Adding Covariates to Loglinear Models for the Study of Social Mobility
Author(s): Thomas A. DiPrete
Source: *American Sociological Review*, Vol. 55, No. 5 (Oct., 1990), pp. 757-773
Published by: American Sociological Association
Stable URL: https://www.jstor.org/stable/2095870

ADDING COVARIATES TO LOGLINEAR MODELS FOR THE STUDY OF SOCIAL MOBILITY*

THOMAS A. DIPRETE
Duke University

Two strategies, linear regression and loglinear models, have enabled sociologists to make great progress in the study of social mobility and stratification, but each has deficiencies. Linear regression models are insensitive to the multidimensional character of stratification, while loglinear models do not easily incorporate independent variables. I propose a class of constrained multinomial logit models for the study of social mobility that bridges the gap between these two approaches. Parsimony in specifying intercepts is achieved through standard methods for parameterizing interaction terms in loglinear and related models of social mobility. Parsimony in specifying the effects of covariates is achieved by partitioning covariates into groups within which effects are constrained to be proportional. The resulting specification consists of three types of parameters: (1) a reduced set of intercepts; (2) coefficients that convert the effect of each variable in a group into what may be thought of as a single group-specific metric; and (3) a set of scores for each group that specifies the impact of the group's covariates on outcomes. Examples are provided using data from the 1983 and 1987 Current Population Surveys.

Two approaches to the study of social mobility are prominent in the sociological literature. The first uses linear regression to model outcomes on a single quantitative stratification dimension as a function of covariates, which may measure individual-level resources and liabilities or contextual factors. Status attainment models and earnings functions are examples of this strategy. The second approach uses loglinear and related (e.g., association) models for categorical data to study mobility across stratification categories, such as classes or occupational groups.

While these strategies have allowed sociologists to make enormous progress in the study of stratification and mobility, each has certain deficiencies as well. Linear regression models are insensitive to the multidimensional character of the stratification system (Hodge 1981; Hout 1988, p. 1362 ff.). Loglinear and related models overcome this deficiency, but, unlike linear regression models, they do not allow easy specification of the individual and structural determinants of mobility apart from the information contained in the marginal distributions of origin and destination positions. Consequently, a gap exists in the methodological tools available for mobility research. Models of social mobility are needed that avoid the rigidly hierarchical assumptions of status attainment or earnings functions while retaining the ability of these models to easily incorporate individual- and structural-level covariates.

I propose a class of constrained multinomial logit models for multidimensional, partially-ordered outcomes that can incorporate individual-level and structural effects in addition to the vacancy constraints arising from the marginals of loglinear and related models. These constrained multinomial logit models have several advantages over their unconstrained counterpart. First, they are more parsimonious. Second, maximum likelihood estimates of their parameters are usually attainable, even when the sample size and pattern of associations would cause unconstrained estimation to fail. Third, these models can provide a closer link between parameter values and generalizations about the structure of the mobility process.

* Direct correspondence to Thomas A. DiPrete, Department of Sociology, Duke University, Durham, NC 27706. I would like to thank Margaret Krecker, Michael Hout, Adrian Raftery, and three anonymous reviewers for helpful comments on an earlier version of this paper. I would also like to thank Margaret Krecker for her assistance in preparing the data. This research was supported in part by research funds provided by Duke University, and in part by National Science Foundation grant No. SES-90-12619.

ALTERNATIVE MOBILITY MODELS

Two important tasks of mobility research are describing and explaining the probability of moving to a particular destination category from a particular origin category. With a loglinear model, the log of these probabilities can be expressed as a linear function of origin effects, destination effects, and effects for the interaction between origin and destination:

$$\log P_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \qquad (1)$$

Parsimony in these models is typically achieved by imposing constraints on the interaction terms. For example, Featherman and Hauser (1978) constructed an informative model by constraining interaction effects to be equal within groups of cells that are mutually exclusive and exhaustive of the entire mobility table (see also Goodman 1972). Other scholars have shown how to express interaction terms through row or column scores, which may be imposed *a priori* or estimated from data (Haberman 1974; Goodman 1979, 1987).

Loglinear and related models can provide powerful descriptions of the mobility structure, but in their usual form they do not incorporate covariates that might explain this structure. In recent years, however, mobility researchers have elaborated these models to include aggregate-level covariates. For example, in an application of the Haberman model to the study of intergenerational mobility, Hout (1984) used cell averages of substantively interesting variables to specify the $u_{12(ij)}$ of equation 1 as

$$u_{12(ij)} = b_1 S_i S_j + b_2 A_i A_j + d_1 D_i S_i^2 + d_2 D_i A_i^2 + d_3 D_i T_i \qquad (2)$$

where S_i, A_i, and T_i are the average status, autonomy and training for individuals in occupational category i, and D_i is a dummy variable indicating whether the origin and destination category are the same. In a different specification with aggregate covariates, Grusky and Hauser (1984; see also Hauser and Grusky 1988, p. 734) used country-level variables to explain cross-national variation in the parameters of a three-dimensional mobility table (origin by destination by country).

Models that employ aggregate covariates provide additional insights into the mobility process, but they cannot explain individual-level variation in outcomes given a common origin position. Perhaps the most straightforward way to account for this variation is to add dimensions for relevant individual-level covariates to the mobility table (e.g., Yamaguchi 1983). This is a general solution to the problem, but it is not always satisfactory. The use of additional dimensions, particularly for covariates that are continuous or have many categories, can sharply increase the number of cells in the table (which can lead to statistical problems) and the number of parameters in the model (which can make the results difficult to interpret). Consequently, more parsimonious solutions are desirable.

Logistic and probit regressions models are the most widely-known models for qualitative dependent variables that can easily include individual and structural (possibly continuous) covariates. Logistic regression models are closely related to logit models, which in turn are closely related to loglinear models. Logit models can be generalized to multinomial logit models to handle dependent variables with more than two categories (Fienberg 1980).

In what is sometimes called a universal multinomial model, individual-level covariates are assumed to affect the category a response variable falls into. These covariates have the same value for all alternatives, i.e., they are "generic" in the language of applied discrete choice modeling (Hensher and Johnson 1981).[1] The coefficients, on the other hand, are alternative-specific. In this formulation, the probability that individual "t" will move to destination j′ out of J possible destinations has the form:

$$P_{tij'} = P_{tj'} = \frac{\exp(\alpha_{j'} + \mathbf{x}_t'\boldsymbol{\beta}_{j'})}{\sum_{j=1}^{J} \exp(\alpha_j + \mathbf{x}_t'\boldsymbol{\beta}_j)} \qquad (3)$$

where $\mathbf{x}_t$ is a vector of K independent variables for individual t, α_j is the intercept for alternative j, and $\boldsymbol{\beta}_j$ is a vector of coefficients for alternative j. We can choose $\alpha_J = 0$ and $\boldsymbol{\beta}_J = 0$ as a normalization. To be strictly accurate, the index t in equation 3 runs from $1, \ldots, T_i$ within each origin group i. We can think of t as run-

[1] For an example in which covariate values vary by outcome, consider a discrete choice problem where the outcome categories are various goods that could be purchased, and one of the covariates affecting choice is the price of the alternative goods. For a mobility-related model that includes a covariate whose value varies by outcome, see DiPrete (1987).

ning from 1, . . . ,T, where T is the sample size, if we keep in mind that expressions that depend on the origin category apply only to individuals who belong to that category.

Model 3 contains no origin effects, and thus would almost never fit mobility data. Its deficiency can be remedied by including interactions between covariates (including the intercept) and origin dummy variables. Logan accomplished this task by specifying the parameters of equation 1 as functions of individual-level covariates, obtaining what he called the "Saturated Logistic Multiplicative Model" (Logan 1983, p. 328-9).[2] An equivalent model can be written in terms of the destination probabilities as

$$P_{tij'} = \frac{\exp(\alpha_{ij'} + \mathbf{x}_t'\mathbf{\beta}_{ij'})}{\sum_{j=1}^{J} \exp(\alpha_{ij} + \mathbf{x}_t'\mathbf{\beta}_{ij})} \quad (4)$$

where the set of origin categories is assumed to be the same as the set of destination categories, where $\alpha_{iJ} = 0$ and $\mathbf{\beta}_{iJ} = 0$, i = 1, . . ., J.

Even though model 4 is well understood, I provide an example for comparison with later results. I estimated a model based on model 4 with data from the 1983 and 1987 January Current Population Surveys (U.S. Bureau of the Census 1983, 1987). The data for this and subsequent examples are for full-time workers between the ages of 20 and 64 who changed jobs in the previous year, but remained with the same employer, and whose origin and destination occupations were nonfarm. The origin and destination categories I used are: (1) managerial and professional specialty occupations; (2) technical and administrative support occupations ("technical-clerical"); (3) sales occupations; (4) service occupations; (5) precision production, craft and repair occupations ("crafts"); and (6) operators, fabricators, and laborers ("other blue-collar"). The five independent variables used in this example are race, experience, experience squared, education, and college.

Coefficient estimates derived from equation 4 for male workers only are presented in Table 1. These coefficients are interpreted in standard ways (Aldrich and Nelson 1984). The difference between the $\hat{\beta}$ coefficients for a particular variable (e.g., education) for origin-destination pair ij and origin-destination pair ij′ is the estimated effect of a change in one unit of that variable on the change in the log-odds of moving to destination j as opposed to j′ from origin i. To conserve space, Table 1 reports coefficients only for workers whose origin category was technical-clerical occupations. The complete table includes five additional sets of coefficients, one for each of the five other origin categories.[3]

It is difficult to uncover important patterns when the model contains so many coefficients (180 in this example). The difficulty grows when the sample size is not large. Successful estimation of model 4 requires an enormous amount of data, with a moderate number of cases in each cell. With the 1,682 cases in the sample, convergence could not even be obtained for three of the six origin categories (sales, service, and crafts).[4] Moreover, the great majority of the estimated coefficients for the other three origin categories were not significantly different from zero at conventional standards. Such an outcome is common even with much larger sample sizes. The problems encountered in this example illustrate the limitations of multinomial logit models for the study of social mobility.

CONSTRAINED MULTINOMIAL LOGIT MODELS

An unconstrained multinomial logit mobility model can be both difficult to fit and difficult to interpret. When the categories have an underlying order (as in mobility analysis), more parsimonious models with more readily interpretable coefficients may be obtainable.

The most widely-known strategy for dealing with ordered outcomes is the ordered probit or ordered logit model (see Maddala 1983, or Miller and Volker 1985 for an application to

[2] Another mobility-related model that includes individual-level covariates was proposed by Sobel (1985). But in his model the dependent variable was a behavioral outcome other than mobility (he used fertility, specified as a continuous variable), while the independent variables were person-specific mobility effects. He specified these effects to be functions of individual-level covariates.

[3] The definitions of covariates and occupational categories appear in Appendix Table 1. Appendix Table 2 contains a crosstabulation of origin and destination occupations for the males and females who met the selection criteria.

[4] However, the value of the likelihood function had stabilized for the first five significant digits. Thus, approximate likelihood ratio tests are possible.

Table 1. Coefficients From Unconstrained Multinomial Logit Model: Male Same-Employer Job Changers From Nonfarm Origins, 1983 and 1987 CPS

Origin Occupation		Destination Occupation: Professional/ Managerial	Technical/ Clerical	Sales	Service	Crafts	Other Blue-Collar
Technical/ Clerical	Intercept	-12.00 (-3.9)	-2.37 (-1.0)	-4.66 (-1.4)	6.42 (-1.3)	-4.13 (-1.4)	0
	Race	-.70 (-.8)	-.69 (-.8)	.19 (-.1)	-1.17 (-.9)	.42 (-.3)	0
	Experience	.08 (-1.0)	-.11 (-1.5)	-.10 (-1.1)	-.26 (-.8)	-.10 (-1.2)	0
	Experience^2/100	-.07 (-.4)	.24 (-1.4)	.27 (-1.3)	-.40 (-.2)	.28 (-1.4)	0
	Education	.90 (-4.3)	.35 (-2.1)	.33 (-1.4)	-.35 (-1.0)	.33 (-1.7)	0
	College degree	-.59 (-.6)	.14 (-.2)	.75 (-.7)	.90 (.5)	-.70 (-.7)	0

Note: For the full table, N = 1682; approximate log-likelihood = -2177.2. Values in parentheses are t-statistics. Entries lacking t-statistics are fixed by design.

mobility analysis). According to a common interpretation, these models assume a latent continuous dependent variable. The researcher observes which of several mutually exclusive and exhaustive intervals (whose end points are unknown) the dependent variable falls into. The model provides parameter estimates that are asymptotically equivalent to the estimates that would be obtained if the latent variable could be measured directly.

Ordered logit or probit models would be suitable for mobility research if occupations could be ranked on a single dimension (presumably something other than status or earnings, since more efficient estimates could be obtained using these observed quantities directly as dependent variables). However, the assumption underlying the ordered logit or probit model is too restrictive for mobility analysis: while occupations are ordered, these orders are multidimensional in character.[5] Furthermore, as we shall see, both the orders and the distances between occupations are generally a function of one's origin occupation.

To simplify the exposition, I first consider models for mobility from a single origin category. Suppose that the parameter vectors β_j, $j = 1, \ldots, J$ for the alternative outcomes in equation 3 are linearly related to a smaller underlying set of parameter vectors (Anderson 1984). Specifically, the vector of parameters for each destination might be assumed to be parallel to each other. This constraint can be expressed as

$$\beta_j = \phi_j \beta \tag{5}$$

Anderson proposed the term "stereotype ordered regression" (SOR) for a logistic regression model based on this constraint. The SOR constraint forms the basis for a variety of constrained multinomial mobility models.[6]

Using the constraint in equation 5, the log odds of moving to destination j as opposed to j′ can be expressed as

$$\log \frac{P_{tj}}{P_{tj'}} = (\phi_j - \phi_{j'}) \sum_{k=0}^{K} \beta_k x_{tk} \tag{6}$$

where $x_{t0} = 1$, $t = 1, \ldots, T_i$, and the origin index "i" is suppressed since the subsample members under consideration are in the same origin category.

[5] These orderings are also incomplete, since the occupations contained within any particular occupational category are inevitably heterogeneous with respect to status, earnings and other attributes. Consequently, the distribution for any attribute in category j will typically overlap the distribution for the same attribute in category j′.

[6] The SOR constraint can be used in a multinomial probit or other multinomial model as well. I focus on the multinomial logit model because of its computational advantages and its relationship to loglinear and association models.

In this formulation, the covariate coefficients (β_k, k = 1, . . ., K), can be interpreted as factors that convert the K covariates to a common metric. If each of the variables in the equation could be conceptualized as a resource, for example, this metric might be referred to as the metric of "generic" resources (or liabilities). The effect of a unit increase in education on the log odds of moving to destination j as opposed to j′ could then be interpreted as equivalent to the effect of β_{educ} units of some generic resource. In the metric of generic resources, this effect equals the difference between the scores for destinations j and j′ ($\phi_j - \phi_{j'}$).

This way of interpreting the model parameters does not necessarily imply that the grouped covariates are indicators of a latent variable. One need only assume that the direction of their impact (the order and relative distance between response categories imposed by the group of covariates) is the same. However, such a proportionality constraint on the effects of the covariates might imply the existence of a latent composite variable that affects the respondent's destination category (see Hauser and Goldberger 1971, and Hauser, Tsai, and Sewell 1983 for a structural equation model that contains a latent composite variable).

Another interpretation of the ϕ scores can be obtained by dividing both sides of equation 6 by ($\phi_j - \phi_{j'}$):

$$\frac{\log \dfrac{P_{tj}}{P_{tj'}}}{(\phi_j - \phi_{j'})} = \sum_{k=0}^{K} \beta_k x_{tk} \tag{7}$$

In effect, the ($\phi_j - \phi_{j'}$) factor rescales the log odds implied by $\sum_{k=0}^{K} \beta_k x_{tk}$ for each pair of occupational categories (j, j′).[7] But while the rescaled logits are linear functions of β, the logits themselves are multiplicative, not linear, functions of the model parameters.

Model 6 is not identified unless a normalization is imposed. The overabundance of parameters in model 6 is easily seen by setting J=2 and K=3. In that case, the familiar logistic regression would be

$$\log \frac{P_{tj}}{P_{tj'}} = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} \tag{8}$$

[7] A reviewer suggested this "rescaled logit" interpretation.

where j = 1 and j′ = 2. A comparison of equations 6 and 8 shows that both ϕ coefficients are superfluous.

There are two strategies for normalizing the ϕ vector. The first, which is analogous to the common normalization for the multinomial logit model (see equation 3), fixes a zero point and a "unit" distance for the ϕ scores in terms of which other distances can be measured. To carry out this normalization, one might set $\phi_J = 0$, and $\phi_1 = 1$. In the examples here, (where group 1 is professional-managerial and group J is other blue-collar) this would set the "distance" between professional-managerial and other blue-collar occupations (roughly the top and the bottom of the status hierarchy) equal to one unit. Since mobility between these two groups is relatively rare, most other estimated distances would be smaller. Alternatively, one might set the distance between a pair of "neighboring" occupational categories equal to one unit (e.g., one might set $\phi_{\text{other blue-collar}} = 0$ and $\phi_{\text{crafts}} = 1$). This would cause most other distances to exceed one unit. A second strategy, which is analogous to Goodman's (1979) treatment of association models, imposes a linear constraint on the sum of the ϕ scores (e.g., $\sum\phi = 0$), and a constraint on the magnitudes of the scores (e.g., $\sum\phi^2 = 1$). Other normalizations are possible, and more complex normalizations are required with more complex models.

While both model 6 and the common ordered logit model are "one-dimensional," they are not equivalent. The "natural" logits corresponding to the ordered logit model compare $\sum_{j \geq j'} p_j$ with $\sum_{j<j'} p_j$, for any j′, while the "natural" logits corresponding to multinomial logit models (regardless of whether they are constrained as in equation 6) compare p_j and $p_{j'}$ for any (j, j′) pair. The "natural" latent structure interpretation of the ordered logit model involves a latent dependent variable, while the "natural" latent structure interpretation of the SOR model involves a latent composite independent variable. (Differences between these models are discussed at greater length in the Appendix.)

While the model based on the constraint of equation 5 has a simple interpretation, it may be too parsimonious to fit the data adequately in a substantive application. For instance, equation 6 contains the problematic assumption that the baseline logits (in which all covariates are set to zero) equal a simple multiple (β_0) of differences between the same ϕ scores that apply to the covariates. One way to relax the constraint in equation 5 is to introduce a set of unconstrained intercepts into equation 6:

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + \sum_{k=1}^{K} (\phi_j - \phi_{j'})\beta_k x_k \quad (9)$$

This model has two distinct score vectors: α specifies the intercepts, while ϕ specifies the effects of the other covariates on outcomes. The intercepts are typically normalized by setting $\alpha_J = 0$, though, again, other choices are possible.

Panel A of Table 2 illustrates this specification, obtained through maximum likelihood analysis of equation 9 applied to male technical-clerical workers changing jobs but remaining with the same employer.[8] The most highly ranked occupational group is professional-managerial ($\phi = 1$), followed by sales ($\phi = .48$), technical-clerical ($\phi = .36$), crafts ($\phi = .23$), other blue-collar ($\phi = 0$) and service ($\phi = -.35$) in that order. The probability that a worker will end up in a more highly ranked (by the scores) of any pair of occupations increases with both education and experience (the latter effect being somewhat curvilinear). In addition, the model includes a set of baseline intercepts that need not reflect the hierarchy described by the resource scores. A likelihood ratio tests shows that, relative to model 3, model 9 provides a reasonable fit to the data ($L^2_{16} = 27.6$). A comparison of BIC scores (Raftery 1986a, 1986b) shows model 9 (BIC = -59) to be superior to model 3 (BIC = 0).

Equation 9 uses different normalizations for the intercept scores and the resource scores. While only the J-th α is fixed, equation 9 appears to imply a constraint on both the order and the distance between the first and last resource scores, because the former is fixed at 1, while the latter is fixed at 0. But the normalization used for the ϕ scores is actually no more constraining than that for the α scores. The specification of a unit distance between the top and the bottom category is completely general, since the ϕ scores only determine *relative* distances among categories. One can extract a common scale factor from the components of the β vector to convert these relative magnitudes to absolute magnitudes. Furthermore, the apparent order of the resource scores can be maintained or reversed according to the sign of the corresponding β coefficient.

The different ways that the α and ϕ vectors were normalized in equation 9 suggests that alternative normalizations are possible. As it stands, rescaled (by β) covariates weigh on the ϕ dimension, while an "unrescaled" constant ("1") weighs on the α dimension. Equation 9 could be written differently as

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'})\beta_0 + \sum_{k=1}^{K} (\phi_j - \phi_{j'})\beta_k x_{tk} \quad (10)$$

Here, β_0 functions as a conversion factor, as do $\beta_1, \beta_2, \ldots, \beta_K$. This formulation requires a different normalization. For example, α_1 could be set to 1, and α_J set to 0, so that the normalization of the two dimensions would be identical. But a little reflection reveals that conversion factors make sense only when more than one variable is weighing on a common dimension. If only one variable is weighing on a particular dimension, its natural metric can be used, and its β coefficient is superfluous. The general principle is that a set of J-1 distinct scores can be employed whenever the applicable variables are measured in a metric natural to one variable in the group. The β coefficient for the variable that supplies the metric would normally be set to unity, analogous to the treatment of β_0 in equation 9. Applying this logic, a treatment of the resource scores analogous to the treatment of intercepts in equation 9 is obtained:

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + (\phi_j - \phi_{j'})x_1 + \sum_{k=2}^{K} (\phi_j - \phi_{j'})\beta_k x_k \quad (11)$$

where only α_J and ϕ_J are constrained. This specification provides a kind of standardization for the covariate effects within the group, with the variable that supplies the metric (x_1 in equation 11) being the "scale referent."[9] The ϕ scores describe the effect of these standardized covariates on the outcome.

These three alternative specifications (equations 9, 10, and 11) are algebraically equivalent; the choice of model can be based on their

[8] Estimates were calculated using an algorithm that switches between the Berndt, Hall, Hall, and Hausman method (1974) and the Broyden, Fletcher, Goldfarb, and Shanno method (see Dennis and Schnabel 1983). The algorithm was programmed in GAUSS386 by the author, and is available upon request.

[9] A reviewer suggested this terminology for equation 11.

Table 2. Coefficients of Selected Models: Same Employer Job Changers From Technical/Clerical Origins: 1983 and 1987 CPS

	Male				Male and Female					
	Panel A (Model 9)		Panel B (Model 11)		Panel C (Model 13)			Panel D (Model 15)		
	Inter- cepts	Effects	Inter- cepts	Effects	Inter- cepts	Effects	Effects	Inter- cepts	Male Scores	Effects
Destination	α	ϕ	α	ϕ	α	ϕ_1	ϕ_2	α	ϕ_2	ϕ_1
Professional/ managerial	-13.41 (-4.4)	1	-13.41 (-4.4)	.88 (4.1)	-9.09 (-4.5)	1.82 (1.2)	3.59 (1.5)	-10.29 (-5.4)	-1.71 (-5.0)	1
Technical/ clerical	-4.00 (-4.4)	.36 (3.0)	-4.00 (-4.4)	.32 (2.3)	-.35 (-.2)	1.87 (1.9)	1.45 (.9)	-1.70 (-1.1)	-2.09 (-6.7)	.39 (4.3)
Sales	-6.86 (-1.9)	.48 (2.6)	-6.86 (-1.9)	.42 (2.3)	-4.1 (-2.1)	1.19 (1.5)	1.68 (1.3)	-5.03 (-2.6)	-1.00 (-2.4)	.46 (3.6)
Service	2.96 (.8)	-3.5 (-1.0)	2.96 (.8)	-3.1 (-1.1)	.54 (.2)	1	0	-1.20 (-.5)	-0.74 (-1.3)	.047 (.2)
Crafts	-3.22 (-1.5)	.23 (1.6)	-3.22 (-1.5)	.20 (1.4)	-3.87 (-1.9)	0	1	-3.94 (-2.1)	0.28 (0.6)	.29 (2.2)
Other blue-collar	0	0	0	0	0	0	0	0	0	0
Covariates		ß		ß		γ	δ			ß
White		-.20 (-.3)		-.23 (-.3)		-.47 (-1.3)	.18 (.8)			.26 (.6)
Male		—		—		-1.30 (-2.2)	.14 (.3)			—
Experience		.21 (2.5)		.23 (2.2)		-.035 (-.9)	.035 (1.4)			.093 (2.1)
$\text{Experience}^2/100$		-.0029 (-1.7)		-.0033 (-1.5)		5.46 (0.7)	-4.93 (-1.1)			-.13 (-1.2)
Education		.88 (4.1)		1		-.016 (-.1)	.23 (1.8)			.84 (6.1)
College degree		-.72 (-.9)		-.82 (-1.1)		.094 (.3)	-.20 (-1.1)			-.76 (-1.5)
Number of cases	220		220			812			812	
Log-likelihood	-321.7		-321.7			-934.8			-938.5	

Note: Values in parentheses are t-statistics. Entries lacking t-statistics are fixed by design. See text for additional details.

ability to reveal meaningful aspects of the mobility process. Estimates from equations 9 and 10 will be identical except for the intercepts. In contrast, estimates of the ϕ and ß vectors using equation 11 will differ from the analogous estimates using equation 9 or 10.

Panel B of Table 2 displays results using equation 11 for the same data. Years of education is the scale referent. The point estimates indicate that a year of experience is "worth" about 1/4 of a year of education for relatively young men, and progressively less for older workers. While the effects of race and college are not significant in this model, the point estimates imply that being white "costs" about 1/4 year of education, while college "costs" about 4/5 of a year of education. The scores in panels A and B are slightly different, while the intercepts are identical, as they should be. The log-likelihood values for the two models are also identical. In this case, the covariate coefficients and scores in panel B are similar to those in panel A because the unconstrained effect of education in panel A (.88) is close to its constrained value in panel B (1.0). Generally speaking, however, the choice between these two specifications depends upon whether clarity is best achieved by expressing all effects in "units" of a year of education (or some other covariate), or by scaling the ϕ scores such that the "distance" between two given occupational groups (in this case professional-managerial and other blue-collar) equals unity.

ELABORATION OF THE BASIC MODEL

Models 9 or 11 provide appealingly simple descriptions of the mobility process. However, these highly constrained models may not provide acceptable fits to the data. If the fit is poor, the analyst must reduce the number of constraints by increasing the number of dimensions of scores. Two strategies are available. Anderson (1984) suggested using multiple dimensions of scores for each covariate. An alternative strategy, which may provide more interpretable results, uses separate dimensions for mutually exclusive subsets of covariates. I discuss each of these strategies below.

Anderson suggested that the researcher specify a higher dimensional solution when the one-dimensional solution does not fit. For the two-dimensional solution, the following relationship holds among the parameter vectors:

$$\beta_j = \phi_{1j}\gamma + \phi_{2j}\delta \qquad (12)$$

where the following set of constraints is one possible normalization:[10]

$$\phi_{12} = \phi_{1J} = \phi_{21} = \phi_{2J} = 0$$
$$\phi_{11} = \phi_{22} = 1$$

This formulation can be expanded to three-dimensions and beyond. Each expansion increases the number of constraints needed to normalize the resource scores, even though it reduces the constraints on the β_j vectors, j = 1, . . ., J. When the number of dimensions equals the minimum of J-1 and K (where J is the number of destinations and K is the number of covariates weighing on the resource scores), the resulting pattern of constraints yields a different form of the multinomial logit model.

One interpretation of model 12 is that each covariate contains two conceptually distinct types of resources (alternatively, each variable weighs on two distinct latent composite variables). The γ vector extracts the first dimension, while the δ vector extracts the second dimension. Each resource type has a distinct directional impact on mobility outcomes, as described by the corresponding set of resource scores. The effect of the first dimension of resources on the probability of transition from j to j′ depends on the first dimension of scores (ϕ_1), while the effect of the second dimension of resources on the probability of transition depends on the second dimension of scores (ϕ_2). Finding a substantive interpretation for these dimensions, however, might be difficult. The more complicated normalization also creates difficulties of interpretation. These difficulties increase as additional dimensions are included in the model.

Including women in the analysis and adding gender to the list of covariates illustrates the potential value and attendant difficulties that might arise from such a model. The fit of model 9 to the data for both genders is much worse compared with the fit of model 3 than was the case when only men were in the sample (L^2_{20} = 104.5, estimates not shown), though the BIC score (-29.5) favors the more parsimonious model over the unconstrained multinomial logit model. The fit is poor because the gender effects on the log odds of moving to j as opposed to j′ for each (j, j′) pair are not proportional to the effects for the other variables in the model.

The fit is improved substantially by using the constraint found in equation 12 to allow for two distinct dimensions of scores:

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + \sum_{k=1}^{K} (\phi_{1j} - \phi_{1j'})\gamma_k x_{tk} + \sum_{k=1}^{K} (\phi_{2j} - \phi_{2j'})\, \delta_k x_{tk} \qquad (13)$$

where x_t in this case consists of race, experience, experience-squared, education, college and gender. In applying model 13, I equated the scores of service and other blue-collar occupations on the first dimension, and set the distance between crafts and these two groups to be one unit. I equated the scores of crafts and other blue-collar occupations on the second dimension, and specified the distance between service and these two groups to be one unit. This specification allows the effects of covariates on outcomes to work through two distinct sets of distances among occupational groups. Results are presented in panel C.

Model 13 fits the combined male-female sample better than model 9 (L^2_{12} = 31.6, BIC = -49). The effect of gender is much larger on the first dimension of this model than on the second. On the first dimension, being male in-

[10] An alternative normalization is $\Sigma\phi_1 = \Sigma\phi_2 = \Sigma\phi_1\phi_2 = 0$; $\Sigma\phi_1^2 = \Sigma\phi_2^2 = 1$.

creases the odds of moving to a craft or other blue-collar job relative to other destinations. On the second dimension, the effect of gender is not significant. Education's effect, in contrast, is located primarily on the second dimension, where the distance between professional-managerial and other occupations is larger than on the first dimension. The effects of race and experience have opposite signs on the two dimensions.

But model 13 is unsatisfactory for both statistical and theoretical reasons. None of the estimated scores is significantly different from zero at the .05 level. Aside from the gender effect on the first dimension and the education effect on the second dimension, none of the covariate coefficients are significant at even the .10 level. Model 13 is theoretically unsatisfying, too. A reasonable hypothesis in this case might be that gender's effect on outcomes is expressed on a different dimension than the effects of the other covariates. But model 13 divides the effects of *every* covariate along two distinct dimensions. The interpretation of these separate dimensions is not obvious, and the overall effect of education, experience, or gender on outcomes is not easy to determine without additional algebraic computation.[11] For some problems this model, in which the effects of covariates operate through two distinct latent composite variables, would follow naturally from theory. But in the current case, the poor fit of model 9 is due primarily to gender's unique pattern of effects on outcomes, not to the failure of a unidimensional specification for education or experience.

An alternative modeling strategy, analogous to the earlier special treatment of the intercepts in models 9, 10 and 11, is to restrict variables from weighing on more than one dimension of scores unless theory justifies a more complex specification. With this strategy, parsimony is achieved by grouping variables that measure similar resources or that otherwise have similar effects on the dependent variable. The number of score vectors increases with the number of groups. When the number of groups equals the number of covariates (i.e., each covariate is the sole member of its group), then all of the ß parameters would be normalized to unity (compare the treatment of $ß_0$ in equation 11), and the resulting specification would be identical to the multinomial logit model.

This strategy can be implemented in more than one way. For example, one might hypothesize that "ascription" variables such as race and gender work differently than "achievement" variables such as education and experience. This implies that

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + \sum_{k=1}^{K-2} (\phi_{1j} - \phi_{1j'}) ß_k x_{tk} + \sum_{k=K-1}^{K} (\phi_{2j} - \phi_{2j'}) ß_k x_{tk} \quad (14)$$

where ϕ_1 is the score vector for "achievement" covariates, ϕ_2 is the score vector for "ascription" covariates, x_{tK-1} is race, and x_{tK} is gender. The ϕ scores can be normalized by setting $\phi_{1J} = \phi_{2J} = 0$ and $\phi_{11} = \phi_{21} = 1$. Alternatively (and perhaps more realistically), one might specify the race effect to be proportional to the effects of other variables in the model, but the gender effect to be nonproportional.

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + \sum_{k=1}^{K-1} (\phi_{1j} - \phi_{1j'}) ß_k x_{tk} + (\phi_{2j} - \phi_{2j'}) ß_K x_{tK} \quad (15)$$

The normalizations in equation 15 are the same as in equation 14, except that $ß_K$ instead of ϕ_{21} can be specified to equal 1, because the second group of covariates now consists only of the gender variable.

Panel D provides estimates for model 15, with $ß_K$ set to unity. This model fits better (in a BIC sense: $L^2_{16} = 38.9$, BIC = -68) than model 13.[12] The effects of education and experience are easier to interpret in panel D than they are in panel C (the overall pattern is similar to that of panel A). Furthermore, the gender effects are easily interpreted — male technical-clerical workers changing their jobs with the same employer are less likely to move to a nonmanual occupation and more likely to move to a manual occupation than are female workers. While

[11] Additional interpretive difficulties, related to the more complex normalization required, might be alleviated somewhat by using an alternative normalization. But the model would still be unsatisfactory without a plausible interpretation of the two resource dimensions on which each of the covariates is presumed to weigh.

[12] Models 13 and 15 are not nested, and so likelihood ratio tests are inappropriate. However, the BIC scores of models 3, 9, 13, and 15 can be compared. The L^2 reported in the text compares model 15 to model 3.

this result superficially resembles that for the other covariates (with female gender considered a resource), the occupational distances implied by the gender effect are very different from the distances implied by the other five covariates in the model (compare the second and third columns of panel D).

Interaction effects can enter these models in two ways: they can affect the covariate coefficients, and they can affect the pattern of scores. An elaboration of equation 15 can illustrate these possibilities. Because the CPS does not provide an exact measure of work experience, the examples of this paper use a standard proxy (age – education – 5), which might more accurately be called potential experience. But many women withdraw from the labor force at some point to bear and raise children, so this measure has a weaker relationship with actual experience for women than for men. An interaction between gender and experience might therefore be warranted to take account of this measurement problem. With its inclusion, equation 15 becomes:

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + \sum_{k=1}^{K-1} (\phi_{1j} - \phi_{1j'}) \beta_k x_{tk} + (\phi_{2j} - \phi_{2j'}) \beta_K x_{tK} + (\phi_{1j} - \phi_{1j'}) \beta_{K+1} x_{tk'} x_{tK} \qquad (16)$$

where k' is the index for the experience variable, and K is the index for gender. Broadly speaking, equation 16 implies that the resource scores for covariates (aside from the gender covariate) have the same value for men and women, but that the conversion factors for (some) covariates are gender-specific.

To illustrate the second type of interaction, one might argue that men and women differ not in conversion factors, but in the effect of these converted covariates (equivalently, the effect of the latent composite variable) on outcomes. This hypothesis would be expressed in terms of interactions between gender and the resource scores of equation 15:

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + \sum_{k=1}^{K-1} [(\phi_{1j} - \phi_{1j'}) + (\psi_{1j} - \psi_{1j'}) x_{tK}] \beta_k x_{tk} + (\phi_{2j} - \phi_{2j'}) \beta_K x_{tK} \qquad (17)$$

Here x_{tK} is the gender variable, and the ψ_1 vector of scores contains the gender-specific increments to the ϕ_1 vector of scores. In effect, specification 17 contains two different sets of scores, one for males and one for females.[13] Finally, it is possible to specify and estimate a model that contains gender interactions with covariate coefficients and gender interactions with scores.

It is not necessary for a particular covariate (gender, in this example) to interact with all the components of a score vector. More restricted interactions are also possible. For example, mobility analysts generally argue that "immobility effects" at both the individual and category level are necessary to account for the diagonal frequencies of a mobility table (e.g., Logan 1983; Hout 1984). While the α_j parameter in the above models is a measure of immobility when j = i (the origin category),[14] these specifications do not take account of the possibility that a particular individual-level variable such as experience might have a distinctive immobility effect. To allow for this possibility, we can include an interaction between experience and the score for the origin category of each respondent in equation 9:

$$\log \frac{P_{tj}}{P_{tj'}} = (\alpha_j - \alpha_{j'}) + \sum_{k=1}^{K} (\phi_j - \phi_{j'}) \beta_k x_{tk} + (\psi_j \delta_{ij} - \psi_{j'} \delta_{ij'}) x_{tk'} \qquad (18)$$

where k' is the index for the experience variable, and $\delta_{ij} = 1$ if j = i (the origin category), 0 otherwise. An example of an experience-origin interaction is presented later.

SIMULTANEOUS ESTIMATION WITH MULTIPLE ORIGIN CATEGORIES

The models discussed to this point pertain to a single origin category. One way of analyzing the entire sample is to estimate the above models separately for each origin category. This straightforward elaboration occasionally may be the only strategy that yields a good fit and a

[13] However, the constraints on the ψ_1 vector are not the same as those on the ϕ_1 vector. While equation 17 implies that $\phi_{11} = 1$, there is no comparable restriction on ψ_{11}. In other words, the specification in equation 17 allows the distance between scores for destinations 1 and J to be different for males and females.

[14] Because the sample is limited to job changers, immobility does not have the same interpretation here that it would have in an analysis of the complete sample.

Table 3a. Interaction Effects and Scores from Unconstrained Intercepts Model: Male Same-Employer Job Changers from Nonfarm Origins, 1983 and 1987 CPS

	Origin Occupation	Destination Occupation: Professional/ Managerial	Technical/ Clerical	Sales	Service	Crafts	Other Blue-Collar
Interaction Effects	Professional/ managerial	-8.29 (-4.5)	-3.91 (-2.6)	-5.10 (-3.2)	1.70 (.8)	-1.99 (-1.3)	0
	Technical/ clerical	-10.21 (-5.7)	-3.90 (-2.7)	-6.55 (-4.3)	-.018 (-.0)	-3.00 (-2.0)	0
	Sales	-9.38 (-5.1)	-5.14 (-3.4)	-4.18 (-2.7)	.18 (.1)	-2.45 (2.1)	0
	Service	-7.50 (-5.6)	-5.74 (-5.3)	-7.58 (-6.2)	-.68 (-.8)	-2.67 (-3.8)	0
	Crafts	-7.34 (-6.9)	-5.13 (-4.3)	-7.00 (-6.2)	-2.95 (-3.6)	-1.60 (-2.6)	0
	Other blue-collar	-8.53 (-6.6)	-6.13 (-6.2)	-7.57 (-6.9)	-3.37 (-4.3)	-2.66 (-4.4)	0
Resource Scores	Nonmanual	1	.46 (4.7)	.59 (6.7)	-.14 (-.6)	.28 (2.4)	0
	Manual	1	.68 (6.5)	.84 (6.5)	.21 (1.8)	.31 (3.8)	0

theoretically satisfying model. If the underlying mobility structure allows simplification, however, it is possible to achieve significant gains in parsimony by constraining either the covariate coefficients or the resource scores (or both) across origin categories.

For example, one might constrain resource scores to be the same across all origin categories while allowing the conversion factors to vary by origin:

$$\log \frac{P_{tij}}{P_{tij'}} = (\alpha_{ij}-\alpha_{ij'}) + \sum_{k=1}^{K} (\phi_j-\phi_{j'})\beta_{ki}x_{tk} \qquad (19)$$

where i indexes origin categories, j indexes destination categories, K is the number of covariates, and the i subscript on each ß reflects the existence of J distinct **ß** vectors.

An alternative specification allows resource scores to vary across origin categories while fixing the conversion factors:

$$\log \frac{P_{tij}}{P_{tij'}} = (\alpha_{ij}-\alpha_{ij'}) + \sum_{k=1}^{K} (\phi_{ij}-\phi_{ij'})\beta_k x_{tk} \qquad (20)$$

where the i subscript on each ϕ reflects the existence of J distinct ϕ vectors.

A third strategy imposes constraints on both resource scores and conversion factors. As an example, the following model constrains the resource scores for nonmanual workers to be the same regardless of origin category, and

Table 3b. Covariate Coefficients from Unconstrained Intercepts Model: Male Same-Employer Job Changers from Nonfarm Origins, 1983 and 1987 CPS

Covariate	Covariate Coefficients: Nonmanual Origin	Manual Origin
White	.16 (.3)	.89 (2.1)
Experience	.12 (2.8)	-.072 (-2.1)
Experience²/100	-.15 (-1.4)	.17 (2.0)
Education	.68 (6.5)	.50 (5.5)
College degree	-.076 (-.2)	.65 (1.3)

Note: For the model illustrated by Tables 3a and 3b, N = 1682; log-likelihood = -2252.1. Values in parentheses are t-statistics. Entries lacking t-statistics are fixed by design.

imposes a similar constraint for manual workers. It imposes analogous constraints on the covariate coefficients, so that there is one set for nonmanual workers and another for manual workers. The resulting specification, which might be termed a white-collar/blue-collar model, is

$$\log \frac{P_{tij}}{P_{tij'}} = (\alpha_{ij}-\alpha_{ij'}) + \sum_{k=1}^{K} (\phi_{sj}-\phi_{sj'})\beta_{ks}x_{tk} \qquad (21)$$

where s = 1 for nonmanual workers and s = 2 for manual workers.

Tables 3a and 3b provide an illustration of this model for males in the sample. While it was not possible to estimate all parameters of the unconstrained multinomial logit model with these data (see the discussion of Table 1), estimation of equation 21 poses no problem. This model provides an acceptable fit to the data compared to model 4 (L^2_{132} = 149.9, BIC = -830.6). Furthermore, the resource scores and covariate coefficients have ready interpretations.[15] For example, the estimates reveal systematic differences in the impact of individual covariates on outcomes for nonmanual and manual job changers (Table 3b). Being white (or its unmeasured correlates) appears to be a resource for manual workers, but not for nonmanual workers. Experience protects nonmanual job changers from falling in the hierarchy defined by the resource scores, and it inhibits upward movement from manual origins. Professional-managerial and other nonmanual destinations are further apart on the nonmanual-origin hierarchy than on the manual-origin hierarchy; this implies that resources have greater influence on the odds of moving to professional-managerial vs. a lower nonmanual destination for workers from nonmanual origins than they do for workers from manual origins.[16] Even if it had been possible to estimate the unconstrained multinomial logit model with these data, it would have been difficult to discern such patterns from the huge number of coefficients produced. In contrast, estimation using equation 21 makes such patterns comparatively clear.

ACHIEVING PARSIMONY IN THE SPECIFICATION OF INTERCEPTS

The models discussed to this point achieve parsimony by reducing the number of parameters required to describe the effects of covariates on outcomes. However, the models of the previous section still contain J(J-1) distinct intercept parameters (J-1 for each origin category) in addition to the scores and covariate coefficients. Clearly, it would be desirable to reduce the number of intercept parameters, especially if such a reduction produced additional insights into the structure of mobility. The natural way to achieve this goal is to take advantage of the correspondence between intercepts and effect parameters in loglinear and related mobility models.

For illustrative purposes, I will reparameterize model 21. Eliminating the covariates from model 21 produces:

$$\log \frac{P_{ij}}{P_{ij'}} = (\alpha_{ij} - \alpha_{ij'}) \tag{22}$$

The logits on the left side of equation 22 can also be expressed as functions of the u terms of a standard loglinear model (e.g., Fienberg 1980).

$$\begin{aligned} \log(m_{ij}/m_{iJ}) = \log(P_{ij}/P_{iJ}) &= \alpha_{ij} \\ &= u_{2(j)} - u_{2(J)} + u_{12(ij)} - u_{12(iJ)} \\ &= w_{2(j)} + w_{12(ij)} \end{aligned}$$

Several well-known techniques can be used to reduce the number of u terms needed to describe a mobility table (e.g., Hout 1983). These same techniques can be used to reduce the number of α terms, by expressing the α terms as functions of the reduced set of u terms.

As an example of such a reduction, I apply the strategy employed by Featherman and Hauser (1978, pp. 147-50) in their study of intergenerational mobility. Featherman and Hauser proposed a mutually exclusive and exhaustive partition of the interaction terms in the saturated loglinear model into R categories, where H_r, r = 1, . . ., R is the set of ij indices for the interaction terms in the r-th partition. They specified the loglinear model in terms of this partition as follows:

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(r)}$$

where $u_{12(ij)} = u_{12(r)}$ for ij ε H_r, and $\sum_i \sum_j u_r = \sum u_r f_r = 0$ where f_r is the number of cells located in the r-th partition. We can specify α_{ij} in terms of these reparameterized u terms:

$$\alpha_{ij} = w_{2(j)} + u_{12(r_{ij})} - u_{12(r_{iJ})} \tag{23}$$

[15] A more thorough discussion of the results from this and related models can be found in DiPrete and Krecker (unpublished).

[16] This statement applies to the latent resources composite, not necessarily to the measured covariates themselves, since the covariate coefficients for the two strata can differ. In order to assess the validity of this statement when applied to the measured covariates, one could estimate a version of equation 21 in which the conversion factors were constrained to be the same for nonmanual and manual origins. This more tightly constrained model does not fit these data as well as equation 21.

Table 4a. Interaction Effects, Destination Effects, and Scores from Constrained Intercepts Model, Including Category Persistence Interaction with Experience: Male Same-Employer Job Changers from Nonfarm Origins, 1983 and 1987 CPS

	Origin Occupation	Destination Occupation: Professional/ Managerial	Technical/ Clerical	Sales	Service	Crafts	Other Blue-Collar
Interaction Effects	Professional/ managerial	2.01=**1** (11.2)	.19=**4** (2.7)	**4**	-1.44=**6** (-8.4)	**5**	**6**
	Technical/ clerical	**4**	1.35=**2** (11.6)	**4**	**6**	**5**	**5**
	Sales	.86=**3** (7.0)	-.37=**5** (-6.2)	**1**	**6**	**5**	**6**
	Service	**5**	**5**	**6**	**1**	**5**	**4**
	Crafts	**5**	**4**	**5**	**5**	**2**	**4**
	Other blue-collar	**5**	**4**	**5**	**4**	**3**	**2**
Destination Effects		.26 (2.8)	-.19 (-2.1)	-.15 (-1.5)	-.52 (-4.4)	.14 (1.5)	.46 (6.6)
Resource Scores	Nonmanual	1	.40 (3.6)	.53 (5.3)	-.43 (-2.1)	.19 (1.4)	0
	Manual	1	.70 (6.7)	.89 (8.3)	.19 (1.7)	.34 (4.3)	0

where r_{ij} is the partition containing the ij cell. This new parameterization gives J-1+R-1 distinct parameters for equation 22 instead of J(J-1). If R-1 is much smaller than (J-1)(J-1), a substantial reduction in the number of parameters needed to specify the model results.

The parameter reduction for model 21 was achieved by excluding all covariates from the model. If covariates are present, the value of intercepts may change substantially, altering the simple relationship between the intercepts and the loglinear parameters discussed above. There is, however, a simple solution to this problem: Express the covariates for each individual in the sample as deviations from sample means for the individual's origin category. Using these deviations, equation 21 becomes

$$\log \frac{P_{tij}}{P_{tij'}} = (\alpha_{ij}-\alpha_{ij'}) + \sum_{k=1}^{K} (\phi_{sj}-\phi_{sj'})\beta_{ks}(x_{tk}-\bar{x}_{ik}) + \sum_{k=1}^{K} (\phi_{sj}-\phi_{sj'})\beta_{ks}\bar{x}_{ik}$$

$$= (\alpha_{ij} + \sum_{k=1}^{K} \phi_{sj}\bar{x}_{ik}) - (\alpha_{ij'} + \sum_{k=1}^{K} \phi_{sj'}\bar{x}_{ik}) + \sum_{k=1}^{K} (\phi_{sj}-\phi_{sj'})\beta_{ks}(x_{tk}-\bar{x}_{ik})$$

$$= (\tilde{\alpha}_{ij}-\tilde{\alpha}_{ij'}) + \sum_{k=1}^{K} (\phi_{sj}-\phi_{sj'})\beta_{ks}\tilde{x}_{tk} \qquad (24)$$

Table 4b. Covariate Coefficients from Constrained Intercepts Model: Male Same-Employer Job Changes from Nonfarm Origins, 1983 and 1987 CPS

Covariate	Covariate Coefficients: Nonmanual Origin	Manual Origin
White	.21 (.5)	.91 (2.3)
Experience	.13 (3.2)	-0.79 (-2.2)
Experience²/100	-.17 (-1.7)	.16 (1.9)
Education	.57 (5.2)	.48 (5.9)
College degree	.023 (.1)	.69 (1.5)
Persistence	-.014 (-2.0)	-.011 (-1.9)

Note: For the model illustrated by Tables 4a and 4b, N = 1682; log-likelihood = -2261.2. Values in parentheses are t-statistics. Entries lacking t-statistics are fixed by design. Entries in **bold** type describe the pattern of constraints. See text for details.

where s = 1 if the origin is a nonmanual occupation and s = 2 if the origin is a manual occupation, $\bar{x}_{ik}$ is the mean for the k-th covariate for the i-th origin group, $\tilde{x}_{tk}$ is the deviation from the appropriate origin mean for the k-th covariate for the t-th individual, and $\tilde{\alpha}$ is a function of α and terms involving origin-specific means for the covariates. Consequently, the second term in equation 24 disappears for an individual whose value on each covariate is at the mean of his or her origin group. With this specification, the log odds of moving to one category versus another for an "average" individual in origin category i equals the appropriate difference between the transformed intercepts. The number of intercept parameters can then be reduced substantially by employing existing strategies found in the mobility literature. For example, using the Featherman-Hauser strategy, we could write

$$\tilde{\alpha}_{ij} = w_{2(j)} + u_{12(r_{ij})} - u_{12(r_{iJ})} \tag{25}$$

As an illustration, I estimate the following model:

$$\log \frac{P_{tij}}{P_{tij'}} = (\tilde{\alpha}_{ij}-\tilde{\alpha}_{ij'}) + \sum_{k=1}^{K} (\phi_{sj}-\phi_{sj'})\beta_{ks}\tilde{x}_{tk} + (\gamma_s\tilde{x}_{tk'}\delta_{ij} - \gamma_s\tilde{x}_{tk'}\delta_{ij'}) \tag{26}$$

where $\tilde{\alpha}$ is constrained as in equation 25 with R = 6, where s = 1 or 2 depending upon whether an individual's origin category is nonmanual or manual, and where the last term contains an additional experience effect on the probability of leaving one's origin category (which is assumed to have one value for all nonmanual origins, and another for all manual origins). The partition for $\tilde{\alpha}$ and the estimates for model 26 are shown in Tables 4a and 4b.

In comparison with model 4, model 26 reduces the number of parameters needed to describe the structure of mobility from 180 to 32. The fit of this model is rather good by conventional standards ($L^2_{148} = 168$), and the BIC score suggests it is superior to model 21 as well (BIC = -938.7 for model 26 vs. -830.6 for model 21). The specification of the intercepts in terms of the column and interaction effects of standard loglinear models makes their interpretation noticeably easier. For example, it is much easier to see in Tables 4a and 4b than in Tables 3a and 3b that the "average" job changer is more likely to stay within his origin group than to move to any other single category, that nonmanual job changers are more likely to stay within their stratum than to move to a manual job, and that crafts or other blue-collar job changers are more likely to end up in one of these two groups than to obtain a service, sales, or professional-managerial job. While model 26 constrains these interaction terms along the lines of the Featherman-Hauser model, the scoring models of Haberman or Goodman could have been used instead to reparameterize the interactions. It would even be possible to relate the scores for the covariates and the scores for the baseline interactions if there were a theoretical advantage to doing so. Finally, the equivalent of a saturated model for the intercepts could be estimated, if that model were appropriate.

CONCLUSION

I have discussed a class of constrained multinomial logit models that are practical to use and that offer the possibility of new insights into the process of social mobility. While the sample size used in the examples of this paper is relatively small, similar models with much larger samples are practical, even when using a microcomputer. With more powerful computers, estimating models with more covariates and even larger samples is possible.

To realize the full potential of this modeling strategy, the researcher must specify the constraints in a way that best illuminates the underlying structure of mobility. Goodness-of-fit considerations must also be kept in mind; a badly misspecified model may distort rather than reveal major patterns in the data. While constructing sensible specifications is more work than mechanically estimating an unconstrained multinomial logit model, the result is a gain in knowledge about the influence of individual and structural factors on mobility outcomes. From both a practical and a theoretical point of view, these models help close the gap between the status attainment and the loglinear approach to the study of social mobility.

THOMAS A. DIPRETE is Associate Professor of Sociology at Duke University. His research interests concern the link between labor market structure and job mobility, and the impact of recent changes in the American and world economy on labor markets, employment practices, and work careers.

Appendix. A Brief Comparison of Ordered Logit and SOR Models

A binary logit model can be interpreted as a threshold model for some latent variable y^*. Assume that:

$$y_t^* = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t \tag{27}$$

Given this assumption, we can write

$$\text{Prob}(y_t^* > 0) = \text{Prob}(\varepsilon_t > -\mathbf{x}_t'\boldsymbol{\beta}) = 1 - F(-\mathbf{x}_t'\boldsymbol{\beta}) = F(\mathbf{x}_t'\boldsymbol{\beta}),$$

where F is the logistic distribution function. If we have a dichotomous dependent variable y, we might assume that we observe $y_t = 1$ when $y_t^* > 0$, and $y_t = 0$ otherwise. This assumption establishes a correspondence between the observed variable and the latent variable, and allows us to interpret components of β_k in terms of the effect of x_{tk} on y_t^* as well as in terms of the effect of x_{tk} on changes in $\text{Prob}(y_t = 1)$.

To generalize this model, assume that the underlying model is the same, but that we observe a polychotomous response. Suppose for simplicity that y_t is trichotomous. In the general case (Maddala 1983, p. 46), we might designate these threshold values as $-\mathbf{x}_t'\boldsymbol{\beta}_1$ and $-\mathbf{x}_t'\boldsymbol{\beta}_2$. Under these assumptions, both the latent variable y^* and the threshold values, which determine the mapping from the latent to the observed variable, vary across individuals. It follows that:

$$\text{Prob}(y_t = 1) = P_{t1} = \text{Prob}(y_t^* \le -\mathbf{x}_t'\boldsymbol{\beta}_1) = F(-\mathbf{x}_t'\boldsymbol{\beta}_1)$$

$$\text{Prob}(y_t = 2) = P_{t2} = \text{Prob}(-\mathbf{x}_t'\boldsymbol{\beta}_1 \le y_t^* \le -\mathbf{x}_t'\boldsymbol{\beta}_2) = F(-\mathbf{x}_t'\boldsymbol{\beta}_2) - F(-\mathbf{x}_t'\boldsymbol{\beta}_1)$$

$$\text{Prob}(y_t = 3) = P_{t3} = \text{Prob}(y_t^* > -\mathbf{x}_t'\boldsymbol{\beta}_2) = 1 - F(-\mathbf{x}_t'\boldsymbol{\beta}_2) = F(\mathbf{x}_t'\boldsymbol{\beta}_2)$$

The "global" logits (Agresti 1984, p. 19), $\log((P_{t2} + P_{t3})/P +_{t1})$ and $\log(P_{t3}/(P_{t1} + P_{t2}))$, have simple forms under this model:

Appendix Table 1. Variable Definitions

Variable	Definition
Professional/ managerial	Census Codes 3-37, 43-199
Technical/clerical	Census Codes 203-235,303-389
Sales	Census Codes 243-285
Service	Census Codes 403-469
Crafts	Census Codes 503-699
Other blue-collar	Census Codes 703-899
Race	White = 1; Nonwhite = 0
Education	Highest year of education attended
College degree	Education ≥ 16
Experience	Age – Education – 5 in the previous year
Gender	Male = 1; Female = 0

$$\log((P_{t2} + P_{t3})/P_{t1}) = \mathbf{x}_t'\boldsymbol{\beta}_1$$

$$\log(P_{t3}/(P_{t1} + P_{t2})) = \mathbf{x}_t'\boldsymbol{\beta}_2 \tag{28}$$

In contrast to these simple forms, the "local" logits such as $\log(P_{t2}/P_{t1})$, or $\log(P_{t3}/P_{t2})$ are more complicated. For example,

$$\log(P_{t2}/P_{t1}) = \log\left(\frac{e^{\mathbf{x}_t'\boldsymbol{\beta}_1} - e^{\mathbf{x}_t'\boldsymbol{\beta}_2}}{1 + e^{\mathbf{x}_t'\boldsymbol{\beta}_2}}\right) \tag{29}$$

The usual simplification of the ordered logit model, which reduces the number of parameter vectors in the trichotomous case from two to one, is to assume that :

$$-\mathbf{x}_t'\boldsymbol{\beta}_2 = -\mathbf{x}_t'\boldsymbol{\beta}_1 + c.$$

This constraint does not materially simplify equation 29, but it does lead to a more parsimonious form for 28, namely

$$\log((P_{t2} + P_{t3})/P_{t1}) = \mathbf{x}_t'\boldsymbol{\beta}_1$$

$$\log(P_{t3}/(P_{t1} + P_{t2})) = \mathbf{x}_t'\boldsymbol{\beta}_1 - c$$

The assumption $-\mathbf{x}_t'\boldsymbol{\beta}_2 = -\mathbf{x}_t'\boldsymbol{\beta}_1 + c$ amounts to assuming that the relationship between x_t and the odds does not depend upon the cutpoint.[a] Because the odds that y_t^* lies above vs. below any cutpoint are proportional to $\mathbf{x}_t'\boldsymbol{\beta}_1$ in the standard simplification of the ordered logit model, McCullagh (1980) referred to this standard form as the "proportional odds" model.[b]

In contrast, mobility models based on the multinomial logit model always yield simple forms for "local" logits of the form $\log(P_{tj}/P_{tj}')$, for example,

$$\log\frac{P_{t2}}{P_{t1}} = \mathbf{x}_t'(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)$$

while "global" logits of the form $\log(\sum_{j\ge j'} P_{tj}/\sum_{j<j'} P_{tj})$ are more complicated:

$$\log(P_{t3}/(P_{t1} + P_{t2})) = \log\left(\frac{e^{\mathbf{x}_t'\boldsymbol{\beta}_3}}{1 + e^{\mathbf{x}_t'\boldsymbol{\beta}_2}}\right)$$

In the multinomial logit model, there is no underlying latent dependent variable. The outcome categories are assumed to have an objective existence — they are not artifacts of the method for measuring the dependent variable. The SOR simplification assumes that covariates have parallel effects on outcomes (in other words, one might assume a latent intervening variable between the exogenous variables and the dependent variable) and uses this assumption to reduce the number of parameters in the model. This contrast — the assumption of latent intervening variables in a model for "local" logits vs. the assumption of proportional odds with a latent dependent variable in a model for "global" logits — best captures the difference between the simplifying assumptions of the SOR models considered in this paper and the more familiar ordered logit model.

[a] In other words, the odds-ratio for being in categories 2 or 3 vs. 1 for two individuals is the same as the odds ratio for being in categories 3 vs. 1 or 2 for the same individuals. This ratio depends only on the difference between their covariate vectors.

[b] This proportionality assumption is similar to the proportionality assumption in a proportional hazards model (McCullagh 1980). For an empirical example in which the proportional odds simplification of the general ordered logit model breaks down, see Peterson and Harrell (1990).

Appendix Table 2. Full-Time Workers Who Changed Jobs in the Previous Year But Remained with the Same Employer: 1983, 1987 January CPS

		Destination Occupation						
	Origin Occupation	Professional/ Managerial	Technical/ Clerical	Sales	Service	Crafts	Other blue-collar	Total
Males	Professional/ managerial	245	30	40	13	27	10	365
	Technical/clerical	58	74	21	8	27	32	220
	Sales	46	7	74	3	15	10	155
	Service	22	11	6	62	17	30	148
	Crafts	43	39	19	15	112	76	304
	Other blue-collar	37	42	31	30	117	233	490
	Total	451	203	191	131	315	391	1682
Females	Professional/ managerial	169	62	24	10	4	7	276
	Technical/clerical	145	368	34	11	13	21	592
	Sales	38	47	48	8	2	8	151
	Service	33	33	7	35	3	14	125
	Crafts	2	8	1	2	9	13	35
	Other blue-collar	5	29	13	10	17	95	169
	Total	392	547	127	76	48	158	1348

REFERENCES

Agresti, Alan. 1984. *Analysis of Ordinal Categorical Data*. New York: Wiley.

Aldrich, J. H. and F. D. Nelson. 1984. *Linear Probability, Logit and Probit Models*. Beverly Hills: Sage.

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

Berndt, E., B. Hall, R. Hall, and J. Hausman. 1974. "Estimation and Inference in Nonlinear Structural Models." *Annals of Economic and Social Measurement* 3:653-65.

Dennis, J. E., Jr. and Robert B. Schnabel. 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, N. J.: Prentice-Hall.

DiPrete, Thomas A. 1987. "Horizontal and Vertical Mobility in Organizations." *Administrative Science Quarterly* 32:422-44.

DiPrete, Thomas A. and Margaret L. Krecker. Unpublished. "Occupational Linkages and Job Mobility Within and Across Organizations." Manuscript, Department of Sociology, Duke University.

Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. NY: Academic.

Fienberg, Stephen E. 1980. *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA: Massachusetts Institute of Technology Press.

Goodman, Leo A. 1972. "Some Multiplicative Models for the Analysis of Cross-Classified Data." Pp. 649-96 in Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press.

________. 1979. "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 74:537-52.

________. 1987. New Methods for Analyzing the Intrinsic Character of Qualitative Variables Using Cross-Classified Data." *American Journal of Sociology* 93:529-83.

Grusky, David B. and Robert M. Hauser. 1984. "Comparative Social Mobility Revisited: Models of Convergence and Divergence in 16 Countries." *American Sociological Review* 49:19-38.

Haberman, Shelby J. 1974. "Loglinear Models for Frequency Tables with Ordered Classifications." *Biometrics* 30:589-600.

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

Hauser, Robert M., Shu-Ling Tsai, and William H. Sewell. 1983. "A Model of Stratification with Response Error in Social and Psychological Variables." *Sociology of Education* 56:20-46.

Hauser, Robert M. and David B. Grusky. 1988. Cross-National Variation in Occupational Distributions, Relative Mobility Chances, and Intergenerational Shifts in Occupational Distributions." *American Sociological Review* 53:723-41.

Hensher, David A. and Lester W. Johnson. 1981.

Applied Discrete-Choice Modelling. New York: Halsted Press.

Hodge, Robert W. 1981. "The Measurement of Occupational Status." *Social Science Research* 10: 396-415.

Hout, Michael. 1983. *Mobility Tables*. Beverly Hills: Sage.

________. 1984. "Status, Autonomy and Training in Occupational Mobility." *American Journal of Sociology* 89:1379-1409.

________. 1988. "More Universalism, Less Structural Mobility: The American Occupational Structure of the 1980s." *American Journal of Sociology* 93:1358-1400.

Logan, John A. 1983. "A Multivariate Model for Mobility Tables." *American Journal of Sociology* 89:324-49.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

McCullagh, Peter. 1980. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society*, Series B 42:109-42.

Miller, Paul and Paul A. Volker. 1985. "On the Determination of Occupational Attainment and Mobility." *Journal of Human Resources* 20:197-213.

Peterson, Bercedis and Frank E. Harrell, Jr. 1990. "Partial Proportional Odds Models for Ordinal Response Variables." *Applied Statistics* 39:205-17.

Raftery, Adrian E. 1986a. "Choosing Models for Cross-Classifications." *American Sociological Review* 51:145-46.

________. 1986b. "A Note on Bayes Factors for Log-Linear Contingency Table Models with Vague Prior Information." *Journal of the Royal Statistical Society*, Series B 48:249-50.

Sobel, Michael E. 1985. "Social Mobility and Fertility Revisited: Some New Models for the Analysis of the Mobility Effects Hypothesis." *American Sociological Review* 50:699-712.

U.S. Bureau of the Census. 1983. *Current Population Survey, January, 1983*, (including supplemental data) [MRDF]. U.S. Bureau of the Census [producer]. U.S. Department of Labor, Bureau of Labor Statistics [distributor].

U.S. Bureau of the Census. 1987. *Current Population Survey, January, 1987: Occupational Mobility and Job Tenure* (ICPSR 8913) [MRDF]. U.S. Bureau of the Census [producer], Ann Arbor: Inter-University Consortium [distributor].

Yamaguchi, Kazuo. 1983. "The Structure of Intergenerational Occupational Mobility: Generality and Specificity in Resources, Channels, and Barriers." *American Journal of Sociology* 88:718-45.