposing common adherence does not merely raise the criterion by D; it leaves a model that fits the targeted moments materially worse on its own terms, before any comparison between the two. Table OA.35 reports every targeted moment with its data value, its model value and its share of the criterion, which is the fit evidence the paper asks the reader to weigh.

Are L_U and L_R global minima? D is informative only if L_R is the restricted model's minimum: a restricted search that stopped short would overstate the statistic. I therefore re-estimated the restricted model from – further starting points, each drawn uniformly within fifteen percent of the width of every parameter's range around the estimate, under its own seed, and run for 24 hours with the same differential-evolution search on a 16-core node (a population of 30). The estimate itself is in none of the initial populations, so a run that returns to it is independent evidence rather than a restart. They ended between – and – (median –), against the reported 556.0; – ended below the estimate and – within one point of it. These starts test whether the estimate is the minimum of its own basin, not that no better basin exists elsewhere in the box; closing that gap would require the restricted model to fall by more than the D of Section 5, less the one-percent critical value of 21.7, from an estimate whose population of 94 trial vectors had converged at the end of the production search. The unrestricted model is not multi-started, and the test does not require it: a lower L_U raises D, so failing to find the unrestricted minimum can only understate the statistic. The counterfactuals of Section 6 are computed at $\widehat{\Theta}_U$ and do depend on it; they should be read as the implications of the estimated vector rather than of a verified global optimum.

OA.11.5 A heterogeneous preference channel

In the restricted model ability can move preferences only through the mean utility of a child, $\omega_{0,\theta}$. A referee could argue that the test is then won by functional form: the estimated model gets nine adherence increments, while the restricted one cannot let the *dispersion* of the value of a birth vary with ability, and a smaller dispersion at high ability would also produce fewer births that are unwanted at the time. I therefore re-estimate both models with the scale of the transitory wantedness component indexed by quartile, $\sigma_\omega(\theta) = \sigma_\omega \exp(\delta_\theta)$ with $\delta_{Q1} = 0$ and $\delta_{Q2}, \delta_{Q3}, \delta_{Q4} \in [-2, 2]$, three parameters added to each model, which nests the specification of the paper at $\delta = 0$. The restricted model now has a preference channel that can act on exactly the margin the wantedness moments measure.

With this channel in both models the criterion is 445.7 for the estimated model and 546.0 for the restricted one, so $D = 100.3$ on the same nine restrictions ($p = 0.000$; scaled, 92.1), against $D = 97.5$ in the paper. In the estimated model the three dispersion parameters lower the criterion by 2.8 percent ($p = 0.005$), with $\hat{\delta}_{Q4} = -0.07$ (s.e. 0.23). The channel improves the fit, then, but it closes none of the gap: D is larger with it than without, and the estimated dispersion is not systematically smaller at high ability—$\hat{\delta}_{Q2} = -0.24$, $\hat{\delta}_{Q3} = 0.15$, none of the three different from zero at conventional levels—so a tighter distribution of the value of a birth at high ability is not what the wantedness moments call for.

OA.11.6 Solution, simulation, and optimization

The model is a finite-horizon dynamic program with one continuous state, assets, and discrete states for ability, education, experience, marital status, whether the first birth preceded marriage, and child status. Time is discrete in four-year periods, $\text{age}_t = 10 + 4t$, with $T = 17$. Assets are discretized on a 30-point grid with cubic spacing, which concentrates points near the borrowing constraint; the child-investment and contraception grids also have 30 points, and policies are evaluated off-grid by linear interpolation. The solution is by backward induction: the four retirement periods, a pure consumption–saving problem, are solved by the endogenous grid method; the six non-fertile working periods, in which the discrete labor choice makes the value function non-concave, by DC-EGM with an upper-envelope step; and the seven fertile periods, in which marriage, contraception, schooling and child investment enter as further discrete margins within a three-sub-stage period, by value-function iteration on the grid, with the child-investment first-order condition solved by bisection and the method choice aggregated in closed form as the logit log-sum of Online Appendix OA.8.

Simulation draws 10,000 life histories from pre-drawn uniform random variables with a fixed seed, so that the moments are a deterministic function of the parameter vector. Schooling decisions are evaluated after the fertility draw, conditional on its outcome, as in Online Appendix OA.7. The use and pill-share moments are averages of predicted probabilities across simulated women, and the within-method-failure moments average the conception probability each user faces rather than realized births, which makes them close to smooth in the parameters; the birth, marriage and employment moments are averages of realized outcomes under common random numbers. The contraception-use moments cover ages 22–

38 and exclude the age-18 cell, where the share of women exposed to the risk of conception differs sharply across education groups; the within-method-failure moments cover ages 18–33 (model periods 2–5), as in the main text.

The criterion of Online Appendix OA.11 is minimized by self-adaptive differential evolution (jDE; Brest et al., 2006), which evaluates each generation's trial vectors in parallel across the 47 workers of a 48-core node, with a population of 94 started around the warm start (three quarters as local perturbations, one quarter uniform on the bounds). One solve–simulate–score takes about four seconds on a single-threaded worker, so the search sustains roughly 12 evaluations per second. Each production estimate ran for 48 hours from a warm start on the unified-sample moments of Appendix OA.1.3 (36 hours for the baseline-risk model and 21 for the encompassing model and the pair of Online Appendix OA.11.5, all of which start from the production estimates; over the last five hours of each of the latter three the criterion fell by less than 0.15), with improvements over the final hour below a tenth of a point for the estimated model and below a hundredth for the restricted one, three orders of magnitude below the difference between their criteria on which the test turns; the multi-start evidence above is the convergence check. Standard errors use numerical Jacobians on 160,000 simulated women at two finite-difference steps, because the interpolated policies and the upper-envelope comparison make optimal choices move in small jumps as parameters vary, a non-smoothness that more simulated agents do not remove. The full description of the state space, first-order conditions, algorithms and computational requirements is in the Additional Results, Section S.7; the replication package reproduces every table and figure from the estimated vector in under a minute on a laptop.

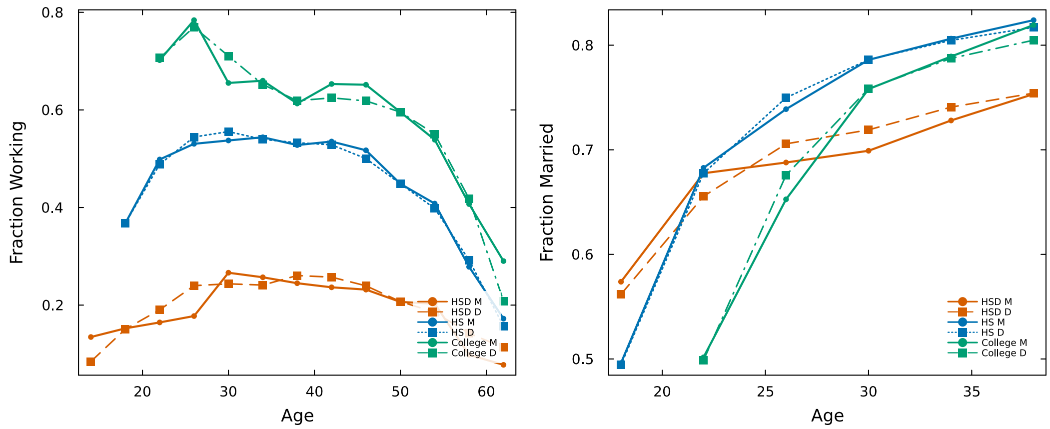
OA.12 Model Fit

OA.12.1 Targeted Moments

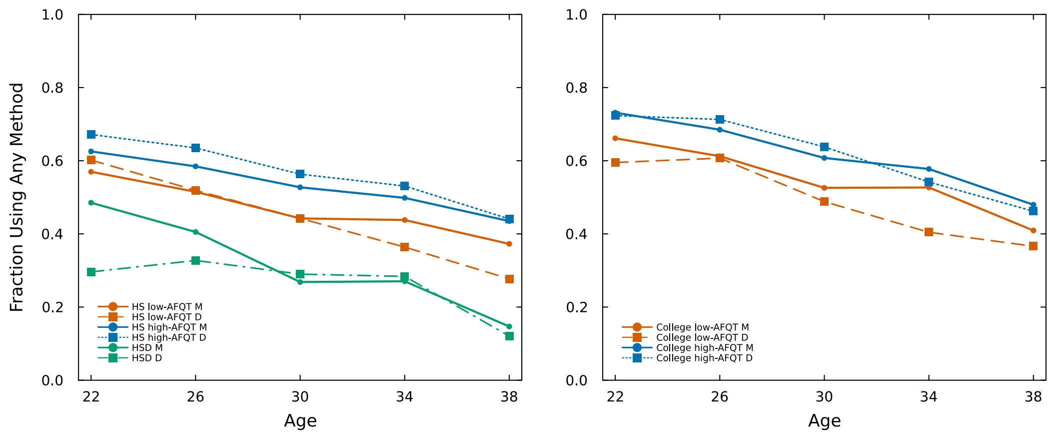
Table OA.35 reports the 138 moments that enter the criterion, each with its data value, its model value and its share of the weighted criterion, in estimation order and grouped into the nine weighted blocks of Section 4. Section 5.1 discusses the first-birth, wantedness and within-method blocks; Figure OA.6 and the paragraphs below cover the rest. The eight zero-weight method-mix moments are in Table OA.37.

Schooling and child investment. The first ten moments describe the joint distribution of schooling and early fertility. The model matches the association between a teen birth and dropping out, and the ability gradient in college attendance, from 11.5% in Q1 to 68.7% in Q4 in the data and from 10.8% to 65.8% in the model. The one departure is the third quartile,

Figure OA.6. Model Fit: Labor-Force Participation, Marriage, and Contraceptive Use



(a) Labor-force participation (left) and marriage (right) by education



(b) Contraceptive use by education and ability half

Notes: Outcomes by completed education (HSD = dropout, HS = high school, College = college graduate); panel (b) splits the HS and college groups at the median AFQT. Ages index four-year periods: 22 is the period 22–25. Solid lines with circular markers are the model, dashed lines with square markers the NLSY79 data.

which the model overpredicts; the ability-indexed college allowance of Section 3 matches the level and slope of attendance, and the miss sits in the interior of the distribution, where it does not affect the fertility gradient. The two child-investment moments are the ratios of net-of-subsidy private expenditure on older children by parental education in Caucutt and Lochner (2020, Table 6): the model reproduces the college-versus-dropout ratio and overshoots the high-school-versus-dropout ratio. Scaled by their magnitude rather than a bootstrap variance, they carry 0.3 percent of the criterion.

Labor supply. The model reproduces the education gradient in participation and its life-cycle shape, rising in the twenties, flat through the fifties and declining after 58 (Figure OA.6a, left). For high-school graduates the model is within two percentage points at every age before 62; the largest misses are at the ends of working life, dropouts at 42 and 62 and college graduates at 62, none above 8 percentage points. Experience is not targeted: the experience profiles that feed wages are disciplined by the 36 employment shares.

Marriage. Marriage rates rise with age in every education group and converge by the early thirties, as in the data (Figure OA.6a, right). No marriage moment departs from the data by more than 2.3 percentage points; the largest miss is dropouts at 22, and college marriage at 26 is slightly underpredicted, closing by 30. The model omits non-pecuniary match quality and the insurance value of a second earner, which is where the remaining discrepancies sit.

Contraception use. The model matches the education gradient in use and its decline with age from 22 on (Figure OA.6b). The largest single discrepancy in the moment vector is dropouts' use at 22, 0.49 in the model against 0.30 in the data, 5.3 percent of the criterion on its own: for a childless dropout the value of avoiding a birth is close to zero, so her use margin is governed by the method costs, which are common across education groups. Among college women the model errs in both directions. Reported use and the model's use margin differ: a woman relying intermittently on a low-tier method may not describe herself as a user, and two users of the same method can differ in how reliably they use it. Birth rates by education and ability are matched throughout.

Table OA.35. Targeted moments: data, model, and share of the criterion

j	Stata j	Moment	Data	Model	Weight share (%)
schooling					
1	1	HS dropout, no birth 14-17	0.070	0.063	0.9
2	2	HS dropout, birth 14-17	0.300	0.336	1.2
3	3	College attendance, no birth 14-17	0.414	0.417	0.0
4	4	College attendance, birth 14-17	0.082	0.091	0.2
5	5	College attendance, AFQT Q1	0.115	0.108	0.1
6	6	College attendance, AFQT Q2	0.247	0.229	0.6
7	7	College attendance, AFQT Q3	0.421	0.463	2.3
8	8	College attendance, AFQT Q4	0.687	0.658	1.4
9	9	College graduation, no birth 18-21	0.619	0.589	1.5
10	10	College graduation, birth 18-21	0.256	0.235	0.1
investment					
11	11	Child investment ratio HS/HSD	1.200	2.583	0.3
12	12	Child investment ratio Coll/HSD	4.600	3.335	0.0
birth_rate					
13	20	Birth rate, AFQT Q1, age 14	0.285	0.318	1.7
14	21	Birth rate, AFQT Q1, age 18	0.631	0.581	3.2
15	22	Birth rate, AFQT Q1, age 22	0.770	0.757	0.3
16	23	Birth rate, AFQT Q1, age 26	0.827	0.810	0.7

continued on the next page

j	Stata j	Moment	Data	Model	Weight share (%)
17	24	Birth rate, AFQT Q1, age 30	0.858	0.865	0.1
18	25	Birth rate, AFQT Q1, age 34	0.872	0.883	0.3
19	26	Birth rate, AFQT Q1, age 38	0.875	0.898	1.6
20	27	Birth rate, AFQT Q2, age 14	0.166	0.200	2.5
21	28	Birth rate, AFQT Q2, age 18	0.480	0.455	0.8
22	29	Birth rate, AFQT Q2, age 22	0.669	0.653	0.4
23	30	Birth rate, AFQT Q2, age 26	0.752	0.724	1.3
24	31	Birth rate, AFQT Q2, age 30	0.795	0.795	0.0
25	32	Birth rate, AFQT Q2, age 34	0.817	0.820	0.0
26	33	Birth rate, AFQT Q2, age 38	0.825	0.853	1.7
27	34	Birth rate, AFQT Q3, age 14	0.088	0.096	0.2
28	35	Birth rate, AFQT Q3, age 18	0.313	0.296	0.4
29	36	Birth rate, AFQT Q3, age 22	0.504	0.513	0.1
30	37	Birth rate, AFQT Q3, age 26	0.633	0.594	2.1
31	38	Birth rate, AFQT Q3, age 30	0.693	0.678	0.4
32	39	Birth rate, AFQT Q3, age 34	0.722	0.709	0.3
33	40	Birth rate, AFQT Q3, age 38	0.734	0.749	0.4
34	41	Birth rate, AFQT Q4, age 14	0.034	0.033	0.0
35	42	Birth rate, AFQT Q4, age 18	0.185	0.202	0.6
36	43	Birth rate, AFQT Q4, age 22	0.366	0.436	6.1
37	44	Birth rate, AFQT Q4, age 26	0.551	0.525	0.8
38	45	Birth rate, AFQT Q4, age 30	0.666	0.626	2.4
39	46	Birth rate, AFQT Q4, age 34	0.707	0.678	1.3
40	47	Birth rate, AFQT Q4, age 38	0.722	0.730	0.1
marriage					
41	48	Share married, educ 1, age 18	0.562	0.574	0.1
42	49	Share married, educ 1, age 22	0.656	0.678	0.3
43	50	Share married, educ 1, age 26	0.706	0.688	0.2
44	51	Share married, educ 1, age 30	0.719	0.699	0.3
45	52	Share married, educ 1, age 34	0.741	0.728	0.1
46	53	Share married, educ 1, age 38	0.754	0.753	0.0
47	54	Share married, educ 2, age 18	0.494	0.496	0.0
48	55	Share married, educ 2, age 22	0.678	0.683	0.1
49	56	Share married, educ 2, age 26	0.750	0.739	0.5
50	57	Share married, educ 2, age 30	0.786	0.786	0.0
51	58	Share married, educ 2, age 34	0.805	0.806	0.0
52	59	Share married, educ 2, age 38	0.817	0.824	0.3
53	60	Share married, educ 3, age 22	0.499	0.502	0.0
54	61	Share married, educ 3, age 26	0.676	0.653	0.6
55	62	Share married, educ 3, age 30	0.758	0.758	0.0
56	63	Share married, educ 3, age 34	0.787	0.789	0.0
57	64	Share married, educ 3, age 38	0.805	0.819	0.3
labor					
58	65	Employment, educ 1, age 14	0.084	0.135	4.7
59	66	Employment, educ 1, age 18	0.151	0.152	0.0
60	67	Employment, educ 1, age 22	0.190	0.164	0.8
61	68	Employment, educ 1, age 26	0.240	0.178	3.2
62	69	Employment, educ 1, age 30	0.244	0.266	0.4
63	70	Employment, educ 1, age 34	0.241	0.257	0.2
64	71	Employment, educ 1, age 38	0.261	0.245	0.2
65	72	Employment, educ 1, age 42	0.257	0.236	0.3
66	73	Employment, educ 1, age 46	0.240	0.232	0.0
67	74	Employment, educ 1, age 50	0.207	0.207	0.0
68	75	Employment, educ 1, age 54	0.185	0.203	0.3

continued on the next page

j	Stata j	Moment	Data	Model	Weight share (%)
69	76	Employment, educ 1, age 58	0.142	0.098	1.2
70	77	Employment, educ 1, age 62	0.113	0.078	0.2
71	78	Employment, educ 2, age 18	0.367	0.368	0.0
72	79	Employment, educ 2, age 22	0.489	0.498	0.5
73	80	Employment, educ 2, age 26	0.544	0.531	0.8
74	81	Employment, educ 2, age 30	0.556	0.537	1.2
75	82	Employment, educ 2, age 34	0.540	0.544	0.1
76	83	Employment, educ 2, age 38	0.533	0.528	0.1
77	84	Employment, educ 2, age 42	0.529	0.535	0.2
78	85	Employment, educ 2, age 46	0.500	0.517	1.1
79	86	Employment, educ 2, age 50	0.449	0.449	0.0
80	87	Employment, educ 2, age 54	0.398	0.408	0.4
81	88	Employment, educ 2, age 58	0.292	0.278	0.5
82	89	Employment, educ 2, age 62	0.157	0.172	0.3
83	90	Employment, educ 3, age 22	0.706	0.702	0.0
84	91	Employment, educ 3, age 26	0.770	0.784	0.4
85	92	Employment, educ 3, age 30	0.711	0.655	4.9
86	93	Employment, educ 3, age 34	0.652	0.660	0.1
87	94	Employment, educ 3, age 38	0.619	0.613	0.0
88	95	Employment, educ 3, age 42	0.625	0.653	1.1
89	96	Employment, educ 3, age 46	0.618	0.652	1.5
90	97	Employment, educ 3, age 50	0.596	0.596	0.0
91	98	Employment, educ 3, age 54	0.551	0.539	0.2
92	99	Employment, educ 3, age 58	0.418	0.407	0.1
93	100	Employment, educ 3, age 62	0.208	0.290	1.7
contraception					
94	101	Contraception use, educ 1, age 22	0.296	0.485	5.3
95	102	Contraception use, educ 1, age 26	0.327	0.405	0.5
96	103	Contraception use, educ 1, age 30	0.290	0.268	0.0
97	104	Contraception use, educ 1, age 34	0.284	0.270	0.0
98	105	Contraception use, educ 1, age 38	0.121	0.147	0.0
99	106	Contraception use, educ 2, AFQT low, age 22	0.602	0.570	0.8
100	107	Contraception use, educ 2, AFQT low, age 26	0.519	0.515	0.0
101	108	Contraception use, educ 2, AFQT low, age 30	0.443	0.442	0.0
102	109	Contraception use, educ 2, AFQT low, age 34	0.364	0.438	1.7
103	110	Contraception use, educ 2, AFQT low, age 38	0.277	0.373	2.2
104	111	Contraception use, educ 2, AFQT high, age 22	0.672	0.625	2.1
105	112	Contraception use, educ 2, AFQT high, age 26	0.635	0.585	1.7
106	113	Contraception use, educ 2, AFQT high, age 30	0.563	0.527	0.5
107	114	Contraception use, educ 2, AFQT high, age 34	0.531	0.498	0.3
108	115	Contraception use, educ 2, AFQT high, age 38	0.441	0.435	0.0
109	116	Contraception use, educ 3, AFQT low, age 22	0.595	0.661	0.4
110	117	Contraception use, educ 3, AFQT low, age 26	0.607	0.612	0.0
111	118	Contraception use, educ 3, AFQT low, age 30	0.488	0.526	0.1
112	119	Contraception use, educ 3, AFQT low, age 34	0.404	0.527	0.7
113	120	Contraception use, educ 3, AFQT low, age 38	0.366	0.409	0.1
114	121	Contraception use, educ 3, AFQT high, age 22	0.724	0.731	0.1
115	122	Contraception use, educ 3, AFQT high, age 26	0.713	0.685	1.0
116	123	Contraception use, educ 3, AFQT high, age 30	0.638	0.608	0.5
117	124	Contraception use, educ 3, AFQT high, age 34	0.542	0.577	0.4
118	125	Contraception use, educ 3, AFQT high, age 38	0.462	0.480	0.1
wantedness					
119	126	Unintended first birth by 21, share of women (own report), AFQT Q1	0.325	0.319	0.1

continued on the next page

j	Stata j	Moment	Data	Model	Weight share (%)
120	127	Unintended first birth by 21, share of women (own report), AFQT Q2	0.253	0.234	0.6
121	128	Unintended first birth by 21, share of women (own report), AFQT Q3	0.170	0.183	0.4
122	129	Unintended first birth by 21, share of women (own report), AFQT Q4	0.102	0.103	0.0
method_fail					
123	130	Pill annual failure, AFQT Q1, ages 18-33	0.020	0.031	0.9
124	131	Pill annual failure, AFQT Q2, ages 18-33	0.010	0.015	0.8
125	132	Pill annual failure, AFQT Q3, ages 18-33	0.010	0.013	0.5
126	133	Pill annual failure, AFQT Q4, ages 18-33	0.007	0.007	0.0
127	134	Low-tier annual failure, AFQT Q1, ages 18-33	0.050	0.073	0.5
128	135	Low-tier annual failure, AFQT Q2, ages 18-33	0.024	0.037	0.6
129	136	Low-tier annual failure, AFQT Q3, ages 18-33	0.015	0.031	2.1
130	137	Low-tier annual failure, AFQT Q4, ages 18-33	0.013	0.019	0.6
unint_fbage					
139	146	Unintended share of first births at 14-17, AFQT Q1	0.626	0.631	0.0
140	147	Unintended share of first births at 14-17, AFQT Q4	0.696	0.683	0.0
141	148	Unintended share of first births at 18-21, AFQT Q1	0.456	0.448	0.0
142	149	Unintended share of first births at 18-21, AFQT Q4	0.564	0.473	1.5
143	150	Unintended share of first births at 22-25, AFQT Q1	0.294	0.391	1.7
144	151	Unintended share of first births at 22-25, AFQT Q4	0.330	0.307	0.1
145	152	Unintended share of first births at 26-41, AFQT Q1	0.256	0.172	1.1
146	153	Unintended share of first births at 26-41, AFQT Q4	0.142	0.193	2.1

Notes: The 138 weighted moments, in estimation order. "Stata j" is the moment's index in the data pipeline's moment file; the gaps are the seven ratios dropped at load as exact functions of moments already in the vector. "Weight share" is the moment's contribution to the weighted criterion, in percent of the total. The eight zero-weight method-mix moments are in Table OA.37.

OA.12.2 The fertility-control moments in the two models

Table OA.36 sets the estimated and the restricted model beside the data on the twenty moments that identify the fertility-control technology, and reports the parameters the restriction moves most. Section 5.3 of the paper summarizes it.

OA.12.3 An untargeted check: the composition of method use

The share of the pill among women who use some method carries zero weight in the estimation, so the model's prediction for it is an untargeted check. If the ability gradient in first births were produced by which method women choose rather than by how well they use it, the pill share would rise steeply with ability. Table OA.37 shows that it does not. Among childless women aged 18–33 the pill share in the data is flat and if anything mildly declining in ability, from 0.835 in the lowest quartile to 0.824 in the highest, with standard errors of about 0.023; the model produces 0.482 to 0.547 across the same cells, flat as well and at a

Table OA.36. Where the Two Models Separate: Fertility-Control Moments and the Parameters the Restriction Moves

		Model		% of own criterion	
	Data	Estimated	Restricted	Estim.	Restr.
A. Unintended first birth by 21, share of women					
Q1	0.325	0.319	0.290	0.1	1.4
Q2	0.253	0.234	0.263	0.6	0.1
Q3	0.170	0.183	0.182	0.4	0.3
Q4	0.102	0.103	0.114	0.0	0.4
B. Annual first-birth rate within method, ages 18–33					
Pill, Q1	0.020	0.031	0.018	0.9	0.0
Pill, Q2	0.010	0.015	0.017	0.8	1.1
Pill, Q3	0.010	0.013	0.016	0.5	1.6
Pill, Q4	0.007	0.007	0.016	0.0	3.7
Low tier, Q1	0.050	0.073	0.015	0.5	1.0
Low tier, Q2	0.024	0.037	0.016	0.6	0.2
Low tier, Q3	0.015	0.031	0.017	2.1	0.0
Low tier, Q4	0.013	0.019	0.015	0.6	0.1
C. Unintended share of first births by age at first birth, percent					
14–17, Q1	62.6	63.1	61.6	0.0	0.0
14–17, Q4	69.6	68.3	80.4	0.0	0.5
18–21, Q1	45.6	44.8	36.8	0.0	2.5
18–21, Q4	56.4	47.3	55.4	1.5	0.0
22–25, Q1	29.4	39.1	32.5	1.7	0.1
22–25, Q4	33.0	30.7	33.9	0.1	0.0
26–41, Q1	25.6	17.2	26.4	1.1	0.0
26–41, Q4	14.2	19.3	19.1	2.1	1.6

	Estimated	Restricted
D. Parameters the restriction moves (standard errors in parentheses)		
Child-utility increment, Q4, $\omega_{0,Q4} - \omega_{0,Q1}$	-0.066 (0.032)	-0.216 (0.048)
Transitory wantedness component, σ_ω	0.212 (0.044)	0.287 (0.035)
Cost of the low-tier method, κ_1	0.016 (0.028)	0.231 (0.057)
Cost of the pill, κ_2	0.037 (0.041)	0.226 (0.027)
Dispersion of method taste shocks, σ_n	0.081 (0.020)	0.017 (0.012)
Adherence intercept, 14–17 (logit)	-1.10 (1.08)	9.99 (–)
Adherence intercept, 26–33 (logit)	4.86 (–)	10.00 (–)
Child penalty, high school, ψ_{lk}	-0.021 (0.009)	-0.053 (0.032)

Notes: Panels A–C: the twenty targeted moments that identify the fertility-control technology (Table OA.33), in the data, in the estimated model and in the restricted model that imposes $c_{\theta g} = c_g$; the last two columns are each moment's share of its own model's weighted criterion, in percent. Panel A is the share of women in each AFQT quartile with a first birth by 21 that they reported as mistimed or unwanted. Panel B is the annual first-birth rate among childless women using each method, ages 18–33, annualized as $1 - (1-p)^{1/4}$ on both sides. Panel C is the percent of first births in each age range that were unintended, lowest and highest quartiles; in the model a first birth is unintended if $\Delta V + \sigma_\omega \varepsilon^\omega < 0$ at the start of its period (Section 3.4). Panel D: the parameters that move under the restriction, with sandwich standard errors; in the restricted model the adherence intercept is common to all quartiles. These parameters are in the model's utility units, except the two adherence intercepts, which are in logits and map into the probabilities of Table 3; a dash in place of a standard error marks a parameter for which none can be computed, as in that table. The full vectors, with standard errors, are in Online Appendix OA.13.

Table OA.37. Untargeted: pill share among method users, by AFQT quartile

	Q1	Q2	Q3	Q4
Ages 18–25, data	0.802	0.777	0.769	0.714
model	0.498	0.499	0.560	0.569
Ages 26–33, data	0.756	0.734	0.724	0.647
model	0.431	0.446	0.486	0.506
Ages 18–33, parity 0, data	0.835	0.843	0.845	0.824
(s.e.)	(0.023)	(0.018)	(0.016)	(0.015)
model	0.482	0.484	0.535	0.547

Notes: Share of the pill among childless women using some method, by AFQT quartile. These moments carry zero weight in the criterion. The first two panels are the eight zero-weight moments of the estimation vector; the third is the method composition built from the NLSY79 contraception battery at parity zero, with bootstrap standard errors.

lower level. Composition therefore works against the within-pill gradient of Section 2.3: Online Appendix OA.6.1 prices the observed method mix at published typical-use failure rates and obtains a Q1/Q4 failure ratio of 1.06 against an observed 2.72. The shortfall in the level reflects the two-method technology: with the estimated failure rates, the unintended births among users in the data require more low-tier users than the data have, and the untargeted mix is where the estimator gives. The Additional Results (Section S.5) re-estimate the five method-cost parameters with the eight pill-share moments switched on and everything else held at the estimate, which brings the share to 0.66–0.76 but raises the criterion on the paper's moments by 18 percent, with the loss concentrated in the unintended-birth blocks: the share of the lowest quartile with an unintended first birth by 21 falls from 0.319 to 0.285 against 0.325 in the data. The experiments of Section 6, re-run at that vector, move by little: the long-acting method at Colorado's take-up lowers first births by 21 by 6.9 percent against 7.1 at the estimate, equalizing adherence by 13.4 against 16.1, and the incidence across quartiles is unchanged. The lowest quartile's welfare gain from the long-acting method is 10.2 percent against 9.9 at the estimate.

OA.12.4 Unintended births: non-use or method failure?

The data place the ability gradient in unintended first births in non-use rather than in failure (Online Appendix OA.3): 85.1 percent of the lowest quartile's unintended first births were conceived with no method in use, against 53.0 percent in the highest. The wantedness moments count unintended births whatever their route, so the model is free to produce them through either. Table OA.38 shows which it uses. In the model a first birth is unintended when the realized value of a birth is negative, and it comes through non-use when the woman chose no method that period and through failure when her method failed; lower

Table OA.38. Unintended first births: conceived with no method or through method failure

	Data	Estimated	Restricted
A. Share of unintended first births (ages 14–41) conceived with no method, %			
Q1	85.1	42.5	87.8
Q4	53.0	71.8	62.4
B. Q1 minus Q4, unintended first birth by 21, share of women, %			
Total		21.6	17.6
through non-use		5.5	20.5
through method failure		16.1	−2.9

Notes: Data: the first birth's own pregnancy record; "no method" includes having stopped a method for reasons other than wanting to conceive. Model: 10,000 simulated women at each model's estimate; a first birth is unintended when the realized value of a birth, $\Delta V + \sigma_\omega \varepsilon^\omega$, is negative at the start of its period. Panel B is the targeted moment of Section 4 split by route; the data do not record the route for the per-woman moment on the same basis, so the comparison is Panel A.

adherence acts on both, since it lowers the return to using any method. On this split the estimated model and the data part company. In the data the share of unintended first births conceived with no method falls with ability; in the estimated model it rises—42.5 percent for the lowest quartile against 71.8 for the highest—and 16.1 of the 21.6 point gap by 21 runs through failure. The restricted model, in which an ability gradient can arise only through non-use, reproduces the data's ordering. At the estimate, then, adherence acts mainly on failure conditional on use and little on the decision to use: the within-method moments that identify it pin the first, and no targeted moment records the route of an unintended birth to discipline the second. The test of Section 5.3 does not turn on the route: the restricted model produces all of this gradient through non-use, with the failure route running the other way, and is rejected on the timing and within-method moments. But the split is a limitation of the estimated technology, and a moment that records it is the natural next target.

OA.13 Estimated Structural Parameters

Table OA.39 reports the estimated vector with the sandwich standard errors of Section 4, alongside the vector estimated under the restriction $c_{\theta g} = c_g$ of Section 5; parameters are grouped by their role in the model. Six parameters of the estimated model carry no standard error (Section 4 and the footnote of Section 5.2): the 26–33 adherence intercept and its second- and third-quartile increments, where adherence is essentially complete in every quartile and the criterion is flat in them; the saturated top-quartile increment at 14–17; the dropout increment to the baseline-risk odds at 14–25, also flat; and the top-quartile increment at 26–33, which sits on the upper bound of its box. In the restricted model the 14–17 and 26–33 adherence intercepts sit on their upper bounds and ψ_{lk} for dropouts is flat. The row labels map to the symbols of Section 3 as follows: the baseline-fertility rows are the odds λ_{be} of Section 3.4 by education and age group, reported as the college level and the non-negative education increments, and below them the implied baseline risk p^0_{be} of each cell, which is the object the moments pin down (the odds at 14–25 are large only because the risk there is close to $p_{\max}$); in each adherence block "α, Q1 intercept" is α_g and "Δ Q2–Q4" are the cumulative increments Δ_{rg}; the subscripts *low* and *pill* on κ and ρ^0 are the methods $n = 1$ and $n = 2$, $\Delta\kappa_{pill}$ HS and Coll are the education shifters $\Delta\kappa_{2,e}$, and σ_ω is the scale of the transitory component of the value of a birth; in the schooling panel $\phi_{k,BAC}$ is the college-attendance term and $\phi_{k,HS}$, $\phi_{k,D}$ and ϕ_{grad} are the continuation wedges of Section 3.3. The adherence parameters are in Table 3 of the main text; Figure OA.7 plots them against the restricted model. The allowances are annual, in thousands of 2016 dollars, like every money amount in the paper; the method costs κ are in units of ΔV rather than money (Section 3.4).

Table OA.39. Estimated structural parameters

	Full model	Restricted: $c_{\theta g} = c_g$
Parameter	Estimate (S.E.)	Estimate (S.E.)
A. Fertility-control technology		
Baseline odds of a first birth with no method, λ_{be}		
College, ages 14–25	10.51 (8.72)§	10.38 (5.99)§
College, ages 26–33	1.13 (0.31)§	1.77 (0.46)
College, ages 34–41	0.50 (0.12)	0.45 (0.13)
Δ HS, ages 14–25	5.45 (8.80)§	2.52 (4.68)
Δ HS, ages 26–33	0.35 (0.45)§	0.54 (0.71)§
Δ HS, ages 34–41	0.00 (0.16)	0.26 (0.21)
Δ HSD, ages 14–25	0.12‡ –	2.30 (7.33)§
Implied baseline risk p^0_{be} (probability of a first birth in a period with no method)		
College, ages 14–25	0.730 (0.053)	0.730 (0.037)
College, ages 26–33	0.424 (0.054)	0.511 (0.048)
College, ages 34–41	0.267 (0.043)	0.248 (0.048)
High school, ages 14–25	0.753 (0.020)	0.742 (0.020)
High school, ages 26–33	0.477 (0.030)	0.558 (0.032)
High school, ages 34–41	0.268 (0.023)	0.333 (0.032)
Dropout, ages 14–25	0.753 –	0.751 (0.019)
Adherence $c_{\theta g}$, ages 14–17 (logit scale)		
α, Q1 intercept	−1.097 (1.081)	9.993† –
Δ Q2	1.450 (0.662)	0^R –
Δ Q3	1.038 (0.531)	0^R –
Δ Q4	4.897‡ –	0^R –
Adherence $c_{\theta g}$, ages 18–25 (logit scale)		
α, Q1 intercept	0.177 (0.497)	0.807 (0.737)
Δ Q2	0.983 (0.262)	0^R –
Δ Q3	0.118 (0.251)	0^R –

Table OA.39, continued

	Full model		Restricted: $c_{\theta g} = c_g$	
Parameter	Estimate	(S.E.)	Estimate	(S.E.)
Δ Q4	0.707	(0.261)	0^R	–
Adherence $c_{\theta g}$, ages 26–33 (logit scale)				
α, Q1 intercept	$4.863^‡$	–	$10.000^†$	–
Δ Q2	$0.262^‡$	–	0^R	–
Δ Q3	$2.045^‡$	–	0^R	–
Δ Q4	$5.000^†$	–	0^R	–
Failure rate at zero adherence				
ρ^0_{low}	0.950	(0.241)	0.204	(0.124)
ρ^0_{pill}	0.433	(0.129)	0.411	(0.201)
Transitory wantedness shock (units of ΔV)				
σ_ω	0.212	$(0.044)^§$	0.287	(0.035)
Method costs (units of ΔV)				
κ_{low}	0.016	(0.028)	0.231	$(0.057)^§$
κ_{pill}	0.037	(0.041)	0.226	$(0.027)^§$
$\Delta\kappa_{pill}$ HS	0.016	(0.037)	−0.168	$(0.024)^§$
$\Delta\kappa_{pill}$ Coll	−0.066	(0.038)	−0.263	$(0.032)^§$
σ_n (method taste shocks)	0.081	(0.020)	0.017	(0.012)
B. Preferences over children				
Utility of a child, $\omega_{0,\theta} + \omega_1 i^{\omega_2}$, with $\omega_{0,\theta} = \omega_0 + \Delta\omega_{0,\theta}$ and $\Delta\omega_{0,Q1} = 0$				
ω_0	−0.078	$(0.125)^§$	−0.319	(0.059)
$\Delta\omega_{0,Q2}$	0.004	(0.041)	−0.154	(0.045)
$\Delta\omega_{0,Q3}$	−0.135	(0.043)	−0.237	(0.048)
$\Delta\omega_{0,Q4}$	−0.066	(0.032)	−0.216	(0.048)
ω_1	0.293	$(0.100)^§$	0.186	(0.059)
ω_2	0.041	(0.029)	0.052	(0.039)
Cost of a child, by education				
$\phi_{k,HS}$	−0.053	(0.014)	−0.072	(0.023)
$\phi_{k,D}$	−0.272	(0.120)	−0.320	(0.175)
$\phi_{k,BAC}$	−0.429	(0.124)	−0.499	(0.127)
Terminal utility with no child				
ϕ_{nk} HSD	−0.496	(0.216)	0.251	(0.159)
ϕ_{nk} HS	−0.022	(0.072)	−0.183	(0.060)
ϕ_{nk} Coll	−0.111	(0.115)	−0.218	(0.098)
C. Schooling				
College psychic cost, $\xi_C/\theta^{\omega_C} + \xi_\theta(\theta - 1)$				
ξ_C (level)	−0.781	(0.188)	−1.339	(0.220)
ω_C (ability curvature)	0.136	(0.173)	0.124	(0.151)
ξ_θ (ability slope)	0.251	(0.079)	0.393	(0.110)
ϕ_{grad} (college graduation cost)	−0.001	(0.029)	−0.109	(0.097)
Allowances while enrolled, annual				
HS allowance	18.82	$(3.92)^§$	27.26	$(16.71)^§$
College allowance	26.05	$(9.98)^§$	39.62	$(6.69)^§$

Table OA.39, continued

Parameter	Full model Estimate (S.E.)	Restricted: $c_{\theta g} = c_g$ Estimate (S.E.)
Scale of the schooling taste shocks		
σ_C (college entry)	0.400 (0.095)	0.613 (0.109)
σ_{CD} (college continuation)	0.211 (0.089)	0.427 (0.094)
σ_{HS} (high-school continuation)	0.086 (0.016)	0.145 (0.020)§
D. Marriage: meeting probability μ		
HSD, ages 18–21	0.556 (0.030)	0.556 (0.028)
HSD, ages 22–25	0.263 (0.032)	0.238 (0.033)
HSD, ages 26–33	0.043 (0.020)	0.066 (0.023)
HSD, ages 34–41	0.084 (0.023)§	0.059 (0.019)
HS, ages 18–21	0.658 (0.015)	0.647 (0.014)
HS, ages 22–25	0.380 (0.014)	0.390 (0.013)
HS, ages 26–33	0.182 (0.011)	0.178 (0.011)
HS, ages 34–41	0.092 (0.010)	0.075 (0.009)
Coll, ages 22–25	0.516 (0.021)	0.516 (0.020)
Coll, ages 26–33	0.317 (0.021)	0.324 (0.022)
Coll, ages 34–41	0.134 (0.023)	0.099 (0.024)
E. Labor supply		
Disutility of work ψ_l*, by education and age*		
ψ_l HSD, ages 14–29	−0.0237 (0.0023)	−0.0257 (0.0028)
ψ_l HSD, ages 30–49	−0.0179 (0.0018)	−0.0196 (0.0024)
ψ_l HSD, ages 54–65	−0.0157 (0.0019)	−0.0181 (0.0024)
ψ_l HS, ages 14–29	−0.0051 (0.0005)	−0.0021 (0.0008)
ψ_l HS, ages 30–49	−0.0038 (0.0003)	−0.0035 (0.0004)
ψ_l HS, ages 54–65	−0.0072 (0.0006)	−0.0081 (0.0009)
ψ_l Coll, ages 14–29	0.0079 (0.0020)	0.0018 (0.0017)
ψ_l Coll, ages 30–49	0.0008 (0.0009)	−0.0002 (0.0010)
ψ_l Coll, ages 54–65	−0.0022 (0.0006)	−0.0020 (0.0008)
ψ_l HSD, ages 50–53	−0.0168 (0.0018)	−0.0193 (0.0031)§
ψ_l HS, ages 50–53	−0.0059 (0.0005)	−0.0064 (0.0005)
ψ_l Coll, ages 50–53	−0.0010 (0.0007)	−0.0005 (0.0010)
Interaction with a child in the household		
ψ_{lk} HSD	−0.061 (0.135)§	−0.714‡ –
ψ_{lk} HS	−0.021 (0.009)	−0.053 (0.032)§
ψ_{lk} Coll	−0.051 (0.019)	0.264 (0.100)
Other		
σ_l (labor taste shock)	0.0081 (0.0012)	0.0097 (0.0020)
ψ_{ret} (ages 58+)	−0.0045 (0.0007)	−0.0051 (0.0011)
Estimated parameters	76	67
Criterion $L(\hat{\Theta})$	458.52	556.03
$D = L_R - L_U$	97.5	on 9 restrictions

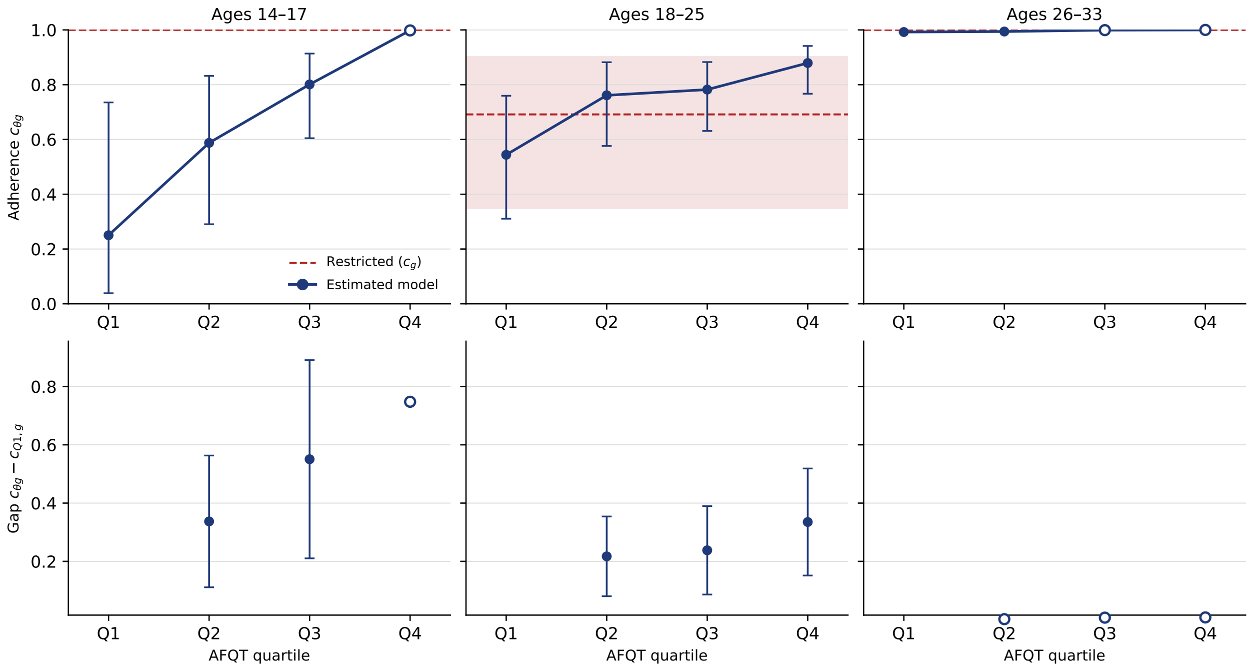


Figure OA.7. Adherence by AFQT quartile, within age group: levels and contrasts

Notes: Top row: estimated adherence $c_{\theta g}$ (solid, navy) with 95 percent intervals computed on the logit scale, on which it is estimated; the dashed line is the restricted model, in which adherence is common across quartiles, with its 95 percent band shaded. Bottom row: the contrasts $c_{\theta g} - c_{Q1,g}$ with the standard error of the difference, which nets out the base logit slot that the four levels of an age group share (Panel C of Table 3). The contrasts are what the nested test of Section 5.3 is about, and they are estimated more precisely than the levels above them: a level moves with the base slot, a gradient does not. A hollow marker carries no interval, either because the logistic is saturated near one so that no standard error is defined, or because the logit standard error exceeds three, at which point the interval covers essentially the whole unit interval and reports nothing. The same rule governs the restricted band, which is drawn only at 18–25; at 14–17 the restricted value is saturated (the logit intercept sits on the upper bound of its box, so it has no standard error).

Notes: Simulated method of moments on the NLSY79. Standard errors in parentheses, from the sandwich of Section 4 with Ω the cluster-bootstrap covariance of the data moments (1,000 replications resampling women) and W its fixed diagonal; the Jacobian is computed by two-sided finite differences on 160,000 simulated women, and the simulation-noise inflation $1+1/S = 1.582$ compares the 10,000 simulated histories the estimator minimized on with the 5,820 women of the estimation sample (Online Appendix OA.11). Parameters are grouped by their role in the model. The implied baseline risk is $p^0_{be} = p_{\max}\lambda_{be}/(1+\lambda_{be})$, with delta-method standard errors; it, not the odds, is the object the moments pin down, and the odds are large at 14–25 only because the estimated risk there is close to the normalisation $p_{\max} = 0.80$. The restricted column imposes $c_{\theta g} = c_g$: the nine adherence increments are fixed at zero and every other parameter is re-estimated. ^{R} restricted to zero by the null. † estimate on a bound of its box, where the estimator is not asymptotically normal; no standard error is reported. ‡ the criterion is flat in the parameter (the weighted Jacobian column is numerically zero), so no standard error is defined. § the standard error moves by more than a quarter between finite-difference steps h and $h/2$ and is reported as unstable.

OA.14 Counterfactual Outcomes by Ability Quartile

Table OA.40 reports, by AFQT quartile, two outcomes of the experiments of Section 6: first births by age 17, whose by-21 counterpart is in Table 6, and the share of method users on the long-acting method at ages 18–25, zero in every experiment that does not add it to the menu. The incidence matches the aggregate results of the main text: closing the adherence gap acts on the lowest quartile, the cut in the cost of use on the highest, and take-up of the long-acting method is highest at the bottom, where avoiding a birth is worth most relative to its cost.

Table OA.40. Counterfactual experiments: outcomes by AFQT quartile

			Cost of use		LARC menu		
	Baseline	Adherence = Q4	No cost of use	Partial cut	Colorado	Pill's cost	All users
Panel A: First birth by 17, by AFQT quartile							
Q1	0.249	0.163	0.224	0.230	0.215	0.205	0.128
Q2	0.158	0.129	0.129	0.137	0.145	0.133	0.097
Q3	0.099	0.079	0.079	0.085	0.082	0.073	0.042
Q4	0.058	0.058	0.045	0.048	0.055	0.048	0.027
Panel B: LARC share among users at 18–25, by AFQT quartile							
Q1	0.000	0.000	0.000	0.000	0.409	0.534	0.000
Q2	0.000	0.000	0.000	0.000	0.260	0.489	0.000
Q3	0.000	0.000	0.000	0.000	0.321	0.578	0.000
Q4	0.000	0.000	0.000	0.000	0.079	0.541	0.000

Notes: Experiments as defined in Section 6 and columns as in Table 6, run at the estimated parameters with common random numbers, holding each quartile's 14–17 adherence at its own 18–25 value. Panel A: share of each quartile's women with a first birth by age 17. Panel B: share of method users at ages 18–25 on the long-acting method, each column taken over its own set of users.

OA.15 Decomposition of the Welfare Gains

Table OA.41 decomposes the ex-ante gains of Table 7 into the windfall to existing users, the saving at the choices made under the experiment, and the remainder. The windfall values the baseline choices at the new costs (Laspeyres); the second saving values the experiment's choices (Paasche). For a cut in the cost of use the windfall is a lower bound on the gain by revealed preference, so the remainder is the value of re-optimizing: more or better fertility control and everything downstream of it. For the long-acting method the windfall is zero, since nobody pays its cost at baseline, and the saving on new choices is measured against the pill's cost, the method it nests with. The Additional Results (Section S.6) report the experiments with 14–17 adherence at its estimated values and the realized-utility version of

Table OA.41. Ex-ante welfare gains: the windfall to existing users and the remainder

	Total	Windfall to existing users		Total net of windfall	
		at baseline choices	at new choices	at baseline choices	at new choices
Adherence = Q4					
Q1	29.7	0.0	0.0	29.7	29.7
Q2	12.1	0.0	0.0	12.1	12.1
Q3	18.1	0.0	0.0	18.1	18.1
Q4	0.0	0.0	0.0	0.0	0.0
All	14.7	0.0	0.0	14.7	14.7
No cost of use					
Q1	46.0	39.9	55.6	6.2	-9.6
Q2	64.8	53.8	79.2	11.0	-14.4
Q3	125.2	95.5	158.9	29.6	-33.7
Q4	140.7	104.5	179.8	36.1	-39.2
All	82.9	67.0	101.8	15.9	-18.9
Partial cut (first births by 21 −10%)					
Q1	28.1	25.8	31.9	2.3	-3.8
Q2	37.6	33.6	42.6	3.9	-5.0
Q3	63.4	54.3	70.8	9.1	-7.4
Q4	68.9	58.3	77.1	10.6	-8.2
All	46.1	40.6	51.8	5.5	-5.6
LARC (Colorado)					
Q1	9.9	0.0	-5.4	9.9	15.3
Q2	3.3	0.0	-4.2	3.3	7.4
Q3	5.7	0.0	-6.4	5.7	12.1
Q4	0.7	0.0	-2.2	0.7	2.9
All	5.0	0.0	-4.6	5.0	9.6
LARC (pill's cost)					
Q1	18.5	0.0	0.0	18.5	18.5
Q2	10.3	0.0	0.0	10.3	10.3
Q3	17.7	0.0	0.0	17.7	17.7
Q4	8.1	0.0	0.0	8.1	8.1
All	13.7	0.0	0.0	13.7	13.7
LARC (all users)					
Q1	53.8	0.0	0.0	53.8	53.8
Q2	34.0	0.0	0.0	34.0	34.0
Q3	67.3	0.0	0.0	67.3	67.3
Q4	28.7	0.0	0.0	28.7	28.7
All	44.8	0.0	0.0	44.8	44.8

Notes: Percent of lifetime consumption, ex ante, by AFQT quartile and for all women, with teen adherence at each quartile's 18–25 level. *Total* is the consumption-equivalent variation of Table 7. *Windfall to existing users* is the expected discounted saving on the methods chosen at baseline when their costs change (at baseline choices) or on the methods chosen under the experiment (at new choices), in the same units; the two *net* columns subtract each from the total. Experiments that change no cost have a windfall of zero.

this measure.

OA.16 Sampling Uncertainty of the Counterfactuals

The counterfactual outcomes of Section 6 are functions of the estimated parameter vector, so they inherit its sampling uncertainty. Table OA.42 reports percentile intervals from a parametric bootstrap: 200 parameter vectors are drawn from the estimated sampling distribution of Section 4, each is clipped to the estimation box, and the experiments are re-run on every draw with common random numbers, so that the dispersion across draws is parameter uncertainty alone. One direction of the parameter space is held at its estimate rather than drawn: the high-school increment to the baseline-risk odds at ages 14–25, whose standard error exceeds the width of its admissible range. It is a direction in which the criterion is nearly flat (its standard error exceeds the width of its box), not sampling uncertainty, and a symmetric draw would place mass where the model is a different model; the other parameters are drawn from their distribution conditional on it, which preserves their cross-covariances. The intervals are therefore conditional on that restriction. Percent falls are computed inside each draw, so the interval is for the fall itself.

Two features of the table carry over to the text. First, the effects on births are estimated with useful precision: the fall in first births by 21 from equalizing adherence lies between 9.6 and 20.3 percent, and the fall from the long-acting method at Colorado's take-up between 4.1 and 8.9 percent, an interval that contains the program's estimated effect on teen births. Second, the welfare *levels* are not: the consumption-equivalent gain to the lowest quartile from equalizing adherence has an interval of 8.9 to 45.8 percent of lifetime consumption, because the value of averting an early birth rests on the baseline-risk odds at ages 14–25, which the data pin down poorly. What the draws do settle is the distribution of the gains, which is what the text claims: the lowest quartile gains most from equalized adherence in 91.5 percent of draws, gains least from removing the cost of use in 99.0 percent, and gains most from the long-acting method at Colorado's take-up in 93.0 percent, with the highest quartile gaining least in 97.0 percent. The ranking of the two headline magnitudes is the one comparison the draws do not settle: removing the cost of use lowers first births by 21 by less than equalizing adherence in 60.0 percent of draws and matches it at a smaller cut in the rest, which is why Section 6 describes the two effects as of the same size and rests the comparison on who gains.

Table OA.42. Counterfactual outcomes: point estimates and bootstrap percentile intervals

	Estimate	5th pct	95th pct
Equalise adherence: fall in first births by 21 (%)	16.1	9.6	20.3
No cost of use: fall in first births by 21 (%)	14.2	3.8	18.5
Equalise adherence: CEV, Q1 (%)	29.7	8.9	45.8
Q2	12.1	3.3	20.7
Q3	18.1	5.2	31.8
Q4	0.0	0.0	0.0
No cost of use: CEV, Q1 (%)	46.0	7.7	66.9
Q4	140.7	22.4	321.3
LARC at Colorado take-up: fall in first births by 21 (%)	7.1	4.1	8.9

Notes: Point estimates from Tables 6 and 7, with teen adherence at each quartile's 18–25 level; 5th and 95th percentiles across 200 parametric bootstrap draws from the estimated sampling distribution, clipped to the estimation box, with one flat direction held at its estimate and the remaining parameters drawn conditionally on it (see text). Experiments re-run on each draw with common random numbers; percent falls computed within each draw. The intervals are conditional on the held direction, not unconditional confidence intervals.

OA.17 Use of Large Language Models

In line with the journal's policy, this note records how a large language model was used. In revising the paper (August–September 2026) I used Anthropic's Claude, through its desktop application, as a coding and editing assistant. On the code side, it helped me debug and extend the Julia estimation and post-estimation scripts and the Stata robustness do-files; I ran every script myself and checked the outputs against independent calculations, for example by reproducing the nested-test statistics on my own machine before quoting them. On the writing side, it checked the manuscript for internal consistency—numbers against tables, tables against results files, the model appendix against the code—and edited text from my drafts and outlines and, where I asked for it, proposed rewordings that I then revised. I also used it as a sounding board, to discuss ideas, alternative specifications and the reading of results; the decisions on every one of them were mine. The research question, the model, the identification strategy, the estimation, the interpretation of the results and the final wording are my own. The model did not have access to the NLSY79 microdata beyond the derived analysis files on my computer.