College, cognitive ability, and socioeconomic disadvantage: policy lessons from the UK in 1960-2004

ONLINE APPENDIX

Andrea Ichino Aldo Rustichini Giulio Zanella

April 24, 2024

Appendix to Section 3.1

The 1970 British Cohort Study

The 1970 British Cohort Study (BCS70) is a longitudinal study following the lives of 17,198 individuals born in England, Scotland, Wales, and Northern Ireland in a single week of April 1970. The BCS70 is representative of the British population at each respective age at which participants are observed. After the initial birth survey, conducted in 1970, data were collected in 10 additional surveys of the same participants. The last data collection was carried out in 2021, when the cohort members were 51 years old. Each survey monitored the cohort members' health, education, and economic outcomes. The data can be downloaded using the UK Data Service at this URL `https://ukdataservice.ac.uk`. For additional details see (Dodgeon et al., 2020).

We use the second wave (1975), third wave (1980), and seventh wave (2004), where cohort members were 5, 10, and 34 years old, respectively. In these waves, cohort members completed tests assessing their verbal and logical/mathematical skills. Excluding observations with missing test scores, the usable sample sizes for these waves are 7,479, 6,927, and 9,448. Excluding high school dropouts, the sample sizes are 5,748, 5,308, and 7,270. After imputing the missing scores as described in the Online Appendix to Section 3.3, the usable sample size of subjects with tertiary education or certification that allows them to apply for tertiary education is 7,369 for all waves.

University Statistical Record

The University Statistical Record (USR) contains administrative information on the universe of students enrolled at UK universities between 1972 and 1993. USR was initiated following the Robbins Report, which had stressed the need for better data for the proper design of higher education policies. It was subsequently discontinued and replaced by the Higher Education Statistics Agency (HESA) in 1993. Unfortunately, pre-1993 USR information was not merged into HESA. Out of the initial 8,103,977 person/year records of students enrolled in a UK higher education institution in this period, we keep the 6,889,425 records of white individuals born in the UK, so as to match the final USoc sample. These records correspond, after some minor data cleaning, to 1,523,192 students born during 1948-1976.

Appendix to Section 3.3

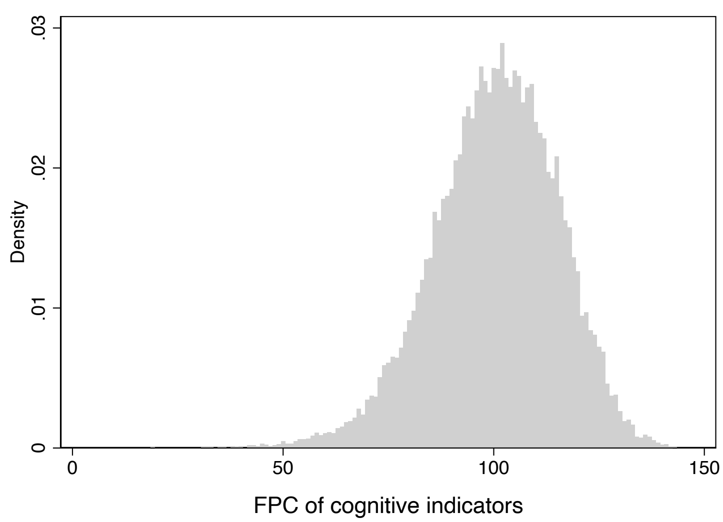
USoc

Table A-1: Eigenvector of the PCA of cognitive ability measures in USoc

Immediate word recall	0.457	Help in immediate word recall	–0.011
Delayed word recall	0.449	Help in delayed word recall	0.004
Correct subtractions	0.318	Help in substractions test	–0.050
Number series	0.413	Help in number series test	–0.034
Verbal ability	0.365	Help in verbal ability test	–0.015
Numeric ability	0.423	Help in numeric ability test	–0.004
		Material aid in recall test	0.011
		Material aid in subtraction test	–0.040

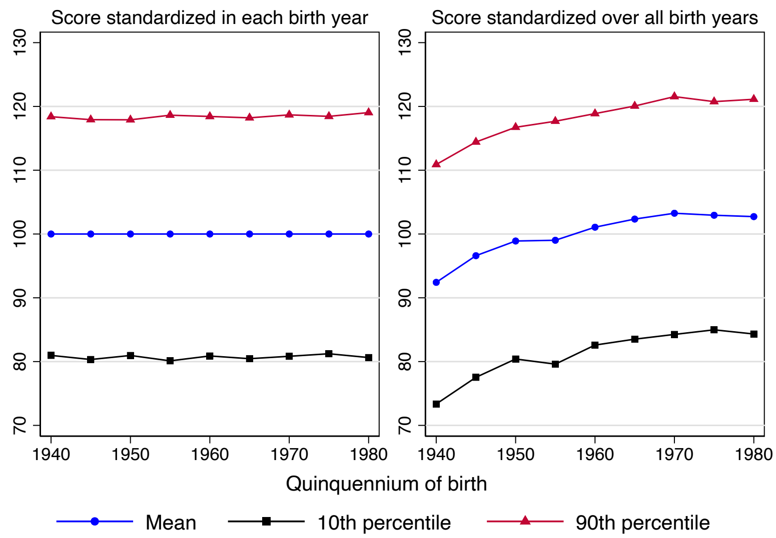
Notes: The table reports the eigenvector of the Principal Components Analysis of the 14 cognitive ability measures contained in the USoc sample with cognitive information described in the central panel of Table 1, which is representative of the UK population. The First Principal Components (FPC) is the measure of ability that we use in our analysis. It has an eigenvalue of 2.55 and explains 18.2% of the data variability. The left panel of the table displays the positive values of the eigenvector terms for the fractions of correct answers in the 6 cognitive questions. The right panel, shows instead that the eigenvector values are negative for 6 out of 8 help dummies. For the the two remaining help dummies the values are positive but close to zero.

Figure A-1: Distribution of cognitive ability in the USoc sample with cognitive information



Notes: The figure illustrates the empirical distribution of the FPC of the 14 cognitive ability variables in the USoc sample with cognitive information described in the central panel of Table 1

Figure A-2: Evolution of the cognitive ability score with different standardizations



Notes: The left panel displays the mean, the 10th and the 90th percentiles of the average ability score standardized within each birth year in the USoc sample of 22,175 white respondents born in the UK between 1940 and 1984, with non-missing education and ability score (see the central panel in Table 1). The right panel displays the mean, the 10th and the 90th percentiles of the average ability score standardized over all birth years in the same USoc sample.

The 1970 British Cohort Study

Imputation in the BCS70

Not all BCS70 subjects have all cognitive scores at all ages and there are several codes for missing information. We explain here in greater detail the different imputation strategies that we follow in the panels of Figure 2 of the main text. In panel (A) or (B) no imputation of such missing information is attempted. In panel (C), the missing answers were coded as follows. First, "Not stated", "no answer", and "more than one answer" were treated as incorrect answers, and "Not applicable" and "not scorable" were coded as missing. Then, the missing answers were imputed as follows:

- If an observation had some missing answers in a specific test, but not all answers were missing, the missing answers were coded to be incorrect.
- If all answers for a test were missing, but the score for at least one test of the same age was available, the score of the missing test was set to be equal to the child's average score from the other tests.
- If all answers to all the tests were missing, the child's score in each test was set to the average score of the other children on the test, conditional on educational attainment.

When creating a general cognitive score using the PCA, a dummy for having all missing in all tests was used.

PCA in the BCS70

Table A-2: Eigenvector of the PCA of cognitive ability measures in BCS70 age 5

	Including imputed scores	Excluding imputed scores
Schonell Reading Test	0.222	0.224
Human Figure Drawing 1	0.515	0.518
Human Figure Drawing 2	0.522	0.525
English Picture Vocabulary Test	0.354	0.340
Complete a Profile	0.319	0.323
Copying Designs	0.432	0.430
Dummy for all tests missing	0.005	–

Notes: The table reports the eigenvector of the Principal Components Analysis of the 6 cognitive ability measures contained in BCS70 sweep 2 (age 5). Two samples are presented: one including imputed scores and one excluding them. For both samples, the First Principal Components (FPC) is the measure of ability that we use in our analysis. It has eigenvalues of 2.56 and 2.51, explaining 36.5% and 41.8% of data variability in the imputed and non-imputed samples, respectively.

Table A-3: Eigenvector of the PCA of cognitive ability measures in BCS70 age 10

	Including imputed scores	Excluding imputed scores
Pictorial Language Comprehension	0.345	0.412
Spelling Dictation Task	0.410	0.395
Friendly Maths Test	0.486	0.469
Edinburgh Reading Test	0.497	0.487
British Ability Scales (BAS)	0.480	0.466
Dummy for all tests missing	−0.002	–

Notes: The table reports the eigenvector of the Principal Components Analysis of the 5 cognitive ability measures contained in BCS70 sweep 3 (age 10). Two samples are presented: one including imputed scores and one excluding them. For both samples, the First Principal Components (FPC) is the measure of ability that we use in our analysis. It has eigenvalues of 2.82 and 3.35, explaining 47% and 67% of data variability in the imputed and non-imputed samples, respectively.

Table A-4: Eigenvector of the PCA of cognitive ability measures in BCS70 age 34

	Including imputed scores	Excluding imputed scores
Literacy Skills	0.705	0.707
Numeracy Skills	0.705	0.707
Dummy for all tests missing	−0.069	–

Notes: The table reports the eigenvector of the Principal Components Analysis of the 2 cognitive ability measures contained in BCS70 sweep 7 (age 34). Two samples are presented: one including imputed scores and one excluding them. For both samples, the First Principal Components (FPC) is the measure of ability that we use in our analysis. It has eigenvalues of 1.67 and 1.65, explaining 55.9% and 82.6% of data variability in the imputed and non-imputed samples, respectively.

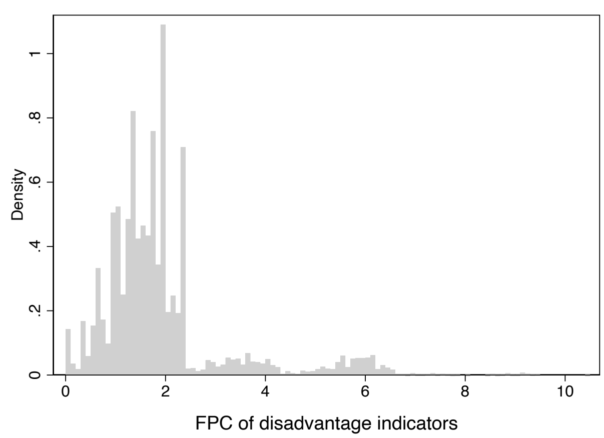
Appendix to Section 3.4

Table A-5: Eigenvector of the PCA of socioeconomic factors generating advantage in college enrollment and graduation in USoc

Father education	0.277
Mother education	0.290
Mother work	0.191
Mother dead	–0.125
Mother absent	–0.224
Father work	0.617
Father dead	–0.416
Father absent	–0.428

Notes: The table reports the eigenvector of the Principal Components Analysis of the eight socioeconomic background variables (referring retrospectively to when the respondent was 14 years of age) on which we base our measure of socioeconomic disadvantage. This analysis is conducted using the USoc sample with cognitive information described in the central panel of Table 1, which is representative of the UK population. The First Principal Component (FPC) has an eigenvalue of 1.76 and explains 22% of the data variability. The table displays negative values for the variables that, as expected, reduce the FPC and increase disadvantage: whether either parent was dead or absent when the respondent was 14 years of age.

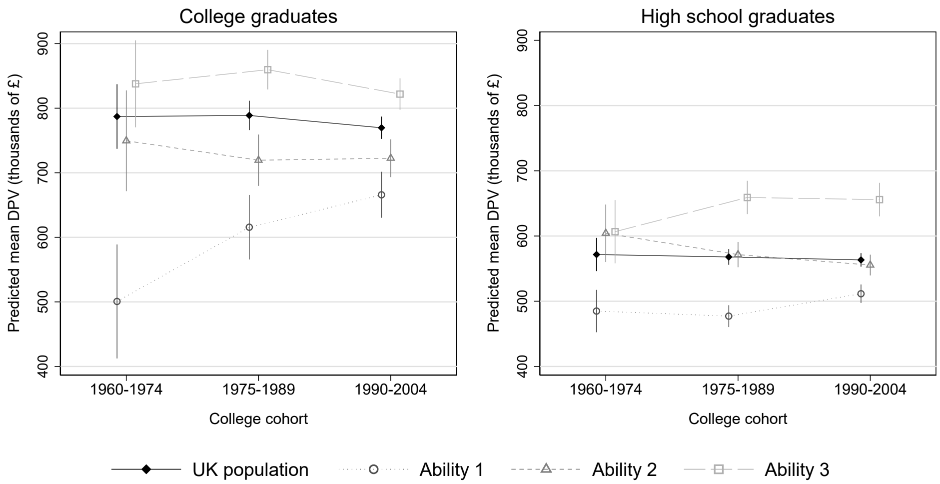
Figure A-3: Distribution of socioeconomic disadvantage



Notes: The figure illustrates the empirical distribution of our measure of disadvantage The measure is the the FPC of 8 socioeconomic variables at age 14, rescaled so that the minimum is zero.

Appendix to Section 4.3

Figure A-4: Evolution of DPV of lifetime earnings by educational attainment and by ability



Notes: The figure shows the the evolution, over the three college cohorts, of the predicted discounted present value (DPV) of lifetime earnings of college graduates (left panel) and high school graduates (right panel), for three ability groups defined by the terciles of the cognitive ability distribution in the UK population. The DPV's are estimated as explained in Section 3.1.

Appendix to Section 5.1

Bootstrap procedure

As mentioned in the text, our parameter estimates and their standard errors are obtained from 1,000 bootstrap replications, in samples obtained from random draws with replacement. We redraw with replacement separately from each college cohort sample. However, the ability and disadvantaged measures are *not* recomputed, so that for each individual in the bootstrap samples these measures are those computed from the PCA in the original sample (as if they were data when bootstrapping). Parameter estimates (as well as model-predicted and empirical values of targeted and untargeted moments) and standard errors are given by, respectively, the mean and standard deviation of each parameter's estimate across the 1,000 replications.

Starting values

It is plausible that our MD estimation criterion function, $\mathcal{C}(\gamma, \tau, \delta, \beta, \alpha_1, \alpha_2, \alpha_3; \rho)$, has local minima, and given that the grid is finite, the "wrong" starting point for the search process may yield estimates that correspond to one of them. This is particularly worrisome because there is no reference scale for policy parameters γ, τ, δ, and β, and no reference value in the literature for for technology parameters α_j estimated in models where productivity depends on both educational attainment and cognitive ability. Thus, one does not know where the grid should be centered in $\mathbb{R}^7$ in order not to get stuck into a local minimum. We solve this problem by noting that: (i) a researcher not interested in disentangling the impact of higher education policy $G = (\gamma, \tau, \delta, \beta)$ from changing technology and socioeconomic characteristics or not interested in using the model for equilibrium policy analysis, can obtain a partial set of estimates by Nonlinear Least Squares (NLS) from the supply-side equation (18), after replacing $\Delta \ln w_j(G)$ with its empirical analog, $\ln \widehat{w}(1, j) - \ln \widehat{w}(0, j)$; and that (ii) the demand-side equation (13) can be used in isolation to calibrate the technological productivity ratios $(\alpha_1, \alpha_2, \alpha_3)$. Specifically, for each j and cohort, we solve equation (13) for α_j and then plug into the resulting equation the empirical values of the odds of college graduation, ξ_j, and the wage ratio, r_j to obtain a numerical value for α_j.

These NLS estimates of $(\gamma, \tau, \delta, \beta)$ and calibrated values for $(\alpha_1, \alpha_2, \alpha_3)$ are reported in Table A-6, and provide a reasonable starting values for our grid-search procedure, even if they ignore the equilibrium effects of higher education policy or of technological progress.

Grid search

Our MD estimates and standard errors are then obtained as follows. Starting from the initial values in Table A-6, we set up a grid to locate the global minimum of criterion function $\mathcal{C}(\gamma, \tau, \delta, \beta, \alpha_1, \alpha_2, \alpha_3; \rho)$, calibrating parameter ρ to the value of 0.584 estimated for the UK by Card and Lemieux (2001). Anchoring the grid search process to these initial values – which are essentially partial-equilibrium guesses of the parameters that we want to estimate – increases our confidence that the MD algorithm – which instead takes into account general equilibrium effects – does not end up at a local minimum.

In order to mitigate the curse of dimensionality, we design an algorithm that starts from a small grid composed by $3^7 = 2,187$ points, 3 for each of the seven parameters (the initial guess and two neighboring points, at distance 0.01 for γ, τ, and the three α_j's, and distance 0.001 for δ and β). We then solve numerically for the model's equilibrium at each point of this grid by finding the unique fixed point of equation (19) for that particular combination of $(\gamma, \tau, \delta, \beta, \alpha_1, \alpha_2, \alpha_3)$, and we obtain a MD estimate by locating the minimum of $\mathcal{C}(\gamma, \tau, \delta, \beta, \alpha_1, \alpha_2, \alpha_3; \rho)$ over the grid. If this MD estimate hits a grid boundary (for example, if the estimate for γ is the minimum or the maximum in the vector of values for γ that is used to build the grid), then a point is added to enlarge that boundary and estimation is repeated over the expanded grid. This process is iterated until the MD estimates are at an interior point of the grid. The grid is expanded considerably before convergence is achieved, which indicates that partial and general equilibrium estimates may differ in

Table A-6: Initial estimates of policy parameters

[A] NLS on supply-side eq. (18)

	College cohort		
	1960-1974	1975-1989	1990-2004
γ	9.836	3.891	3.990
	(0.806)	(0.877)	(0.623)
τ	-5.825	-1.852	-2.103
	(0.674)	(0.623)	(0.440)
δ	0.000	0.046	0.030
	(0.000)	(0.015)	(0.013)
β	0.001	0.006	0.015
	(0.000)	(0.001)	(0.001)
N	4,716	6,922	6,252

[B] Calibration of demand-side eq. (13)

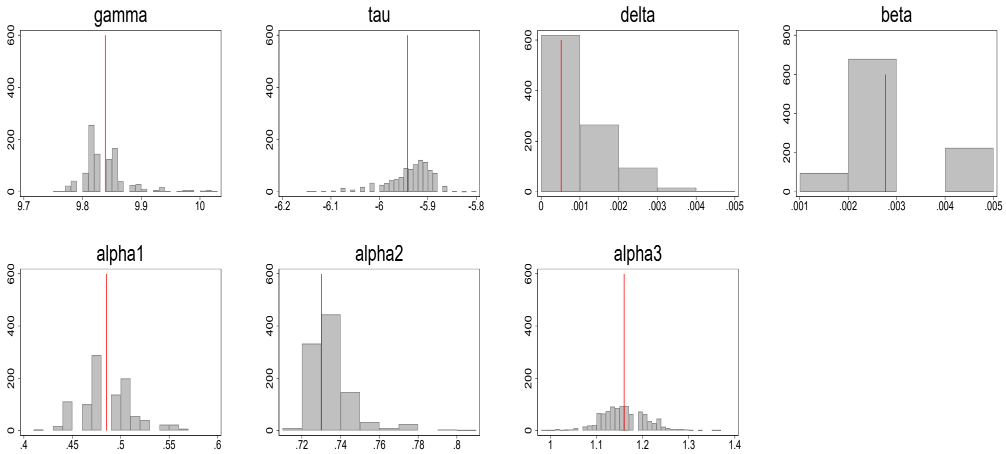
	College cohort		
	1960-1974	1975-1989	1990-2004
α_1	0.448	0.602	0.646
α_2	0.732	0.813	0.949
α_3	1.186	1.225	1.382

Notes: Panel [A] reports Nonlinear Least Squares (NLS) estimates of parameters in equation (18), after replacing $\Delta \ln w_j(G)$ with its empirical counterparts, i.e., $\ln w(1,j) - \ln w(0,j)$, for each ability group j. Sample: USoc, 17,890 white respondents born in the UK in 1940-1984 with non-missing education and ability information and with at least a high school certification that allows for college enrollment (see the right panel of Table 1 in the paper). Panel [B] reports our calibration of the technological productivity ratios across ability groups, α_j, obtaining by solving equation (13) for α_j and then plugging into the resulting equation the empirical values of the odds of college graduation, ξ_j, and the wage ratio, r_j. Parameter ρ is set to the value of 0.584 estimated for the UK by Card and Lemieux (2001).

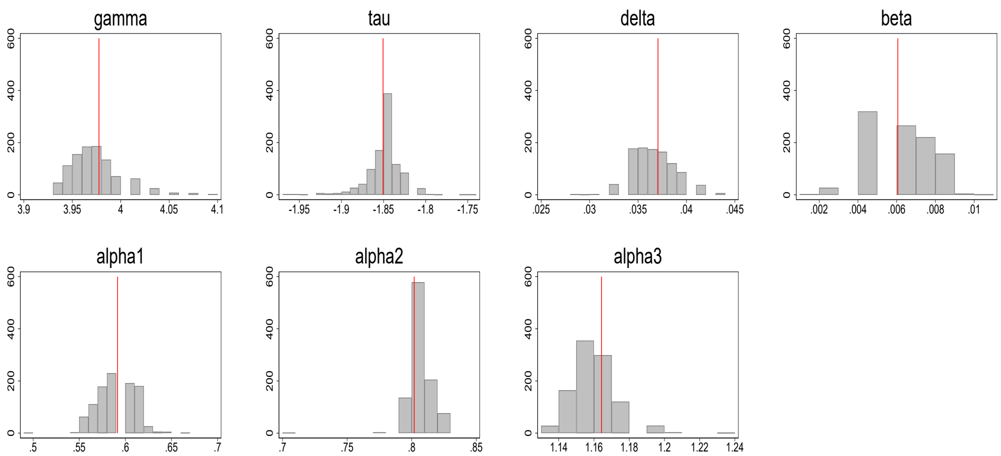
non-negligible ways. Recall that our final estimates result from averaging estimates across 1,000 bootstrap samples. In the initial, actual sample the size of the grids from which we obtain "one-shot" estimates are: 311,850 points for college cohort 1960-1974 (the minimum value of the criterion function is min = 0.0315); 141,004,800 points for cohort 1975-1989 (min = 0.0092); and 19,568,640 points for cohort 1990-2004 (min = 0.0775). Then, to produce bootstrap estimates, we build a new initial grid centered around these one-shot estimates and we iterate according to the "no boundary estimates" rule described above in the 1,000 bootstrap samples. The distribution of the resulting 1,000 bootstrap estimates of each parameter is illustrated in Figure A-5 for three three college cohorts. The vertical lines mark the averages that we report as our point estimates in Table 3 of the main text, and the standard deviations are our bootstrap standard errors in that table.

Figure A-5: Distribution of MD estimates across 1,000 bootstrap samples

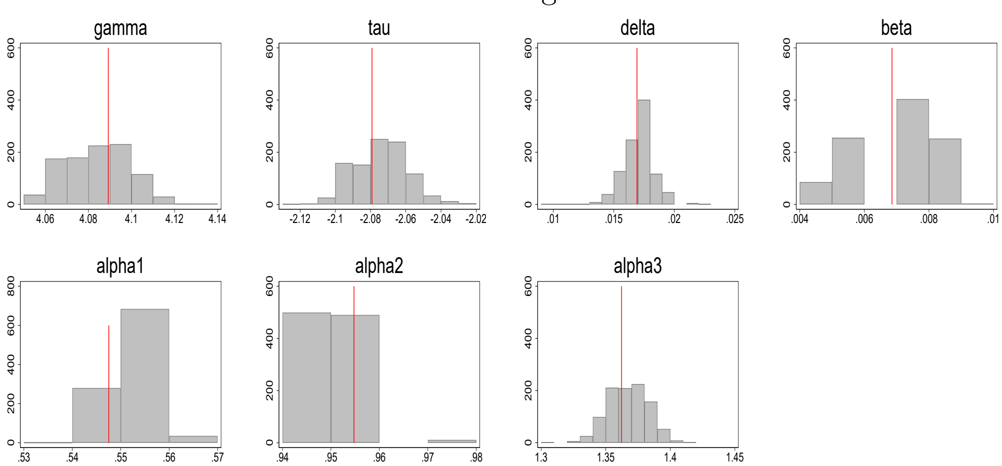
1960-1974 college cohort



1975-1989 college cohort



1990-2004 college cohort



Notes: The figure illustrates the distribution of our MD estimates of the seven structural parameters of interest across 1,000 bootstrap samples, by college cohort. The vertical line is the mean of the distribution. The point estimates and standard errors reported in Table 3 of the main text are the means and standard deviations, respectively, of these distributions.

Identification

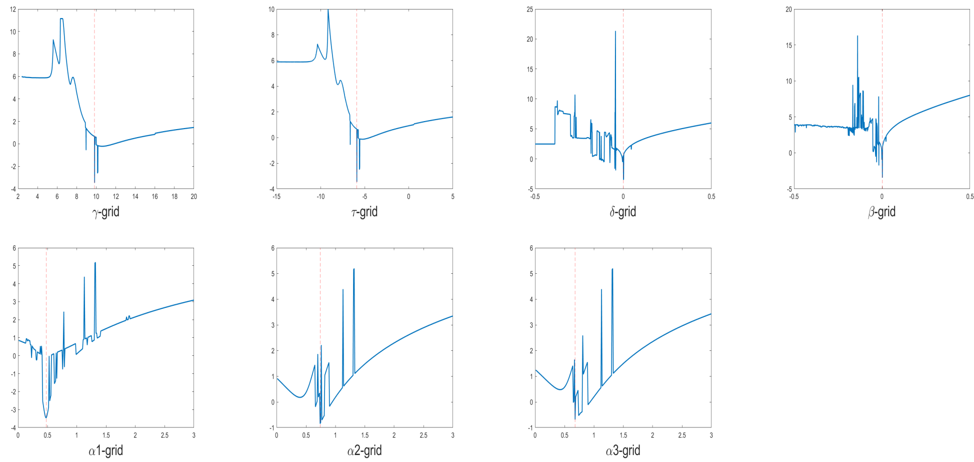
A crucial question in our econometric analysis is whether the MD algorithm that we employ produces estimates that correspond to a global minimum or not. To increase our confidence that it does, we inspect two- and three-dimensional sections of the criterion function over a much wider grid than the one used by the computational algorithm. The two-dimensional sections are shown in Figure A-6 for each college cohort. Each panel in this figure plots the value of the log of the MD criterion, i.e., $\ln \mathcal{C}(\gamma, \tau, \delta, \beta, \alpha_1, \alpha_2, \alpha_3; \rho)$, letting only one parameter at the time vary, while keeping the remaining six parameters (in addition to ρ) fixed, at the value of the "one-shot" MD estimates. Both the global minimum and local minima are clearly visible in each panel. The vertical line corresponds to the free parameter. Note that despite the appearance of a cusp, the function is smooth around the minimum – such appearance is due to the the log scale, which is convenient but produces a large negative value at the minimum, where it is very close to zero. Local minima are clearly visible, but in all cases our MD estimate corresponds to the global minimum.

This exercise can be extended to allowing two parameters to vary while keeping the remaining five (in addition to ρ) fixed at the "one-shot" MD estimates, thus producing three-dimensional sections of the log criterion function that can be represented graphically. Such sections are shown in Figure A-7 for college cohort 1960-1974 and in Figure A-8 for college cohort 1990-2004. The possible $\binom{7}{2} = 21$ combinations are represented for either cohort, and each panel illustrates the contour lines of the log MD criterion. The intersections of the straight continuous and dashed lines represent our MD estimates and the global minimum, respectively. Again, local minima are visible and we avoid them with the only exception of the first panel in Figure A-7. In this case, our MD algorithm selects a local minimum for parameters γ and τ. Yet this local minimum is sufficient close to the global one to leave our conclusion that from the 1960-1974 cohort to the 1990-2004 cohort γ declined and τ increased substantially (which results a in a non-meritocratic expansion policy) unchanged.

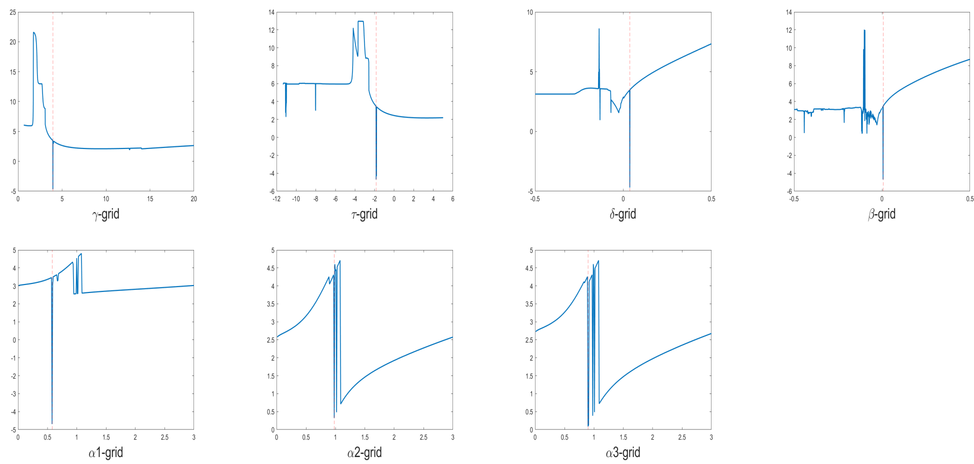
Finally, we perform the following check on our identification: first we set the parameters at some value, and we use these parameters to generate a simulated data set from the model; then we use our MD algorithm to estimate the parameters on the simulated data to see whether the algorithm recovers the parameters that generated the data. For this exercise, we use the parameters that we employed to simulate the counterfactual ME policy represented in the left panel of Figure 9. Table Table A-7 contrasts the "true" and the estimated parameters for college cohort 1990-2004. Although there is a notable discrepancy for β (which is over-estimated due to the nonlinear dependence of τ on students' ability in a context characterized by $\eta < 0$), the MD algorithm does a satisfactory job. This reinforces our confidence that the targeted moments are sufficiently informative in our model to identify the parameters. The table's notes provide more details.

Figure A-6: 2D sections of criterion function, log scale

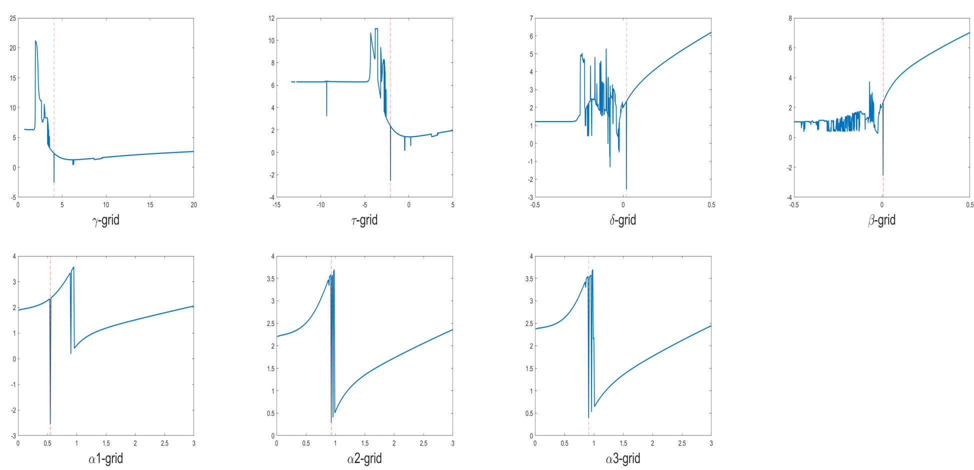
1960-1974 college cohort



1975-1989 college cohort

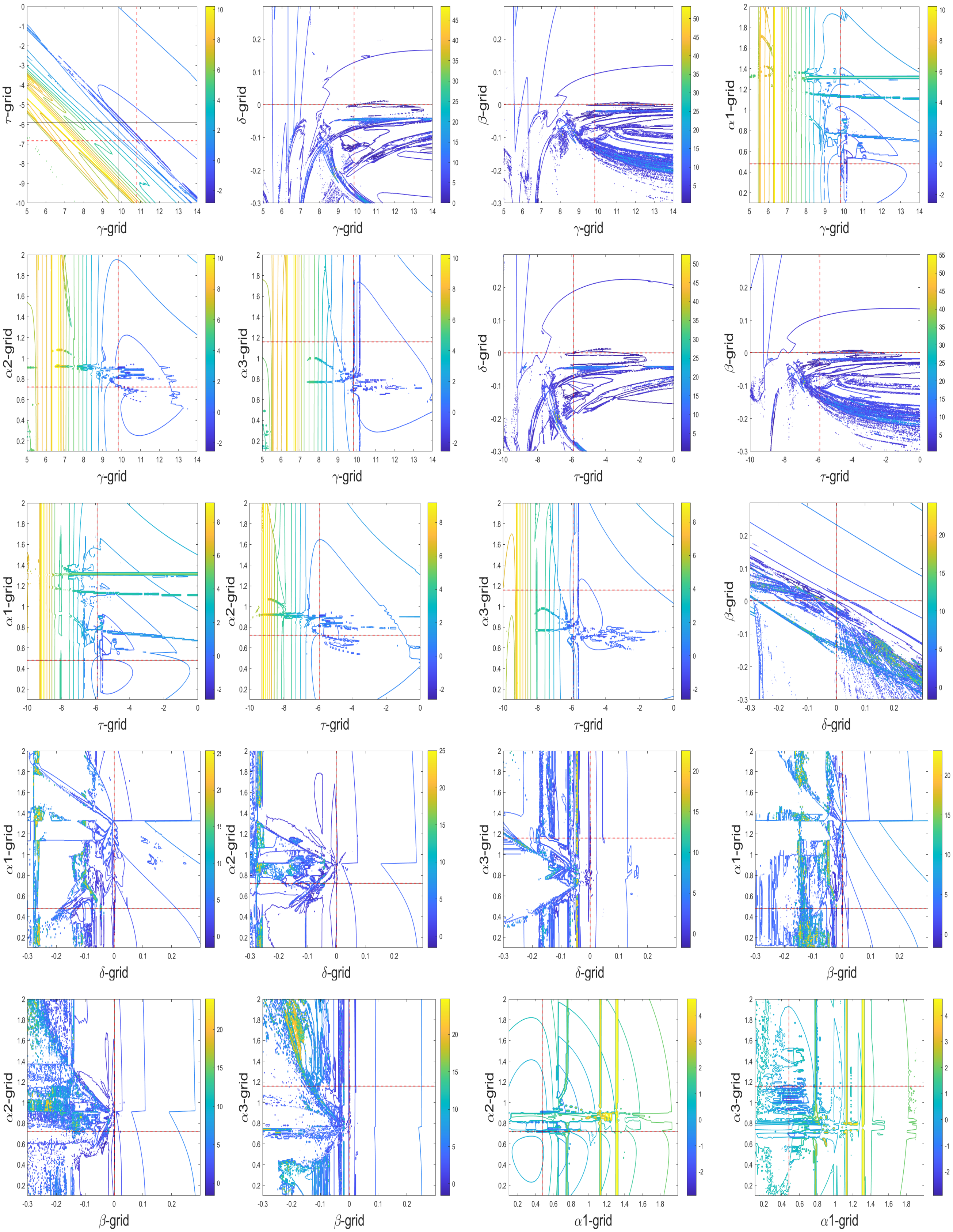


1990-2004 college cohort



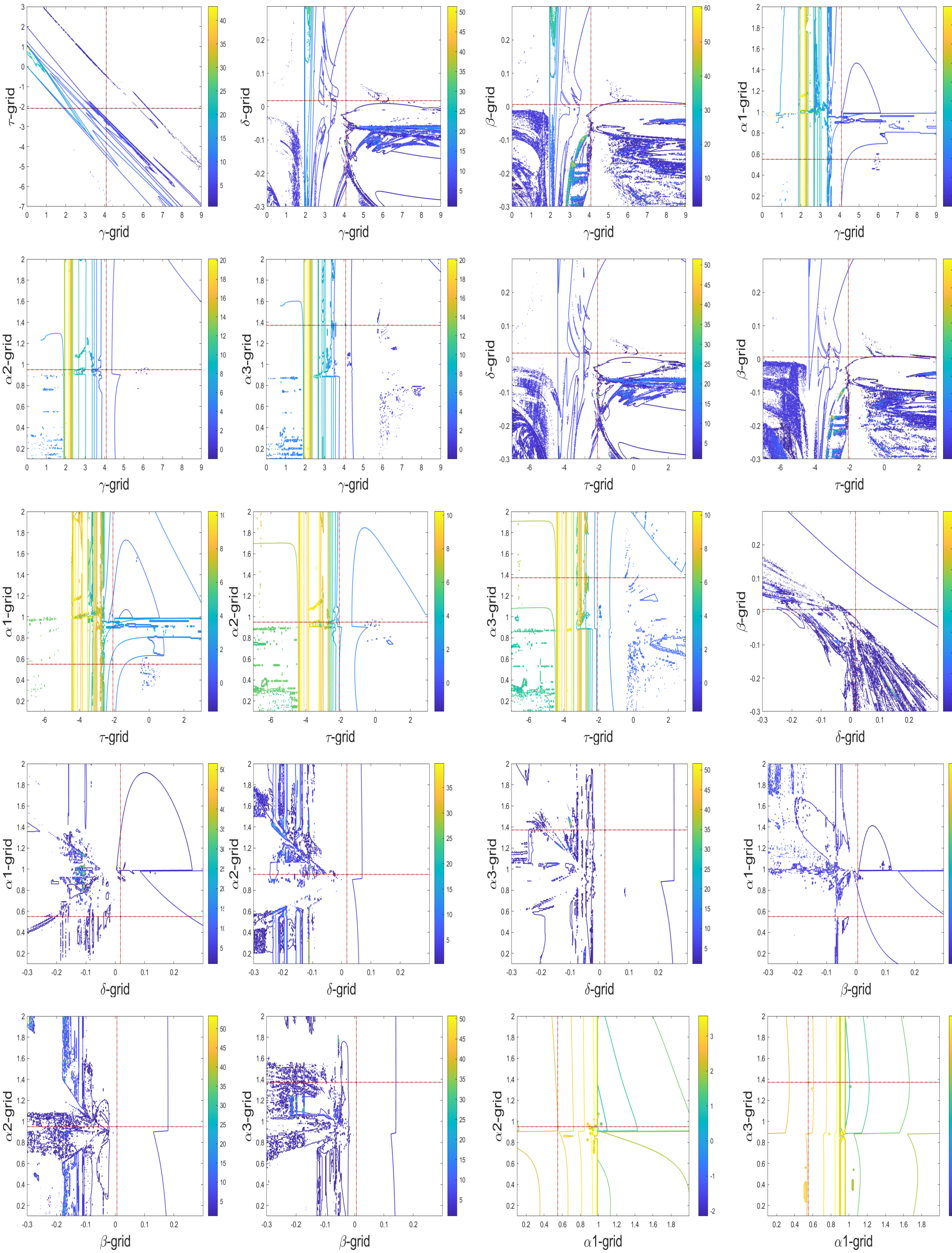
Notes: Each panel plots the value of the log of the MD criterion as a function of one parameter, keeping the remaining six parameters fixed at the MD estimates obtained in the actual (as opposed to bootstrap) sample. The dashed line marks the global minimum, which corresponds to our "one-shot" MD estimate. Local minima are clearly visible, and anchoring the grid to our initial estimates (see Table A-6) helps avoiding them. Despite the appearance of cusps, the function is smooth around the global minimum, which takes a large negative value on the log scale because it is close to zero.

Figure A-7: 3D sections of criterion function for cohort 1960-1974, log scale



Notes: see Appendix text

Figure A-8: 3D sections of criterion function for cohort 1990-2004, log scale



Notes: see Appendix text

Table A-7: Estimating known parameters from simulated data

	College cohort: 1990-2004	
	True (data-generating) parameters	Estimated parameters
γ	9.078	9.170
τ	$-6.3 \cdot \mathbb{I}[\theta > \mathbb{E}[\Theta]]$	-6.260
δ	0.000	0.001
β	0.000	0.011
α_1	0.55	0.42
α_2	0.96	0.90
α_3	1.36	1.27
N	6,252	6,252

Notes: The table shows the result of a check on our identification. We first set the parameters at the values that generate the data for the effects of the counterfactual ME policy on the 1990-2004 cohort represented in the left panel of Figure 9. Parameter ρ is still set to the value of 0.584 estimated for the UK by Card and Lemieux (2001). Then, we use our MD algorithm to estimate the parameters on the simulated data. The grid is centered around the true parameters (6.3 for τ), and the algorithm expands the initial grid of $3^7 = 2,187$ points to a grid of 56,310,800 points over 18 iterations. The algorithm does a satisfactory job at recovering the parameters that generated the data, indicating that the targeted moments are sufficiently informative in our model to identify the parameters. The overestimate of parameter β reflects the nonlinear dependence of τ on students' ability in a context characterized by $\eta < 0$. That is, the penalty on low-ability students implied by $\tau = -6.3 \cdot \mathbb{I}[\theta > \mathbb{E}[\Theta]]$ is picked up by a strictly positive β that penalizes disadvantaged students, in line with the negative empirical correlation between ability and disadvantage in our data.

Appendix to Section 5.2

Table A-8: Empirical vs model-predicted targeted moments

	[A] College graduation odds				[B] Wage ratio		
	1960-1974	1975-1989	1990-2004		1960-1974	1975-1989	1990-2004
				Ability group 1			
		ξ_1				r_1	
model	0.137	0.167	0.174	model	1.110	1.248	1.134
	(0.013)	(0.025)	(0.005)		(0.036)	(0.034)	(0.005)
data	0.140	0.162	0.186	data	1.028	1.289	1.297
	(0.014)	(0.012)	(0.013)		(0.099)	(0.058)	(0.040)
				Ability group 2			
		ξ_2				r_2	
model	0.298	0.352	0.562	model	1.208	1.239	1.214
	(0.012)	(0.035)	(0.007)		(0.028)	(0.027)	(0.009)
data	0.276	0.343	0.467	data	1.271	1.262	1.300
	(0.003)	(0.018)	(0.024)		(0.083)	(0.041)	(0.033)
				Ability group 3			
		ξ_3				r_3	
model	0.641	0.799	1.297	model	1.396	1.279	1.223
	(0.022)	(0.015)	(0.039)		(0.057)	(0.010)	(0.008)
data	0.693	0.873	1.268	data	1.379	1.299	1.254
	(0.003)	(0.034)	(0.062)		(0.084)	(0.036)	(0.030)

	[C] Cognitive Ability				[D] SES disadvantage		
	1960-1974	1975-1989	1990-2004		1960-1974	1975-1989	1990-2004
				College graduates			
		$\mathbb{E}(\Theta \vert K=1)$				$\mathbb{E}(\Lambda \vert K=1)$	
model	110.3	108.9	108.4	model	1.97	1.58	1.32
	(0.4)	(0.4)	(0.3)		(0.03)	(0.03)	(0.03)
data	110.3	109.0	108.2	data	1.85	1.58	1.21
	(0.4)	(0.3)	(0.03)		(0.03)	(0.03)	(0.03)
				High school graduates			
		$\mathbb{E}(\Theta \vert K=0)$				$\mathbb{E}(\Lambda \vert K=0)$	
model	101.8	99.6	96.9	model	2.17	1.88	1.74
	(0.3)	(0.2)	(0.3)		(0.02)	(0.02)	(0.03)
data	101.8	99.5	97.3	data	2.22	1.88	1.79
	(0.2)	(0.2)	(0.2)		(0.02)	(0.02)	(0.02)

Notes: The table reports the model-predicted and the empirical values of the ten targeted moments in our minimum distance (MD) estimation of the policy and technology parameters reported in Table 3 in the paper: the college graduation odds and wage ratio in the ability groups defined by the tercile of the ability distribution in the UK population, ξ_j and r_j, respectively, for $i = 1, 2, 3$); and the average ability and disadvantage of college and high school graduates. The MD criterion function is given by equation (20) in the paper, and the weighting matrix is the identity matrix. The college cohorts to which columns refer are defined by the period of actual or potential college attendance, which is an individual's age plus 20. Cross-sectional response weights are applied. Sample: USoc, 17,890 white respondents born in the UK in 1940-1984 with non-missing education and ability information and with at least a high school certification that allows for college enrollment (see the right panel of Table 1 in the paper.)

Table A-9: Empirical vs model-predicted untargeted moments

	Ability distribution			**Disadvantage distribution**		
	College cohort			College cohort		
	1960-1974	1975-1989	1990-2004	1960-1974	1975-1989	1990-2004
	College, 25th percentile			*College, 25th percentile*		
model	102.5	101.2	100.3	1.51	1.02	0.79
	(0.5)	(0.5)	(0.4)	(0.03)	(0.02)	(0.03)
data	103.0	101.6	100.3	1.31	0.98	0.64
	(0.5)	(0.3)	(0.4)	(0.03)	(0.01)	(0.01)
	College, 75th percentile			*College, 75th percentile*		
model	118.2	117.6	117.0	2.22	1.84	1.50
	(0.5)	(0.4)	(0.3)	(0.06)	(0.02)	(0.02)
data	117.7	117.6	117.0	2.04	1.79	1.46
	(0.4)	(0.3)	(0.3)	(0.04)	(0.02)	(0.01)
	High school , 25th percentile			*High school , 25th percentile*		
model	93.8	91.3	88.3	1.61	1.24	0.98
	(0.3)	(0.3)	(0.3)	(0.02)	(0.02)	(0.00)
data	93.9	91.3	88.9	1.64	1.33	1.02
	(0.3)	(0.3)	(0.3)	(0.00)	(0.02)	(0.02)
	High school , 75th percentile			*High school , 75th percentile*		
model	111.3	109.4	106.7	2.31	1.95	1.74
	(0.3)	(0.2)	(0.3)	(0.00)	(0.01)	(0.01)
data	111.2	109.4	107.0	2.31	1.95	1.74
	(0.2)	(0.2)	(0.3)	(0.00)	(0.01)	(0.01)

Notes: The table reports the mean and standard deviation over 1,000 bootstrap samples (at the respective minimum-distance estimates) of model-predicted vs empirical values of eight untargeted moments targets. A college cohort is defined by the period of actual of potential college attendance, which is an individual's age plus 20. Cross-sectional response weights are applied. Sample: USoc, 17,890 white respondents born in the UK in 1940-1984 with non-missing education and ability information and with at least a high school degree that permits college enrollment (see the right panel of Table 1) in the paper).

Robustness to the inclusion of the "Big Five" personality traits

Using the "Big Five" (B5) personality traits to produce alternative measures of ability or disadvantage does not alter our conclusions. This is shown in Table A-10 and Figure A-9.

In the table, row 1 reproduces (for convenience) the correlations displayed in Table 2 between ability Θ and disdvantage Λ, as measured in the main text without using the B5. The remaining rows report this correlations when: the B5 traits are added to the socioeconomic indicators in the PCA for disadvantge (row 2), only the B5 variables contribute to the PCA for disadvantage (row 3), the B5 variables are added to the cognitive ability indicators in the PCA to produce a broader ability measure (row 4).

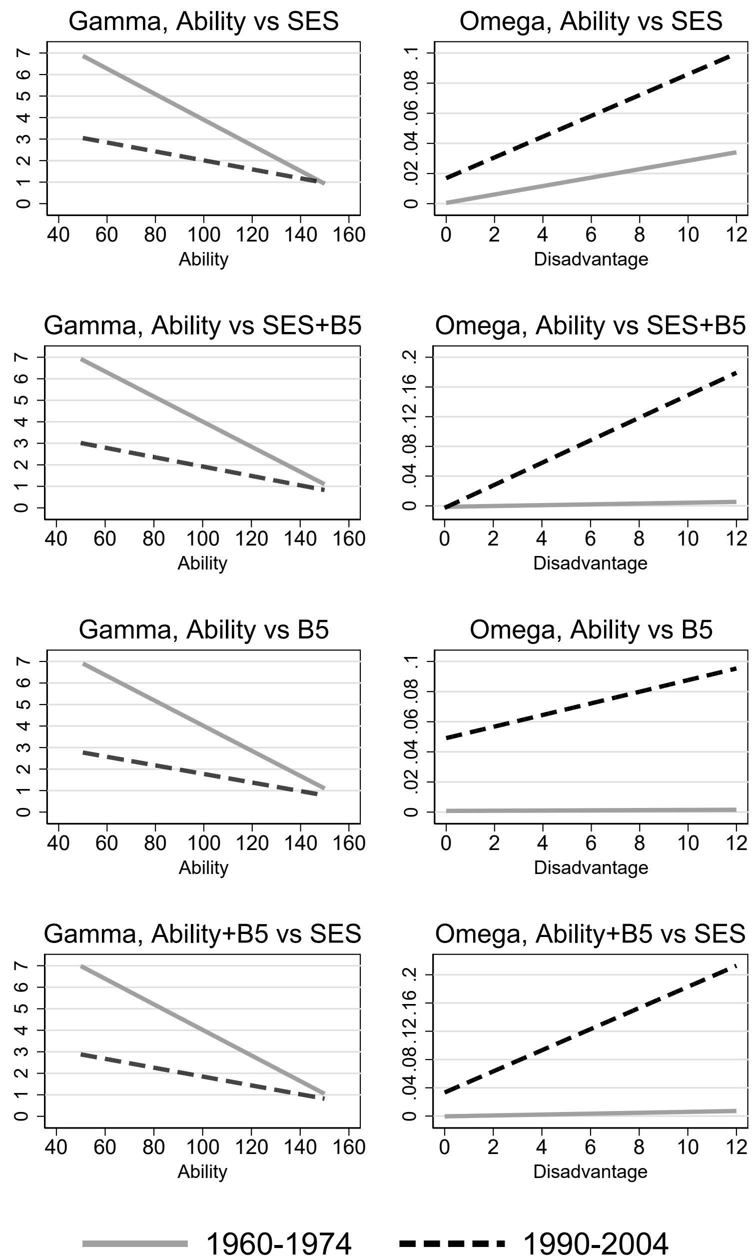
Figure A-9 follows the same exposition pattern to describe how the cost shifts $\Omega(\Lambda)$ and $\Gamma(\Theta)$ changed between the first and the final college cohorts, depending on whether and where the B5 variables are included. In all cases our central conclusion stands: the cost of study effort decreased more for low-ability students than for high-ability ones and increased more for high-disadvantage students than for low-disadvantage ones.

Table A-10: Correlation η between alternative measures of cognitive ability Θ and socioeconomic disadvantage Λ in USoc

	College cohort		
	1960-1974	1975-1989	1990-2004
Θ = FPC(cognitive variables); Λ = FPC(SES variables)	−0.129	−0.143	−0.139
	(0.016)	(0.014)	(0.014)
Θ = FPC(cognitive variables); Λ = FPC(SES & B5 variables)	−0.131	−0.151	−0.150
	(0.016)	(0.014)	(0.014)
Θ = FPC(cognitive variables); Λ = FPC(B5 variables)	−0.029	−0.047	−0.041
	(0.016)	(0.014)	(0.015)
Θ = FPC(cognitive & B5 variables); Λ = FPC(SES variables)	−0.127	−0.150	−0.141
	(0.016)	(0.014)	(0.013)
N	4,716	6,922	6,252

Notes: The table reports the correlation between alternative measures of ability (Θ) and disadvantage (Λ) that combine in different ways cognitive ability variables, socioeconomic status (SES) variables at the time the respondent was 14, and the "Big Five" (B5) personality variables available in USoc: in the first row, Θ is the FPC of cognitive variables and Λ is the FPC of SES variables (this first row is identical to Table 2 in the paper); in the second row, Θ is the FPC of cognitive variables and Λ is the FPC of SES and B5 variables; in the third row, Θ is the FPC of cognitive variables and Λ is the FPC of B5 variables; in the fourth row, Θ is the FPC of cognitive and B5 variables, and Λ is the FPC of SES variables. Standard errors are produced via the delta method. Cross-sectional response weights are applied. Sample: USoc, 17,890 white respondents born in the UK in 1940-1984 with non-missing education and ability information and with at least a high school degree that permits college enrollment (see the right panel of Table 1) in the paper).

Figure A-9: Estimated study effort cost shifts under alternative definitions of ability and disadvantage that include the “Big Five” (B5) personality traits



Notes: The figure shows the study effort cost shifts $\Gamma(\cdot)$ and $\Omega(\cdot)$ implied by NLS estimates of the policy parameters, for the 1960-1974 or 1990-2004 college cohorts, as a function of ability (left panel) or disadvantage (right panel), using alternative definitions of ability and disadvantage that combine in different ways cognitive ability variables, socioeconomic status (SES) variables at the time the respondent was 14, and the “Big Five” (B5) personality variables available in USoc, following the same order of Table A-10. To facilitate a comparison, the first row corresponds to the definition employed in the paper, i.e., ability is the FPC of cognitive variables and disadvantage is the FPC of socioeconomic variables. This first row is identical to Figure 7 in the paper. In the second row, ability is the FPC of cognitive variables and disadvantage is the FPC of SES and B5 variables; in the third row, ability is the FPC of cognitive variables and disadvantage is the FPC of B5 variables; in the fourth row, ability is the FPC of cognitive and B5 variables, and disadvantage is the FPC of SES variables.

Appendix References

Card, D. and T. Lemieux (2001): "Can falling supply explain the rising return to college for younger men? A cohort-based analysis," *The Quarterly Journal of Economics*, 116, 705–746.

Dodgeon, B., J. Elliott, M. Hancock, J. Johnson, S. Parsons, A. Peters, and P. Shepherd (2020): "1970 British Cohort Study Age 34 Sweep User Guide," UCL Centre for Longitudinal Studies.