1

ANOTHER LOOK AT THE STRATIFICATION OF EDUCATIONAL TRANSITIONS: THE LOGISTIC RESPONSE MODEL WITH PARTIAL PROPORTIONALITY CONSTRAINTS

*Robert M. Hauser**
*Megan Andrew**

In this paper we reanalyze Robert D. Mare's highly influential work on educational transitions among American men born in the first half of the 20th century. Contrary to previous belief, Mare found that the effects of socioeconomic background variables decline regularly across educational transitions in conditional logistic regression analyses. We have reconfirmed Mare's findings and tested

An earlier version of this paper was prepared for the May 2005 meetings of the Research Committee on Social Stratification, International Sociological Association in Oslo, Norway. The research reported herein was supported by the William Vilas Estate Trust and by the Graduate School of the University of Wisconsin–Madison. Computation was carried out using facilities of the Center for Demography and Ecology at the University of Wisconsin–Madison, which are supported by Center Grants from the National Institute of Child Health and Human Development and from the National Institute on Aging. We thank Maarten Buis, Jeremy Freese, Carl Frederick, Harry Ganzeboom, Michael Hout, and John A. Logan. We thank Robert D. Mare especially for correcting an error in an earlier draft of this paper and for his elegant and useful discussion of our work. The opinions expressed here are those of the authors. Address correspondence to Robert M. Hauser or Megan Andrew, Center for Demography and Ecology, University of Wisconsin–Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 or e-mail to hauser@ssc.wisc.edu or mandrew@ssc.wisc.edu.

*University of Wisconsin-Madison

them by introducing a modified logistic response model that constrains selected social background effects to vary proportionally across educational transitions. We refer to our preferred model as the logistic response model with partial proportionality constraints (LRPPC). The model can easily be estimated in Stata or using other standard statistical software. Partial proportionality constraints may also prove useful in interpopulation comparisons based on other linear models.

Robert Mare's (1979, 1980, 1981) innovative analyses of American educational transitions in the 1973 Occupational Changes in a Generation (OCG) survey were among the most important and influential contributions to research on social stratification in the past three decades. Prior to the introduction of Mare's model of educational transitions in 1980, social stratification research typically employed linear probability models of school continuation and linear models of highest grade completed (e.g., Hauser and Featherman 1976). This research uniformly emphasized the stability of the stratification process in general and the effects of parental socioeconomic status on educational attainment. In his analyses, Mare applied a logistic response model to school continuation, restricting the base population at risk for each successive transition to those who had completed the prior educational transition. Contrary to prior supposition, Mare's estimates suggested the effects of some socioeconomic background variables declined across six successive transitions including completion of elementary school through entry into graduate school. Mare's (1980) original estimates are reproduced in Table 1.

Mare's studies of educational transitions have been both influential and controversial. His work spawned theories on the transition rates and odds ratios within educational systems, most notably the theories of maximally maintained inequality (Raftery and Hout 1993) and effectively maintained inequality (Lucas 2001). Mare's models were also the basis of a widely cited international comparative study of educational attainment (Shavit and Blossfeld 1993). However, the thesis that effects of social background decline across educational transitions has also been attacked by prominent labor economists (Cameron and Heckman 1998). They suggested, among other things, that Mare's logistic response model is only loosely motivated behaviorally and that the general decline of social background effects is a statistical artifact of the (logistic) parameterization of the model. However, Stolzenberg's (1994) probit

TABLE 1
Coefficients Representing Effects of Social Background Characteristics on School Continuation

Variable	Completes Elementary (0–8)		Attends High School Given Completes Elementary (8–9)		Completes High School Given Attends High School (9–12)		Attends College and Completes High School (12–13)		Completes College Given Attends College (13–16)		Attends Post-college Given Completes College (16–17)	
	b	b/se(b)	b	b/se(b)	b	b/se(b)	b	b/se(b)	b	b/se(b)	b	b/se(b)
Intercept	0.9886	4.22	1.2410	5.40	−0.1778	−1.48	−1.7440	−15.98	−0.6434	−6.33	−0.4669	−3.42
FASEI	0.0075	1.42	0.0041	0.87	0.0154	7.82	0.0145	10.49	0.0115	9.28	0.0070	4.24
SIBS	−0.1325	−5.67	−0.1444	−6.40	−0.1335	−11.39	−0.1067	−9.53	−0.0737	−6.30	−0.0138	−0.84
FAMINC	0.1067	5.36	0.0587	3.79	0.0655	8.57	0.0444	9.24	0.0097	2.68	−0.0110	−2.44
FED	0.1188	4.79	0.0939	3.96	0.0784	6.77	0.0420	4.47	0.0071	0.84	−0.0050	−0.45
MED	0.1677	7.16	0.1243	5.56	0.0815	7.11	0.0940	9.29	0.0361	3.86	0.0383	3.11
BROKEN	−0.3163	−1.71	−0.1256	−0.64	−0.2192	−2.30	−0.0078	−0.09	−0.1567	−1.84	−0.3713	−3.08
FARM	−0.6060	−4.54	−1.0560	−7.94	0.3013	3.88	0.0107	0.14	0.1138	1.41	0.1826	1.55
SOUTH	−0.5948	−4.70	0.4182	3.03	−0.0973	−1.45	0.0309	0.53	−0.0604	−1.08	−0.2736	−3.66
"R2"		0.270		0.178		0.120		0.091		0.026		0.008
$\chi^2(8)$		770.4		497.2		1226.4		1332.2		390.5		68.3
N		5,368		5,009		9,301		7,732		7,674		4,185
Subsample (%)		25		25		50		50		100		100

Source: Mare (1980:301).

Note: Dependent variables are the log odds of continuing from one schooling level to the next. Estimates are based on a 1973 sample of the U.S. white male civilian noninstitutional population born between1907 and 1951. Independent variables are FASEI: father's occupational Duncan socioeconomic index when respondent was 16; SIBS: number of siblings; FAMINC: annual income of family in thousands of constant (1967) dollars when respondent was 16; FED: father's grades of school completed; MED: mother's grades of school completed; BROKEN: absence of one or both parents from respondent's household most of the time to age 16; FARM: respondent lived on a farm at age 16; SOUTH: respondent born in the South census region. χ^2 tests null hypothesis that all coefficients are zero.

analysis reconfirmed Mare's finding that socioeconomic background has little or no influence on transitions from college to graduate training in an American cohort that completed college in the mid-1970s, and he explains the null finding by decay in the effects of SES on aspirations for further schooling. Sociologists have also criticized Mare's logistic response model of educational transitions. Applying multinomial logit models to longitudinal data from Sweden, Breen and Jonsson (2000) showed that class-origin effects on transition probabilities varied according to the particular choice made at a given transition point and that the probability of making a particular choice was path dependent.

Given the impact of Mare's work and the continuing controversy surrounding logistic response models of educational transitions, we return to the data originally analyzed by Mare in an appreciative effort to validate and extend his model. We introduce a modified version of his model that explicitly expresses and estimates changes in social origin effects across educational transitions. Rather than analyzing each educational transition separately as Mare did, we estimate a single model across all educational transitions. In this model, the relative effects of some (but not all) background variables are the same at each transition, and multiplicative scalars express proportional change in the effect of those variables across successive transitions.

1. MODELS OF EDUCATIONAL STRATIFICATION: A REVIEW

Aside from linear regression and linear probability models, researchers in educational stratification have employed a number of more appropriate models to explore the effects of social background on educational transitions, including logistic response, log-linear, and multinomial models. Each model has advantages and disadvantages in the study of educational transitions. Logistic response or continuation odds models employ conditional samples across successive educational transitions and allow for the estimation of robust coefficients that are invariant to marginal changes in educational attainment. In these models, continuation probabilities are asymptotically independent; a model may be estimated separately for each transition, or multiple transitions may be analyzed within a single model (Bishop, Fienberg, and Holland 1975; Fienberg 1977). Although logistic response models have been widely used in stratification research, these models require a large number of parameters, allowing effects of covariates to fluctuate freely across transitions whether or not they vary in the population.

Erikson and Goldthorpe's (1992:91–92) model of uniform differences in parameters of social mobility, Xie's (1992) log-multiplicative layer model for comparing mobility tables, and Hout, Brooks, and Manza's (1995:812) model of trends in class voting in the United States are popular variants of log linear models used to study educational transitions and each resembles the model introduced here by imposing proportionality constraints on the coefficients of a model across multiple populations .[1] These models provide population average estimates of changes in an outcome over time or place, but they do not easily allow for the introduction of individual-level covariates. Moreover, estimation problems may occur when log-linear models are extended to higher dimensions or in instances where the analyst wishes to consider more than three or four transitions, whether they be educational transitions or transitions among occupational groups.

Multinomial models have also been used to study educational transitions, though perhaps less often than logistic response and log-linear models. Multinomial models provide the opportunity to assess horizontal stratification in tandem with vertical stratification, allowing for a richer analysis in some instances (Breen and Jonsson 2000). Anderson's (1984) well-known stereotype regression model provides a flexible means to consider proportionality of the effect of a group of covariates on ordered outcomes but has been relatively ignored in educational stratification research. This model is an important analytic tool in many respects. For example, DiPrete's exemplary paper (1990) uses stereotype regression models to introduce individual-level covariates in social mobility analyses. Yet, multinomial models are still of limited value in the analysis of educational transitions. They specify multiple and possibly ordered categorical outcomes but do not model the conditional risk of transitions. With the exception of stereotype regression models, multinomial models also include a large number of coefficients. As in the case of unrestricted logistic response models, these may add unnecessary complexity and may make it difficult to interpret findings.

[1]We thank Michael Hout for bringing these similarities to our attention.

Given the various weaknesses of existing models of educational transitions and in an effort to extend statistical models of educational stratification, we propose a logistic response model with partial proportionality constraints. We believe that the model proposed here provides a parsimonious and powerful description of changes in the effects of socioeconomic background on educational attainment and, equally important, that it has wide application in studies of changes and differentials in social stratification.

2. LOGISTIC RESPONSE MODELS WITH PARTIAL PROPORTIONALITY CONSTRAINTS

We begin with a logistic response model in Mare's (1980:297) original notation:

$$\log_e\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{j0} + \sum_k \beta_{jk} X_{ijk}, \tag{1}$$

where p_{ij} is the probability that the ith person will complete the jth school transition, X_{ijk} is the value of the kth explanatory variable for the ith person who is at risk of the jth transition, and the β_{jk} are parameters to be estimated. That is, a logistic response model is estimated for persons at risk of completing each transition with no constraints on any parameters across transitions. Delineated above, the logistic response model has two important properties. First, the effects in equation 1 are invariant to the marginal distribution of schooling outcomes. That is, for $k > 0$, a given set of β_{jk} is consistent with any rate of completion of a transition. Second, the continuation probabilities are asymptotically independent of one another (Fienberg 1977). Thus, the model may be estimated separately for each transition, or multiple transitions may be analyzed within a single model.

Suppose that instead of analyzing the data for each transition separately the data are converted to person-transition records. Thus, a record appears for each transition for which each individual is eligible. In each record, there is a single outcome variable—say, y_{ij}, where $y_{ij} = 1$ if the transition is completed and $y_{ij} = 0$ if it is not completed. Since the transitions are ordered and each transition is conditional on completion of prior transitions, there is at most one record for which $y_{ij} = 0$ for each individual—namely, for the last transition for which that individual is eligible; at all prior transitions, $y_{ij} = 1$. In this setup, for example, one could estimate a model that is similar to equation 1, except there is only one set of regression parameters, which apply equally to each transition:

$$\log_e\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{j0} + \sum_k \beta_k X_{ijk}. \tag{2}$$

This model, like that of equation (1), may be estimated with any software that supports logistic regression analysis.

However, the hypothesis that socioeconomic background effects decline across transitions might suggest a different and more parsimonious model than equation (1). We refer to this as the logistic response model with proportionality constraints (LRPC):

$$\log_e\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{j0} + \lambda_j \sum_k \beta_k X_{ijk}, \tag{3}$$

The first term on the right-hand side of equation (3) says there may be a different intercept at each transition. The summation represents effects of the X_{ijk} that are invariant across all transitions; notice there is no j index on β_k. However, there is a multiplicative scalar for each transition, λ_j, which rescales the β_k at each transition subject to the normalizing constraint that $\lambda_1 = 1$. That is, the λ_j introduce proportional increases or decreases in the β_k across transitions; thus equation (3) implies proportional changes in main effects across transitions. The proportionally constrained covariates determine a composite variable, which can be interpreted in reference to a theoretical construct. In this instance, the main effects of the covariates can be interpreted as the weights of these covariates in that composite. Although equation (3) may appear to have more terms than equation (1), it is actually more parsimonious because there are only as many multiplicative terms as transitions. Equation (1) has as many interaction terms as the product of the number of explanatory variables and the number of transitions.

Conceptually, the model in equation (3) is similar to the well-known MIMIC (multiple indicator, multiple cause) model of Hauser, Goldberger, and Jöreskog (Hauser and Goldberger 1971; Hauser 1973; Jöreskog and Goldberger 1975; Hauser and Goldberger 1975). However, proportionality constraints appear within a single-equation model estimated simultaneously in two or more populations whereas the

MIMIC model imposes proportionality constraints across the coefficients of two or more equations estimated within a single population.

Several readers have suggested that the model of equation (3) is the same as the stereotype model. Recall that this model imposes proportionality constraints on coefficients across response categories in a multinomial logistic model; however, the stereotype model pertains to a single population and does not account for the conditional risk of transitions from one category to the next. If there were more than two potential outcomes at each transition, one could possibly combine features of the stereotype and LRPC models.

Our argument is that one ought to estimate a model like that in equation (3) before proposing more complex explanations of change in the effects of socioeconomic background variables across educational transitions. However, one cannot simply jump from the finding that the model of equation (3) fits a set of data to the conclusion that the effects of background variables are the same across transitions up to a coefficient of proportionality. Allison (1999) shows that, if the model of equation (3) fits a set of data, one cannot distinguish empirically between the hypothesis of uniform proportionality of effects across transitions and the hypothesis that group differences between parameters of binary regressions are artifacts of heterogeneity between groups in residual variation.[2]

It is also possible to mix the features of equations (2) and (3) to permit some variables to interact freely with a given transition while others follow a model of proportional change. We refer to this model as the logistic response model with partial proportionality constraints (LRPPC):

$$\log_e\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{j0} + \lambda_j \sum_{k=1}^{k'} \beta_k X_{ijk} + \sum_{k'+1}^{K} \beta_{jk} X_{ijk}. \qquad (4)$$

Equation (4) says that for some variables, X_k, where $k = 1, \ldots, k'$, there is proportional change in effects across transitions, while for other X_k, where $k = k' + 1, \ldots,$ K, the effects interact freely with transition level. For example, equation (4) could apply to Mare's analysis where effects

[2]Ballarino and Schadee (2005) advance this argument in the context of an analysis of educational transitions in Italy and several other nations.

of socioeconomic variables appear to decline across transitions while those of farm origin, one-parent family, and Southern birth vary in other ways.

Equation (4) may be generalized to cover multiple cohorts as well as multiple transitions. For example, equation (5) is one such generalization:

$$\log_e\left(\frac{p_{ijt}}{1-p_{ijt}}\right) = \beta_{jt0} + \lambda_j \sum_{k=1}^{k'} \beta_k X_{ijtk} + \sum_{k'+1}^{K} \beta_{jk} X_{ijtk} + \gamma_t \sum_{k=1}^{k''} \beta_k X_{ijtk} + \sum_{k''+1}^{K} \beta_{tk} X_{ijtk}. \quad (5)$$

Here, effects change proportionally across transition levels, j, for one set of variables; effects change proportionally across time for another (possibly overlapping) set of variables, indexed by t. The effects of the remaining sets of variables may interact freely with transition level j and with period t.

Estimation of equations (3) through (5) is not as simple as that of equations (1) and (2). The same linear expression, such as $\sum_k \beta_k X_{ijk}$, appears twice in the former equations—once with freely estimated coefficients and again as a linear composite in $\lambda_j \sum_k \beta_k X_{ijk}$. The problem is to estimate the models in a way that will yield the same estimates of the β_k in both expressions. One way to accomplish this is simply to iterate. First, estimate the β_k in a model with no interactions. Then, estimate the model again with an interaction in the composite estimated in the previous step, and continue until the fit and parameter values change very little from one iteration to the next (Allison 1999; MacLean 2005). We have used another method, writing the equations of the models and estimating them directly by maximum likelihood in Stata (see Appendix).[3] It may also be possible to estimate this model as if it were a MIMIC model (with uncorrelated outcome variables) using MPlus or other software for estimation of structural equation models.

In summary, this framework allows for the estimation of parsimonious, single equation models of educational transitions and addresses

[3]We thank Jeremy Freese for writing a Stata macro to estimate the model by maximum likelihood. For similar specifications in Stata and other statistical packages, see Allison (1999).

many of the shortcomings in extant models of educational transitions. The models are invariant to change in the marginal distribution, but they use fewer parameters and are characterized by ease in the introduction of individual-level covariates. The models can include uniform and/or partial proportionality constraints across model covariates and transitions of interest while freely estimating effects of other model covariates. Moreover, proportionality constraints in the models can be interpreted as indicative of a latent construct similar to that in the MIMIC model. Yet another possibility is to constrain the effects of a subset of variables to be invariant across populations. Finally, these models of educational transitions more easily allow for nuanced and powerful tests of the significance of estimated differences in the effects of model covariates across educational transitions and, thus, more informative inferences about the effects of social background on educational transitions. We next apply this framework in a replication of original work by Mare (1980, 1981) on educational transitions.

3. REPLICATING AND EXTENDING LOGISTIC RESPONSE MODELS USING THE 1973 OCG SURVEY

In replicating Mare's (1980) original analysis, we use the 1973 Occupational Changes in a Generation (OCG) survey data. The OCG survey was carried out as a supplement to the March 1973 Current Population Survey (CPS). It was carried out via mail in September 1973 after all of the households participating in the March demographic supplement had rotated out of the CPS. The response rate was 83 percent among target males. The present analysis includes 21,682 white men 21 to 65 years old in the civilian noninstitutional population who responded to all of the social background questions in the OCG supplement. We convert these individual records to 88,768 person-transition records.

Father's occupational status (FASEI) is the value of the Duncan Socioeconomic Index for Occupations (Duncan 1961); scale values were assigned using an adaptation of the original scale to codes for occupation, industry, and class of worker that were used in the 1970 Census.[4]

[4]Our code for the Duncan SEI differs from that used by Mare. Our code is more fine-grained than that which was available to Mare (1980, 1981), though the results are largely the same.

Number of siblings (SIBS) is based on questions about the number of older and younger siblings of each sex. The count of siblings was top-coded at 9. Family income (FAMINC) was represented by categorical responses to the question, “When you were about 16 years old, what was your family’s annual income?” Pretest respondents indicated that they answered in contemporary rather than price-adjusted dollars, so we adjusted the midpoints of responses from dollars in the year that the participant turned 16 to 1967 dollars using the Consumer Price Index. We assigned reports of “no income or loss” to the lowest reporting category ($1–499). After examining scatterplots of the relationship between various transformations of income and the probabilities of educational transitions, we top-coded the adjusted incomes at 2.5 standard deviations above the mean and re-expressed the variable in natural logs. Father’s education (FED) and mother’s education (MED) are expressed in years of regular schooling completed. Participants were coded as living in a broken family (BROKEN) if they responded “no” to the question, “Were you living with both your parents most of the time up to age 16?” A dummy variable for farm origin (FARM) was coded 1 if the participant said that he had not moved since age 16 (in the OCG survey) and currently lived on a farm (as reported in the CPS) or if he reported in the OCG survey that he had lived on a farm when he was 16 years old; otherwise, FARM was coded 0. State of birth was ascertained in the OCG survey, and a dummy variable (SOUTH) was created for men born in a Southern state as defined by the U.S. Bureau of the Census. With the exception of family income, all of the continuous background variables have approximately linear relationships in their original metrics with each of the educational transitions.

Educational attainment was ascertained in the March CPS. Respondents (who may or may not have been the OCG target male) reported both the highest grade in regular school that the OCG participant attended and whether or not he had completed that grade. With that protocol, it was possible to create plausible definitions of six key educational transitions in populations at risk of those transitions: (1) completing elementary school (grade 8); (2) attending high school (grade 9) among those who completed elementary school; (3) completing high school (grade 12) among those who attended high school; (4) attending college among those who completed high school; (5) completing college among those who attended college;[5] and (6) attending some form of postgraduate education among those who graduated from college. Unfortunately, it would not be possible to carry out a comparable analysis of educational transitions with Census data after 1990, when the Bureau chose a one-question item on educational attainment (Hauser 1997).

Table 2 shows descriptive statistics for continuous variables used in the present analysis, and Table 3 displays descriptive statistics for the three discrete background variables. Without exception, each successive transition yields a more selective and successful set of students. At each transition, successful students have more highly educated parents with higher occupational status and income and come from smaller families than students who did not complete that transition. They are less likely to have been raised by a single parent, less likely to live or have lived on a farm, and less likely to have been born in a Southern state.[6]

3.1. *Replicating Traditional Logistic Response Models*

Table 4 reports our unconstrained estimates of the effects of social background characteristics on school continuation in the 1973 OCG data. These estimates are based on equation (1) and replicate Mare's (1980) original logistic response model. The coefficients differ somewhat from those estimated by Mare (1980) for several reasons. Either we did not use the same version of the 1973 OCG data file or our sample definition was different from Mare's in some way that we have been unable to determine. Unlike Mare (1980), we also use a more refined scheme for scaling father's occupational status and a logarithmic transformation of family income. Moreover, we used all of the available cases at every transition and defined the base population for college graduation to include those who attended college but did not complete at least one year of college.

[5] Based on comparisons of our sample counts with those reported by Mare (1980:301), we strongly suspect that he defined the base population for completion of college to exclude persons who entered college but did not complete at least one year of college work. We defined the base for college completion to include that group.

[6] We find substantially fewer men than Mare found with farm background or Southern birth.

TABLE 2
Means and Standard Deviations of Social Background Characteristics at Selected Levels of Schooling: 1973 Occupational Changes in a Generation Survey

Level of Schooling	Father's Occupation (FASEI)	Family Income (FAMINC)	Father's Schooling (FED)	Mother's Schooling (MED)	Number of Siblings (SIBS)	N
School entry	31.4	1.773	8.6	9.1	3.7	21682
	22.9	0.900	4.1	3.8	2.6	
Elementary school completion	32.6	1.857	9.0	9.5	3.5	20058
	23.1	0.832	3.9	3.6	2.5	
High school attendance	33.7	1.906	9.2	9.7	3.4	18725
	23.3	0.800	3.9	3.5	2.5	
High school graduation	36.0	1.987	9.7	10.1	3.1	15602
	23.8	0.764	3.8	3.4	2.4	
College attendance	42.8	2.165	10.7	11.0	2.6	8462
	25.1	0.710	3.9	3.3	2.1	
College graduation	46.6	2.226	11.1	11.3	2.4	4239
	25.3	0.698	3.9	3.3	2.0	

TABLE 3
Percentages of Men with Selected Background Characteristics by Levels of Schooling: 1973 Occupational Changes in a Generation Survey

	Broken Family (BROKEN)	Farm Background (FARM)	Southern Birth (SOUTH)	% Continuing	% of All Men
School entry	10.4	16.6	27.0	92.5	100.0
Elementary completion	9.9	15.4	25.1	93.4	92.5
HS attendance	9.8	13.9	24.9	83.3	86.4
HS graduation	9.2	13.3	23.8	54.2	72.0
College attendance	8.2	9.9	22.4	50.1	39.0
College graduation	7.4	9.2	22.1	49.7	19.6

Despite these differences in variable definitions and case selection, the estimates in Table 4 follow the main patterns of Mare's original estimates (see Table 1). Social background explains less of the variation at each higher educational transition, and the effects of the socioeconomic background variables as defined by Mare (FASEI, FAMINC, FED, MED, and SIBS) typically decline from lower to higher educational transitions. Similarly, effects of the other background variables (BROKEN, FARM, and SOUTH) do not show the same pattern of decline across transitions.

3.2. *Extending Traditional Logistic Response Models: The Logistic Response Model with Partial Proportionality Constraints*

Table 5 describes the fit of single-equation models simultaneously estimated for all six educational transitions. Model 1 is a null baseline in which no parameters are fitted except the grand mean. The likelihood ratio test statistic under this model defines the denominator of the pseudo-R^2 statistics and can be used to measure improvements in the fit of more complex models. Model 2 fits an intercept for each transition, and it yields a substantial improvement in fit. Model 3 adds invariant effects of social background variables to Model 2. This follows the model in equation (2). Overall rates of transition vary across levels of schooling, but the effects of social background do not vary across transitions. This modification also yields a substantial improvement in fit.

TABLE 4
Coefficients Representing Effects of Social Background Characteristics on School Continuation: 1973 Occupational Changes in a Generation Survey

Variable	Completes Elementary (0–8)		Attends High School Given Completes Elementary (8–9)		Completes High School Given Attends High School (9–12)		Attends College Given Completes High School (12–13)		Completes College Given Attends College (13–16)		Attends Post-college Given Completes College (16–17)	
	b	b/se(b)	b	b/se(b)	b	b/se(b)	b	b/se(b)	b	b/se(b)	b	b/se(b)
Intercept	0.4655	4.45	0.8376	7.32	−0.1509	−1.78	−1.8544	−22.45	−0.4352	−4.04	−0.1628	−1.05
FASEI	0.1731	6.87	0.2308	9.40	0.1575	11.98	0.1650	17.56	0.1162	10.22	0.0678	4.22
SIBS	−0.1234	−10.83	−0.1397	−12.05	−0.1260	−15.21	−0.1042	−13.24	−0.0889	−8.02	−0.0048	−0.29
FAMINC	0.5690	18.29	0.4013	11.87	0.2953	10.92	0.2976	11.04	0.0213	0.59	−0.1552	−3.01
FED	0.1216	10.13	0.0627	5.22	0.0566	7.10	0.0465	7.12	−0.0114	−1.40	−0.0075	−0.67
MED	0.1496	13.15	0.1083	9.34	0.0911	11.46	0.0748	10.66	0.0225	2.52	0.0327	2.67
BROKEN	−0.3121	−3.66	−0.0554	−0.57	−0.2228	−3.38	−0.1001	−1.64	−0.1473	−1.79	−0.5010	−4.08
FARM	−0.1413	−2.08	−0.6979	−10.37	0.2811	4.80	−0.0278	−0.52	0.2306	2.92	−0.0152	−0.13
SOUTH	−0.6419	−10.57	0.3268	4.71	−0.0771	−1.63	0.0131	0.31	−0.0286	−0.54	−0.2649	−3.52
"R2"		0.313		0.191		0.134		0.120		0.025		0.011
$\chi^2(8)$		3611.6		1868.4		2263.9		2582.1		293.6		64.8
N		21,682		20,058		18,725		15,602		8,462		4,239

Note: Dependent variables are the log odds of continuing from one schooling level to the next. Estimates are based on a 1973 sample of the U.S. white male civilian noninstitutional population born between 1907 and 1951. Independent variables are FASEI: father's occupational Duncan socioeconomic index when respondent was 16; SIBS: number of siblings; FAMINC: natural log of truncated annual income of family in thousands of constant (1967) dollars when respondent was 16; FED: father's grades of school completed; MED: mother's grades of school completed; BROKEN: absence of one or both parents from respondent's household most of the time to age 16; FARM: respondent lived on a farm at age 16; SOUTH: respondent born in the South census region. χ^2 tests null hypothesis that all coefficients are zero.

TABLE 5

Fit of Selected Models of Educational Transitions: 1973 Occupational Changes in a Generation Survey

Model	Description	Log-Likelihood	DF for Model	Model Chi-square	Contrast	Contrast Chi-square	Contrast BIC	Pseudo R-squared
1	Fit the grand mean	−46830.8	0	—		—		0
2	An intercept for each transition	−38674.3	5	16313.0	2 vs. 1	16313.0	16256.0	0.17
3	An intercept for each transition and constant social background effects	−34333.3	13	24995.0	3 vs. 2	8682.0	8590.8	0.27
4	An intercept for each transition and proportional social background effects	−33529.7	19	26602.2	4 vs. 3	1607.3	1538.9	0.28
5	An intercept for each transition, constant effects of socioeconomic variables, interactions of BROKEN, FARM, and SOUTH with transition	−34112.0	28	25437.6	5 vs. 3	442.6	271.7	0.27
6	An intercept for each transition, proportional effects of socioeconomic variables, interactions of BROKEN, FARM, and SOUTH with transition	−33399.7	34	26862.1	6 vs. 5	1424.6	1356.2	0.29
7	Saturated model: Intercepts for each transition and interactions of all social background variables with transition	−33332.2	53	26997.2	7 vs. 6	135.1	−81.4	0.29

Model 4 is based on equation (3). Rather than specifying constant effects of the eight social background variables, it says that all of the background effects vary in the same proportion across each transition. With six more degrees of freedom, the parameters of proportional change yield an improved model fit (change in the chi-square statistic of 1607.3). Evidently, substantial variation in effects of background across educational transitions is captured by the model of proportional change. However, Model 4 does not fit the data well. Model 7 fits all of the interactions between social background variables and transitions, and its fit is significantly better than that of Model 4 (chi-square of 395 with 34 degrees of freedom). Moreover, as noted above, the improvement of fit in Model 4 relative to Model 3 does not tell us whether the background effects actually vary proportionally across transitions or whether there are corresponding differences in residual variation across transitions.

Model 5 specifies constant effects of the socioeconomic background variables (FASEI, FAMINC, MED, and FED) and number of siblings (SIBS), but it permits effects of the other three background variables (BROKEN, FARM, and SOUTH) to interact freely with transition level. Note that Model 5 is a special case of equation 2—because some effects do not vary across educational transitions—and that Model 4 is not nested within Model 5. Comparing Model 5 to Model 3, we find a significant improvement in fit, but it is only about a quarter of the improvement from Model 3 to Model 4 and uses 15 degrees of freedom.

Model 6 is based on the specification of Model 5, but it adds proportional change across transitions in the effects of FASEI, FAMINC, MED, FED, and SIBS. It is an example of the model specified in equation (4). The improvement of fit in Model 6 relative to Model 5 is almost as large as in that of Model 4 relative to Model 3. That is, even after the introduction of freely estimated interaction effects between each transition level and BROKEN, FARM, and SOUTH, fit is improved substantially by the specification of proportional change in effects of socioeconomic background across educational transitions. However, because proportionality constraints apply only to a subset of the background effects in Model 6, we know that variation in coefficients across transitions is not entirely due to residual variation in transitions.

Finally, Model 7 is an example of the specification in equation (2) in which effects of all background variables are permitted to interact with educational transition levels. Although it uses 19 more parameters than Model 6, the improvement in fit is negligible by comparison with the other contrasts in Table 5. Thus, we prefer Model 6 to Model 7 and the other models listed in Table 5. This decision is confirmed by the BIC statistics for contrasts between models, which are also reported in Table 5. Excepting the contrast between Models 6 and 7, there is a substantial improvement in BIC between each successive model.

Table 6 shows the estimated parameters of Model 6. Recall that this model includes additive effects of all of the social background variables, freely estimated interactions of broken family, farm background, and Southern birth with transition level, and multiplicative effects of transition level with a linear composite of the socioeconomic variables and number of siblings.

Table 6 identifies four groups of parameters: (a) freely estimated additive effects, (b) freely estimated interaction effects, (c) multiplicative effects, and (d) additive effects and multiplicative composite. However, the first two groups are not distinct for purposes of estimation. In this instance, group (a) includes the main effects of each transition level and the main effects of broken family, farm background, and Southern birth while group (b) comprises the interaction effects of broken family, farm background, and Southern birth with each educational transition. There are no interaction effects with the first transition because the main effects of broken family, farm background, and Southern birth are defined to reference that transition. Group (c) specifies the multiplicative effects of the second through sixth transitions relative to the effects of socioeconomic background in the first transition. Finally, group (d) includes the main effects of the socioeconomic variables and number of siblings, which are also the weights of those variables in the composite that interact with educational transition level.

The estimates of direct interest in Table 6 are the main effects of the socioeconomic variables (d) and the multiplicative effects of the transition levels (c). As one should expect, the main effects of father's occupational status, family income, mother's education, and father's education are positive and highly significant, while that of number of siblings is negative and highly significant. The multiplicative effects of transition level are also highly significant, and they are increasingly negative at higher level transitions. These coefficients may appear anomalous at first sight, but increasingly negative effects are exactly what one should expect. That is, at each higher transition level, there is a larger proportional decrement in the main effects of the socioeconomic variables.

TABLE 6
Estimated Parameters of LRPPC Model (M6): 1973 Occupational Changes in a Generation Survey

Variable	Coefficient	Standard Error	t-Statistic
a. Freely Estimated Additive Effects			
Completes elementary (0-8, TRANS1)	0.7815	0.0777	10.06
Attends high school if completes elementary (8-9, TRANS2)	0.7738	0.0655	11.82
Completes high school if attends high school (9-12, TRANS3)	−0.3125	0.0670	−4.66
Attends college if completes high school (12-13, TRANS4)	−1.9488	0.0613	−31.79
Completes college if attends college (13-16, TRANS5)	−0.9565	0.0734	−13.03
Attends post-college if completes college (16-17, TRANS6)	−0.3145	0.1085	−2.90
Nonintact family (BROKEN)	−0.3453	0.0833	−4.15
Farm background (FARM)	−0.1010	0.0672	−1.50
Southern birth (SOUTH)	−0.6276	0.0612	−10.25
b. Freely Estimated Interaction Effects			
TRANS2 × BROKEN	0.2913	0.1265	2.30
TRANS2 × FARM	−0.6159	0.0929	−6.63
TRANS2 × SOUTH	0.9564	0.0914	10.47
TRANS3 × BROKEN	0.1390	0.1053	1.32
TRANS3 × FARM	0.3894	0.0885	4.40
TRANS3 × SOUTH	0.5488	0.0776	7.08
TRANS4 × BROKEN	0.2449	0.1024	2.39
TRANS4 × FARM	0.0506	0.0849	0.60
TRANS4 × SOUTH	0.6431	0.0740	8.69
TRANS5 × BROKEN	0.2326	0.1160	2.01
TRANS5 × FARM	0.2259	0.1018	2.22
TRANS5 × SOUTH	0.6058	0.0811	7.47
TRANS6 × BROKEN	−0.0866	0.1473	−0.59
TRANS6 × FARM	0.0767	0.1303	0.59
TRANS6 × SOUTH	0.3855	0.0967	3.99
c. Multiplicative Effects			
Attends high school if completes elementary (8-9, TRANS2)	−0.2257	0.0238	−9.50
Completes high school if attends high school (9-12, TRANS3)	−0.3704	0.0221	−16.76
Attends college if completes high school (12-13, TRANS4)	−0.4312	0.0180	−23.95

(Continued)

TABLE 6 (Continued)

Variable	Coefficient	Standard Error	t-Statistic
Completes college if attends college (13-16, TRANS5)	−0.7804	0.0157	−49.68
Attends post-college if completes college (16-17, TRANS6)	−0.9159	0.0214	−42.73
d. Additive Effects and Multiplicative Composite			
FASEI	0.2662	0.0123	21.55
SIBS	−0.1698	0.0072	−23.71
FAMINC	0.5233	0.0209	25.07
FED	0.0911	0.0066	13.90
MED	0.1435	0.0064	22.25

For example, the estimate of −.2257 for the transition to high school conditional upon completion of elementary school says that the total effect of each socioeconomic background variable is 20.3 percent smaller at the transition from elementary school to high school than it is at the transition from school entry to the completion of elementary school. Note that the decrements in the multiplicative effects are not equal across transitions. The largest decrement is that between college entry and college completion (−0.7804 – (−0.4312) = −0.3492), and the second largest is between elementary school completion and high school entry (−0.2257).

To illustrate the implications of these estimates for the total effects of each of the social background variables, we insert the parameter estimates into the linear model (equation 4) and rearrange terms. The first panel of Table 7 displays the total effects in Model 6 that are implied by the parameter estimates. As expected, the effects of father's occupational status, number of siblings, family income, and parents' education decline regularly across transitions, while those of broken family, farm background, and Southern birth do not. In the second panel, for purposes of comparison, we show the corresponding unconstrained estimates from Model 7.[7] The latter are necessarily identical to those in Table 4 because Model 7 fits all of the background by level

[7]We are concerned here mainly with the deviations in the effects of social background variables, not with the intercepts.

TABLE 7
Comparison of Estimates from LRPPC Model (M6) with Estimates from Saturated Model (7)

Variable	Completes Elementary (0–8)	Attends High School Given Completes Elementary (8–9)	Completes High School Given Attends High School (9–12)	Attends College Given Completes High School (12–13)	Completes College Given Attends College (13–16)	Attends Post-college Given Completes College (16–17)
Model 6: Intercept for each transition, proportional effects of socioeconomic variables, and interactions of BROKEN, FARM, and SOUTH with each transition						
FASEI	0.266	0.206	0.168	0.151	0.058	0.022
SIBS	−0.170	−0.132	−0.107	−0.097	−0.037	−0.014
FAMINC	0.523	0.405	0.329	0.298	0.115	0.044
FED	0.091	0.071	0.057	0.052	0.020	0.008
MED	0.143	0.111	0.090	0.082	0.032	0.012
BROKEN	−0.345	−0.054	−0.206	−0.100	-0.113	−0.432
FARM	−0.101	−0.717	0.288	−0.050	0.125	−0.024
SOUTH	−0.628	0.329	−0.079	0.016	−0.022	−0.242
Intercept	0.781	0.774	−0.313	−1.949	−0.956	−0.314
Model 7: Saturated model with intercepts for each transition and interactions of all social background variables with each transition						
FASEI	0.173	0.231	0.158	0.165	0.116	0.068
SIBS	−0.123	−0.140	−0.126	−0.104	−0.089	−0.005
FAMINC	0.569	0.401	0.295	0.298	0.021	−0.155
FED	0.122	0.063	0.057	0.047	−0.011	−0.007
MED	0.150	0.108	0.091	0.075	0.023	0.033
BROKEN	−0.312	−0.055	−0.223	−0.100	−0.147	−0.501
FARM	−0.141	−0.698	0.281	−0.028	0.231	−0.015
SOUTH	−0.642	0.327	−0.077	0.013	−0.029	−0.265
Intercept	0.465	0.838	−0.151	−1.854	−0.435	−0.163

(Continued)

TABLE 7 (Continued)

Variable	Completes Elementary (0–8)	Attends High School Given Completes Elementary (8–9)	Completes High School Given Attends High School (9–12)	Attends College Given Completes High School (12–13)	Completes College Given Attends College (13–16)	Attends Post-college Given Completes College (16–17)
Deviations (Model 7–Model 6)						
FASEI	−0.093	0.025	−0.010	0.014	0.058	0.045
SIBS	0.046	−0.008	−0.019	−0.008	−0.052	0.010
FAMINC	0.046	−0.004	−0.034	0.000	−0.094	−0.199
FED	0.030	−0.008	−0.001	−0.005	−0.031	−0.015
MED	0.006	−0.003	0.001	−0.007	−0.009	0.021
BROKEN	0.033	−0.001	−0.016	0.000	−0.035	−0.069
FARM	−0.040	0.019	−0.007	0.023	0.106	0.009
SOUTH	−0.014	−0.002	0.002	−0.002	−0.007	−0.023
Intercept	−0.316	0.064	0.162	0.094	0.521	0.152

interactions. The third panel shows the deviations of the estimates in Model 6 from those in Model 7.

As we should expect from the contrast reported in Table 5 between Models 6 and 7, the deviations in the third panel are small in most cases. The largest single deviation (−0.199) pertains to the anomalously large and negative effect of family income on the transition from college to graduate school. A second large deviation is an underprediction of the salutary effect of farm background on the completion of college. One notable pattern in the deviations is that the model underestimates the effect of father's occupational status on higher-level transitions, and it overestimates the effects of the other background variables in the linear composite at those levels.[8]

Conceivably, one might argue that father's occupational status should be removed from the linear composite in the other social background variables and interacted freely with transition levels. However, such a decision—along with the inclusion of number of siblings in the linear composite—would be inconsistent with the more general finding that socioeconomic effects vary proportionally across transition levels. We should then be left with a less parsimonious and less appealing claim that some socioeconomic effects vary proportionally while others do not. In addition, because there is only one discrepant effect, we think that the evidence of nonproportionality is weak, and we have not modified the model.

4. DISCUSSION

In this analysis, we have shown that specification of interaction effects with a linear composite is a parsimonious way to represent and assess the way in which social background effects vary across educational transitions. Compared to Robert Mare's analysis of educational transitions among American men, we have shown that the more parsimonious LRPPC model reproduces most of the empirical features of his estimates. While estimation of the LRPPC model is not as straightforward as an ordinary regression analysis, it can be done easily and quickly with standard statistical software.

[8]This is consistent with Mare's (1980:302–3) observations about the pattern of coefficients in the unrestricted estimates.

We can think of several other ways in which models with interaction effects with a linear composite could be useful. We have already noted earlier applications of a similar idea to comparative analysis of mobility tables (Erikson and Goldthorpe 1992; Xie 1992) and to trends in class voting (Hout et al. 1995). One obvious extension is to analyses, like those in Mare (1979) and in Shavit and Blossfeld (1993), where multiple educational transitions are observed in several cohorts.[9] Such models, we believe, would also be useful to structure and discipline international comparative analyses, such as pooled analyses of the data used by Shavit and Blossfeld (1993) or other cross-national studies that have been designed specifically to create comparable data. We hope that similar models and methods may prove useful across a wide range of research questions in the social and behavioral sciences.

APPENDIX: STATA CODE FOR PREFERRED MODEL OF EDUCATIONAL TRANSITIONS

```
**Model 6: Logistic Response Model with Partial Proportionality Constraints **
capture program drop pplogit
        program define pplogit
                tempname theta
        version 6
        args lnf theta1 theta2 theta3
        gen `theta' = `theta1' + `theta3' + (`theta2'*`theta3')
        quietly replace `lnf' = ln(exp(`theta')/(1+exp(`theta'))) if
                $ML_y1==1
        quietly replace `lnf' = ln(1/(1+exp(`theta'))) if $ML_y1==0
        end
        ml model lf pplogit (outcome = trans1 trans2 trans3 trans4 trans5
trans6 broken farm16 south trans2Xbroken trans2Xfarm16 trans2Xsouth
trans3Xbroken trans3Xfarm16 trans3Xsouth trans4Xbroken trans4Xfarm16
trans4Xsouth trans5Xbroken trans5Xfarm16 trans5Xsouth trans6Xbroken
trans6Xfarm16 trans6Xsouth, nocons) (trans2 trans3 trans4 trans5 trans6,
nocons) (dunc sibsttl9 ln_inc_trunc edhifaom edhimoom, nocons)
**The following set of starting values are not essential**
**but estimation is much faster when starting values are assigned **
        ml init eq1:trans1 =.4041526 eq1:trans2 =.7751678 eq1:trans3 =-
.3116492 eq1:trans4 =-1.948527 eq1:trans5 =-.9562174 eq1:trans6 =-.3119838
eq1:trans2Xbroken =-.0519577 eq1:trans2Xfarm16 =-.7136086 eq1:trans2Xsouth
=.3302017 eq1:trans3Xbroken = -.2044395 eq1:trans3Xfarm16 = .2908632
eq1:trans3Xsouth =-.0779455 eq1:trans4Xbroken =-.0985229 eq1:trans4Xfarm16 =-
.0484765 eq1:trans4Xsouth =.0161895 eq1:trans5Xbroken =-.1119574
eq1:trans5Xfarm16 =.1257399 eq1:trans5Xsouth = -.0215616 eq1:trans6Xbroken =-
.4318161 eq1:trans6Xfarm16 =-.0246239 eq1:trans6Xsouth =-.2418691 eq2:trans2 =-
.2524217 eq2:trans3 =-.3919983 eq2:trans4 =-.4505379 eq2:trans5 =-.7878735
eq2:trans6 =-.9192267 eq3:dunc =.2752204 eq3:sibsttl9 =-.1762127
eq3:ln_inc_trunc =.554373 eq3:edhifaom =.0953649 eq3:edhimoom =.1451568
        ml maximize
```

[9]Equation (5) of this paper provides a template for such analyses.

REFERENCES

Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods Research* 28(2):186–208.

Anderson, J. A. 1984. "Regression and Ordered Categorical Variables." *Journal of the Royal Statistical Society*. Series B (Methodological) 46(1):1–30.

Ballarino, Gabriele, and Hans Schadee. 2005. "Really Persisting Inequalities?" Research Committee on Social Stratification, International Sociological Association. University of California, Los Angeles.

Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Breen, Richard, and Jan O. Jonsson. 2000. "Analyzing Educational Careers: A Multinomial Transition Model." *American Sociological Review* 65(5):754–72.

Cameron, Stephen V., and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy* 106(2):262–333.

DiPrete, Thomas A. 1990. "Adding Covariates to Loglinear Models for the Study of Social Mobility." *American Sociological Review* 55(5):757–73.

Duncan, Otis D. 1961. "A Socioeconomic Index for All Occupations." Pp. 109–38 in *Occupations and Social Status*, edited by Albert J. Reiss, Jr. New York: Free Press.

Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford, England: Clarendon.

Fienberg, Stephen E. 1977. *The Analysis of Cross-Classified Categorical Data: Theory and Practice*. Cambridge, MA: MIT Press.

Hauser, Robert M. 1973. "Disaggregating a Social-Psychological Model of Educational Attainment." Pp. 255–84 in *Structural Equation Models in the Social Sciences*, edited by A. S. Goldberger and O. D. Duncan. New York: Seminar Press.

———. 1997. "Indicators of High School Completion and Dropout." Pp. 152–84 in *Indicators of Children's Well-Being*, edited by Robert M. Hauser, Brett V. Brown, and William R. Prosser. New York: Russell Sage Foundation.

Hauser, Robert M., and David L. Featherman. 1976. "Equality of Schooling: Trends and Prospects." *Sociology of Education* 49(2):99–120.

Hauser, Robert M., and Arthur S. Goldberger. 1971. "The Treatment of Unobservable Variables in Path Analysis." Pp. 81–117 in *Sociological Methodology*, vol. 1, edited by Herbert L. Costner. San Francisco: Jossey-Bass.

———. 1975. "Correction of 'The Treatment of Unobservable Variables in Path Analysis'." Pp. 212–13 in *Sociological Methodology*, vol. 5, edited by David R. Heise. San Francisco: Jossey-Bass.

Hout, Michael, Clem Brooks, and Jeff Manza. 1995. "The Democratic Class Struggle in the United States, 1948–1992." *American Sociological Review* 60(6):805–28.

Jöreskog, Karl G., and Arthur S. Goldberger. 1975. "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable." *Journal of the American Statistical Association* 70:631–39.

Lucas, Samuel R. 2001. “Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects.” *American Journal of Sociology* 106(6):1642–90.

Maclean, Alair. 2005. “Lessons from the Cold War: Military Service and College Education.” *Sociology of Education* 78:250–66.

Mare, Robert D. 1979. “Social Background Composition and Educational Growth.” *Demography* 16(1):55–71.

———. 1980. “Social Background and School Continuation Decisions.” *Journal of the American Statistical Association* 75:293–305.

———. 1981. “Change and Stability in Educational Stratification.” *American Sociological Review* 46(1):72–87.

Raftery, Adrian E., and Michael Hout. 1993. “Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education, 1921–75.” *Sociology of Education* 66(1):41–62.

Shavit, Yossi, and Hans-Peter Blossfeld. 1993. *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview.

Stolzenberg, Ross M. 1994. “Educational Continuation by College Graduates.” *American Journal of Sociology* 99(4):1042–77.

Xie, Yu. 1992. “The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables.” *American Sociological Review* 57(3):380–95.