Inequality of Conditions and Intergenerational Mobility:
Changing Patterns of Educational Attainment in the United States

Stephen L. Morgan
Cornell University

Young-Mi Kim
Cornell University

Supplementary Appendix

Selection of the Sample

We analyze data from the Survey of Income and Program Participation (SIPP), collected by the U.S. Census Bureau in 1986 and 1996 (see U.S. Dept. of Commerce 2001). In particular, we analyze March and November reference month data from the first year of the separate 1986 and 1996 SIPPs.[1] As shown in Table S1, the first wave of the 1986 panel includes 30,577 respondents from 11,454 households, while the first wave of the 1996 panel includes 95,141 respondents from 36,730 households.

[INSERT TABLE S1 ABOUT HERE]

From this baseline sample, we selected for our analysis of college enrollment patterns all households that include SIPP respondents who (1) were between the ages of 17 and 21 in March, (2) who remained in the sample for both March and November, and (3) who were unmarried in November.[2] As shown in Table S1, in 1986 there were 1,900 respondents meeting these criteria (living in 1,500 households), while in 1996 there were 4,994 respondents (living in 4,128 households).

Table S2 breaks down the selected age group (i.e., the group of households in the third line of Table S1) by family structure, presenting the mean age and college enrollment patterns for respondents from each type of family. The table shows to some extent why we decided to exclude married individuals between the ages of 17 and 21 from our analysis. These respondents tended to be older and had substantially lower rates of college enrollment. Thus, even if we had been able to ascertain the family income and wealth of their parents (which, as we describe below, we could not), it is unclear whether married SIPP respondents between the ages of 17 and 21 should be thought of as members of the set of adolescents at risk of entering college. If we had included these married respondents in that age range, the effect of the covariates in our models would likely have been larger to some small degree (as marital status is negatively associated with those variables and with college enrollments). Thus, any bias is a conservative bias (at least with

---

[1] For the 1986 SIPP, we analyzed the wave 1 through 4 core files and the wave 4 assets and liability topical module file. For the 1996 SIPP we analyzed the wave 1 through 3 core files and the wave 3 assets and liability topical module file. All 4 rotation groups are used for both panels.

[2] We also excluded respondents who entered the panel after the first wave.

respect to rejecting null hypotheses of zero coefficients), and, most important for us, should be similar across cohorts.

[INSERT TABLE S2 ABOUT HERE]

Some detail on the rules for determining residence in the SIPP should be noted. When a SIPP interviewer approaches a household and develops a roster for all individuals in the household, individuals who are away from home and in college are included on the roster. Likewise, when a SIPP interviewer approaches a household of college students living together, students are eliminated from the roster for that household if they could be listed as permanent members of their parents' households (which may reduce the number of individuals in the nominal household to zero, thereby ending the interview). Finally, if a respondent between the age of 17 and 21 is living without his or her parents and is not enrolled in college, then he or she is considered independent of his or her parents.

As a result, most college students between the ages of 17 and 21 in the SIPP are listed as members of their parents' households. However, a fair number of our older respondents who are not enrolled in college are living away from their parents. As we describe below, we had to impute the parents' education, income, and wealth variables for this latter group of individuals, as our primary research questions focus on the associations between parental resources and the college enrollment behavior of their children.

Coding of Variables

Imputation of Parents' Education, Income, and Wealth. Values for the composite parents' income and wealth variables were imputed separately for (1) respondents living in the same household with their parents, (2) respondents who did not live with their parents but who represented the primary family of the household in which they resided, and (3) respondents who did not live with their parents and who were not the primary family of the household in which they resided.

Respondents who lived with at least one of their parents (about 87% of the 1986 sample and about 85% of the 1996 sample) had very few missing values on these variables, and these values were imputed based on the information that other respondents having the same family structure provided (using best-subset regression imputation with variables for race, parents' education, family income, SEI score of parents' occupation, and wealth).

For respondents who did not live with their parents, and thus who did not have parents' income and education information, we imputed the missing values for parents' education and family income using best-subset regression imputation with the variables for respondent's sex, race, respondents' years of education, respondents' monthly income and home ownership. The sample used for the imputation included households where respondents were living with their parents. Thus, if it is the case (as seems likely) that respondents living away from home and not enrolled in college are different from those still living at home but enrolled in college, then our imputations are biased toward the values of households in which the children are living with their parents. It is hard to know the consequences of this bias for our results.

There was one additional complication for imputing the parents' wealth variables. The

wealth variable is a household level variable rather than an individual level variable. Thus, many respondents living in a household without their parents had seemingly valid wealth variables, but these referred to their own wealth rather than those of their parents. We therefore set these seemingly valid wealth variables equal to missing and imputed wealth just as parents' education and family income.[3]

Topcodes. All of the composite income and wealth variables in the SIPP are topcoded by the Census Bureau to protect the anonymity of respondents. This topcoding creates two problems: (1) the upper tails of the distributions of the variables are distorted, making it impossible to calculate the means of the variables and (2) the topcodes change from year to year, potentially generating misleading over time comparisons.

To minimize these problems, we developed our own common topcodes for all income and wealth variables in 1986 and 1996 (in order to minimize problem (2)) and then we calculated yearly multipliers for the topcodes using Pareto-inspired-imputation (see Klein, 1962:150-4) (in order to minimize problem (1)). Table S3 shows that our procedure places 1-2 percent of the sample in 1986 and 2-4 percent of the sample in 1996 at the respective topcodes. These individuals are then given values on the variable which equal the topcode multiplied by the values in the final column of Table S3, which range from 1.5 to 2.66. However, as described in the main text, we then mechanically censored the income and wealth variables at their 95th percentiles for the logit models. See Table S4.

[INSERT TABLES S3 AND S4 ABOUT HERE]

Basic Descriptive Patterns

Our primary outcome variable is college enrollment status as of November 1986 and November 1996 respectively. The first row of Table S5 presents the mean of this dummy variable for both cohorts. The remaining rows of Table S5 present the means (and, where relevant, the standard deviations) of all of the other variable used to model college enrollment patterns.

[INSERT TABLE S5 ABOUT HERE]

Table S6 cross-classifies the income and wealth variables by family structure, before imputations but after adjustments for topcoding (and also without any mechanical censoring at the

---

[3] Further complications emerge for respondents who are not members of the primary family of the household in which they live. In this case, it is even harder to determine whether the reported household wealth variable is a valid measure of parents' wealth. To deal with these problems, we chose different imputation equations for the two groups after setting any reported household wealth variables equal to missing. If respondents were designated as members of the primary family of the household (about 7.5% of 1986 sample and 14.5% of 1996 sample), respondent's sex, age, race, imputed parents' education, income and SEI are used to estimate parents' wealth levels. If they are the secondary family of the household (about 4.5% of 1986 sample and 0.3% of 1996 sample), we used only the respondent's individual characteristics such as sex, age, race, years of education, and family monthly income.

95th percentile). The income and wealth of two-parent families is considerably higher than those of single-parent families. Respondents living away from their parents had substantially lower mean income and wealth than all other types of families, and, as we discussed earlier in the imputation section, we did not interpret these seemingly valid values as accurate measures of parental characteristics. So, we set them equal to missing and imputed as described earlier. After imputation, the means of these variables increased. For example, net worth and home equity increased, even though imputed values were still below the means for the full sample. The imputed mean net worth for this group is $67,118 for 1986 and $86,770 for 1996. And the imputed mean home equity is $31,657 for 1986 and $39,353 for 1996.

[INSERT TABLE S6 ABOUT HERE]

Table S7 reveals some of the complexity that we encountered when developing a variable for prior enrollment in March of each year. We needed this variable in order to generate results as close as possible to typical estimates of college entry rates. This was, however, somewhat difficult with the SIPP because there are slight changes in the questionnaires eliciting enrollment status between the 1986 and 1996 panels. For the 1986 panel, the question elicited the highest grade attended by March 1986, while for 1996 the analogous question elicited the highest grade completed. Thus, while the 1986 questionnaire gave a straightforward way to measure enrollment status, we had to make additional untestable assumptions for the 1996 data. For example, we used the November college enrollment response data in combination with the highest grade completed response to sort respondents. A respondent who indicated that he or she had completed the 11th grade as of March 1996, but who then reported attending college in November of 1996, was assumed to have been enrolled in the 12th grade in March of 1996. However, as can be seen in Table S7, we were able to obtain a distribution of prior enrollment for the 1996 survey which was somewhat close to that of the 1986 survey (although inevitably some differences remained, especially for high school seniors).

[INSERT TABLE S7 ABOUT HERE]

Estimation of Social Class as a Predictor of College Enrollment

In order to estimate the effect of the class position of a college-eligible SIPP respondent, we first decided to restrict the sample to respondents living with their parents. Although we felt somewhat comfortable imputing parents' education, income, and wealth for prospective students living away from home, we were less comfortable imputing a categorical variable such as social class, which we would then be parameterizing in the enrollment models with dummy variables. As a result, we dropped between 13 and 15 percent of the sample.

We then attempted to assess the consequences of focusing on this subsample. When estimating the models from Tables 4 and 5 in the main paper with this reduced sample, we obtained the same basic results. This gave us a bit of confidence, since those living away from their parents (by SIPP design) were not generally enrolled in college. But, it must be recognized that the meaning of this lack of change (i.e., using a subsample of students still categorized as

living with their parents, even though they may be away at school) is contingent on whether the imputation scheme for parental characteristics was effective for the results in Tables 4 and 5. This is, by its very nature, unknowable.

We also dropped respondents who lived in households with their parents but whose parents were out of the labor force and did not have values for current occupation. Again, estimating the models from Tables 4 and 5 on this further reduced sample yielded the same basic results. But, in the end, it must be recognized that the models for Table 6 are estimated for a non-random subsample of 70% of the respondents for the models in Tables 4 and 5.

As for coding this 70% of the sample, we implemented our own coding of what has become the dominant schema in the social mobility literature – variously referred to as the EGP schema (after Erikson, Goldthorpe, Portacarero 1979), Goldthorpe's class schema (after Goldthorpe 1987), or the CASMIN coding (after Goldthorpe and Müller 1982) – which has been effectively deployed in a wide variety of substantive contexts, most prominently in studies of social mobility (e.g., Erikson and Goldthorpe 1992; Hout 1989) and voting (see Heath, Jowell, and Curtice 1985; Manza and Brooks 1999).

The following excerpts from *The Constant Flux* (Erikson and Goldthorpe 1992:41-42) describe each of the EGP classes.[4] Following most of the excerpts are examples, or other information, intended to give the reader a sense of each class.

Class I. "Higher-grade professionals, administrators, and officials; managers in large industrial establishments; large proprietors." This class includes: (1) professional occupations regardless of employer size and (2) managers regardless of employer size (excluding some service managers; see discussion below).[5]

Class II. "Lower-grade professionals, administrators, and officials; higher-grade technicians; managers in small industrial establishments; supervisors of non-manual employees." This class includes: (1) Occupations that are considered white collar but are not as esteemed as what are normally thought of as the professions (e.g., nurses and medical technicians); (2) private or publicly employed managers; (3) some service managers regardless of employer size; (4) supervisors of non-manual workers (e.g., supervisors of financial records processing).

Class III. "Routine non-manual employees, higher-grade (administration and commerce) and "Routine non-manual employees, lower-grade (sales and service)." Examples: Secretaries, sales workers (retail and personal), attendants at amusement and recreation facilities.

Class IV. "Small proprietors, artisans, etc., with employees," "Small proprietors, artisans, etc., without employees," "Farmers and small-holders; other self-employed workers in primary production."

Class V. "Lower-grade technicians; supervisors of manual workers." Examples: Dental

---

[4] We were unable to use employer size in our coding because the SIPP does not have employer-size variables. This limitation only effects managers, who we then had to assign to class I.

[5] We are unable to include in Class I Goldthorpe's large proprietors (e.g., individuals of any occupation who employ 25 or more employees) because SIPP do not have employer-size variables.

hygienists, supervisors of material moving equipment operators.

Class VI. "Skilled manual workers." Examples: Mining machine operators, tool and die makers.[6]

Class VIIa. "Semi- and unskilled manual workers (not in agriculture)." Example: Lathe and turning machine operators, textile sewing machine operators.[7]

Class VIIb. "Agricultural and other workers in primary production."

Our detailed coding of the unit-level 1980/1990 COC into each class is available in supplementary appendices for papers which are part of the same social class project. See, in particular, the supplementary appendix for the 2004 article by Morgan and McKerrow (posted on Morgan's website).

[6] Class VI includes occupations that are broadly classified as "Precision production, craft, and repair occupations" in the COC; however, there are numerous exceptions.

[7] Class VIIa includes occupations that are broadly classified as "Operators, fabricators, and laborers" in the COC; however, there are numerous exceptions.

References Cited in the Supplement

Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux*. Oxford: Clarendon Press.

Erikson, Robert, John H. Goldthorpe, and Lucianne Portocarero. 1979. "Intergenerational Class Mobility in Three Western European Societies: England, France, and Sweden." *British Journal of Sociology* 30:415-30.

Goldthorpe, John H. 1987. *Social Mobility and Class Structure in Britain. Second Edition.* Oxford: Clarendon Press.

Goldthorpe, John H. and Anthony Heath. 1992. *Revised Class Schema 1992*. Centre for Research into Elections and Social Trends, Working Paper No. 13.

Goldthorpe, John H. and Walter Müller. 1982. *Social Mobility and Class Formation in Industrial Nations: Proposal for a Comparative Research Project*. Oxford/Mannheim.

Heath, Anthony, Roger Jowell, and John Curtice. 1985. *How Britain Votes*. London: Pergamon.

Klein, Lawrence. 1962. *An introduction to econometrics*. Englewood Cliffs, N.J., Prentice-Hall.

Manza, Jeff and Clem Brooks. 1999. *Social Cleavages and Political Change: Voter Alignments and U.S. Party Coalitions*. Oxford: Oxford University Press.

Morgan, Stephen L. and Mark W. McKerrow. 2004. "Social Class, Rent Destruction, and the Earnings of Black and White Men, 1982-2000." In Kalleberg, Arne L., Stephen L. Morgan, John F. Myles, and Rachel A. Rosenfeld, eds., *Inequality: Structures, Dynamics and Mechanisms – Essays in Honor of Aage B. Sørensen*. *Research in Social Stratification and Mobility*, Volume 21. Amsterdam: Elsevier.

Table S1. Number of respondents and households in the 1986 and 1996 SIPP

	1986		1996	
	Respondents	Households	Respondents	Households
Full SIPP sample:				
Original sample size in the first wave	30,577	11,454	95,141	36,730
Sub-sample for College Enrollment Models:				
Respondents aged 17 to 21 in March	2,158	1,683	6,359	5,145
Respondents aged 17 to 21 in March and Still in the Sample in November	2,061	1,619	5,289	4,347
Respondents aged 17 to 21 in March and Still in the Sample in November and Unmarried in November	1,900	1,500	4,994	4,128

Table S2. Means of respondent's age and the percentage of respondents in college as of March and November by family structure variable

Family structure	Number (%) of respondents		Mean age		% college in March		% college in November	
	1986	1996	1986	1996	1986	1996	1986	1996
Unmarried								
Living with parents	1189 (57.68)	3074 (58.13)	18.75	18.66	28	27	37	37
Living with mother	474 (23.00)	1043 (19.71)	18.70	18.62	19	18	26	26
Living with father	10 (0.47)	179 (3.39)	18.70	18.69	0	16	20	26
Living with no parent	219 (10.61)	702 (13.27)	19.68	19.69	20	25	20	24
Married								
Living with parents	8 (0.40)	15 (0.28)	20.00	20.14	0	7	0	7
Living with mother	2 (0.10)	1 (0.01)	20.50	20.00	0	0	0	0
Living with father	.	1 (0.03)	.	21.00	.	0	.	0
Living with no parent	159 (7.74)	274 (5.19)	20.22	20.07	9	14	7	9
Total/Mean	2061 (100.00)	5289(100.00)	18.90	18.92	23	25	31	33

Notes: All frequencies and means are weighted using the monthly weight of November 1986 and 1996, respectively. Predicted frequencies are rounded to integers.

Table S3. Details of adjustments for topcodes of income and wealth, based on the full baseline SIPP sample (i.e., the first row from Table S1)

	Common top-code (in 1996 dollars)	Number of respondents at the common top-code		Percentage of respondents at the common top-code		Pareto-calculated top-code multiplier	
		1986	1996	1986	1996	1986	1996
Monthly Income	$10,000	287	1164	2.67	3.17	2.66	2.51
Net worth	$500,000	310	1452	2.89	3.95	1.98	1.50
Home equity	$275,027	121	612	1.13	1.67	2.56	2.10

Table S4. Summary statistics of different forms of resource variables (original variables, variables censored at the 95th percentile, and logged variables)

		Monthly Income				Total Net Worth	
		1986	1996			1986	1996
Original variable	Mean	4,282.07	4,751.11	Original variable	Mean	111,603.90	116,495.10
	Median	3,736.00	3,910.00		Median	65,327.28	52,828.00
	95th percentile	9,526.00	10,698.00		95th percentile	392,806.70	425,926.00
Original variable censored at 95th percentile	Mean	4,128.59	4,343.79	Original variable censored at 95th percentile	Mean	100,174.10	91,813.54
	Median	3,736.00	3,910.00		Median	65,327.28	52,828.00
	95th percentile	9,526.00	10,698.00		95th percentile	392,806.70	425,926.00
Natural logarithm of original variable	Mean	8.04	8.10	Natural logarithm of original variable	Mean	9.52	9.36
	Median	8.23	8.27		Median	11.09	10.87
	95th percentile	9.16	9.28		95th percentile	12.88	12.96

Table S5. Means and Standard Deviations of Variables Used to Model Educational Attainment

		1986		1996	
		Mean	S.D	Mean	S.D.
Dependent variable	Enrolled in college in November	.33		.34	
Sex	Female	.49		.48	
Race	Black	.13		.14	
	Hispanic	.09		.13	
	Asian	.02		.04	
	Mixed	.007		.023	
	Other	.004		.015	
Age	Age in March	18.20	1.63	18.20	1.77
Enrollment status as of March	Not enrolled	.30		.26	
	Freshmen or Sophomore	.08		.09	
	Junior	.18		.17	
	Senior	.18		.20	
	Enrolled in college/vocational school	.26		.27	
Family background, income, and wealth	Father's years of education	13.00	3.10	13.09	2.46
	Mother's years of education	12.56	2.52	12.73	2.28
	Total family monthly income[a]	4128.59	2380.12	4343.79	2653.98
	Zero family monthly income	.01		.01	
	Total net worth[b]	100,174.10	111,246.80	91,813.54	114,163.00
	Zero net worth	.04		.04	
	Negative net worth	.07		.10	
	Total home equity[c]	53,604.56	58,024.81	45,673.89	50,291.08
	Zero home equity	.27		.25	
	Negative home equity	.03		.02	
Class	Class I	.18		.19	
	Class II	.18		.19	

Class III	.18	.15
Class IV	.02	.01
Class V	.07	.05
Class VI	.13	.13
Class VIIa	.24	.26
Class VIIb	.01	.02

[a] Total family income is censored at the 97th percentile of the original variable. See table S4 for dollar values of the 95th percentile in each year.

[b] Total net worth is censored at the 95th percentile of the original variable. See table S4 for dollar values of the 95th percentile in each year.

[c] Total home equity censored at the 95th percentile. The 95th percentile in 1986 and 1996 is $198,019 and $170,000, respectively.

Table S6. Means of family monthly income, net worth, and home equity by family structure before imputation but after adjustments for topcoding

	Mean monthly income as of March (N)		Mean of net worth (N)		Mean of home equity (N)	
	1986	1996	1986	1996	1986	1996
Unmarried and living with both parents	5197.00 (1196)	5785.28 (2955)	148,603.76 (1195)	182,966.88 (2939)	70,463.44 (1195)	65,933.36 (2939)
with mother	2784.31 (454)	2502.93 (1125)	71,104.89 (454)	46,750.02 (1112)	38,198.06 (454)	28,792.18 (1112)
with father	3790.91 (10)	3891.28 (176)	47,587.95 (10)	88,214.28 (173)	27,879.38 (10)	35,689.82 (173)
alone	1241.65 (240)	1340.35 (738)	47,633.33 (239)	29,759.94 (730)	19,158.68 (239)	15,880.69 (730)
All unmarried (total N)	4130.13 (1900)	4408.45 (4994)	116,996.71 (1898)	129,899.60 (4954)	56,233.33 (1898)	50,150.76 (4954)

Notes: All means are weighted using the monthly weight of November. All topcoded values were adjusted as described in the text of the data appendix.

Table S7. Proportion Enrolled in College in November of 1986 and 1996 (and the total cell frequency), by Enrollment Status and Age in March of the Same Year

Enrollment in March	Year	Age 17	Age 18	Age 19	Age 20	Age 21
Not enrolled	1986	.000 (36)	.056(54)	.129(132)	.081(161)	.066(182)
	1996	.020(98)	.121(132)	.166(302)	.010(373)	.093(409)
Freshmen or Sophomore	1986	.000(121)	.000(21)	.178(6)	.000(0)	.000(2)
	1996	.000(417)	.013(80)	.000(12)	.000(9)	.000(2)
Junior	1986	.008(248)	.076(79)	.000(10)	.000(2)	.000(2)
	1996	.002(555)	.004(261)	.000(57)	.000(16)	.000(2)
Senior	1986	.594(32)	.594(229)	.440(164)	.177(17)	.000(1)
	1996	.096(125)	.561(588)	.455(209)	.314(51)	.154(13)
Enrolled in college/vocational school	1986	.000(0)	.810(26)	.811(164)	.687(163)	.740(146)
	1996	.538(13)	.760(49)	.801(372)	.816(456)	.761(393)