STATISTICAL MODELS OF EDUCATIONAL STRATIFICATION:
HAUSER AND ANDREW'S MODEL FOR SCHOOL TRANSITIONS

Robert D. Mare
University of California, Los Angeles
May 7, 2006

*Address correspondence to Robert D. Mare, Department of Sociology, UCLA, 264 Haines Hall, Box 951551, Los Angeles, CA 90095-1551; E-mail: mare@ucla.edu. This work was supported by the Russell Sage Foundation and benefited from the facilities of the California Center for Population Research, which is supported by the National Institute of Child Health and Human Development. In preparing this comment I benefited from conversations with Robert M. Hauser, Vida Maralani, William M. Mason, and Christine R. Schwartz.
Hauser and Andrew marry two good ideas that emerged within quantitative social science during the 1970s, namely the multiple-indicator-multiple-cause (MIMIC) model for multivariate responses and the logistic response model for transitions between stages of educational attainment. The MIMIC model, now part of standard textbooks for structural equation models with latent variables (e.g., Bollen, 1989), is a natural, rigorous, and efficient way to represent the effects of exogenous variables on multiple endogenous variables (Hauser and Goldberger 1971). It is also a tool for testing whether a set of variables "behave as a scale" in their relationship to other variables in a multivariate model and thus for avoiding ad hoc, unsubstantiated index construction (e.g., Hauser 1973). As Hauser and Andrew note, the logistic response model for school transitions is a widely used tool for the analysis of educational stratification, allowing investigators of family background effects on educational attainment to recognize that schooling is a sequence of events in time rather than a single status, and that the sources of inequality of educational opportunity and outcome may be different at different stages of schooling. It is gratifying to see that these ideas have continued to inspire application, discussion, and improvement approximately 30 years after they were introduced. In this comment, I place my contributions in the context in which they were conceived, note their continued potential for research on educational stratification, and discuss the strengths and weaknesses of Hauser and Andrew's contribution.

My studies of school transitions built upon approaches that already appeared in the literature and were motivated in part by contemporary arguments about how to interpret inequality of educational opportunities and outcomes. In a pioneering and meticulous
investigation, Beverly Duncan used school progression ratios to describe trends in educational attainment, an illuminating complement to her investigations of family background effects on years of school completed (Duncan 1965; 1967; 1968). Spady (1967) analyzed the associations of father's educational attainment and son's school transitions, using published tabulations and the comparatively primitive statistical methods that were available for such data in the 1960s. Relying largely on simulated data, Boudon (1974) formulated a model for inequality of educational opportunity that treated schooling as a sequence of grade progressions. Hauser (1976) criticized Boudon's approach on several grounds but, most relevant for the present discussion, showed the vulnerability of Boudon's and Spady's interpretations of differences in school progression rates to artifactual ceiling and floor effects. I applied the logistic response model for dichotomous dependent variables to the effects of family background on progression through school and showed the connections between this approach and others that were in widespread use (Mare 1977, 1979, 1980, 1981a). This work took advantage of recent innovations in statistical methods (Cox 1970; Allison 1982), improvements in high speed computing, and the newly available Occupational Changes in a Generation II Survey (Featherman and Hauser 1975).

From a conceptual standpoint, the significance of the model for school transitions lies mainly in viewing educational attainment as a process in time and allowing for rigorous empirical examination the effects of potentially time varying characteristics of individuals, families, and institutions as children and adolescents move through school (Mare 1981b; Lucas 2001; Mare and Chang 2006). This is true even though much of this work's influence has been on traditional studies that gauge inequalities among persons who are classified by relatively static "social background" characteristics and who live in a homogeneous institutional
environment (e.g., Shavit and Blossfeld 1993). School transition models also allow for distinct social processes to govern school progression at different stages, including the possibility that school progression results from the behavior of a multiplicity of actors who vary in importance across time, place, and institutional context. For example, Gamoran and Mare (1989) explored the effects of the competing objectives of families and schools in track placement and their consequences for students. In an independent line of work, Manski and Wise (1983) used variants of school transition models to show how college attendance and graduation result from the interdependent yet distinct actions of families, students, and college admissions officers. Breen and Jonsson (2000) very usefully generalized the school transition model to encompass a more complex institutional framework that includes multiple parallel pathways in education systems. Unlike these works, however, much of the most innovative recent research on educational attainment has adopted relatively asociological approaches in which educational stratification results from information processing and rational calculation by atomized families and individuals. This focus is shared by researchers who are both sympathetic to and critical of the school transition approach (e.g., Cameron and Heckman 1998; Breen and Goldthorpe 1997; Morgan 2005). It remains to be seen whether future researchers will return to a broader set of concerns about the multiple interdependent decision makers at various institutional levels who may be responsible for levels and variations in educational attainment.

Hauser and Andrew's family of models for proportionality restrictions is a significant contribution to the statistical analysis of school transitions within the framework of a discrete time hazard model (Allison 1982). They illustrate the value of proportionality restrictions on the coefficients of the school transition model, both as a way of achieving parsimony of parameters and also as a way of distilling the essence of a set of complex multivariate results. Although
Hauser and Andrew make a strong case for their LRPC and LRPPC models, I think that they understate the power of their family of models. At the same time, in other ways, they may oversell it as well. I develop these points in the balance of this comment.

A key issue in interpreting variation in the effects of social background across school transitions is the identifiability of parameters in binary response models. As we now understand much better than when my work on school transition models first appeared, in logit and other models for dichotomous dependent variables, the variance of the latent dependent variable and thus the scale of the estimated coefficients are not identified (Long 1997; Cameron and Heckman 1998; Allison 1999). To estimate these models, a normalizing constraint is required. In typical software for estimating binary response models, this constraint is to fix the variance of the errors of the equation at a constant, although other constraints, such as fixing the variance of the latent dependent variable itself are also possible (Winship and Mare 1983; 1984). With either type of constraint, within a single equation for a binary response, one can only identify the relative sizes of the coefficients. Across equations for the same dependent variable estimated on different samples, or for different dependent variables estimated on the same sample, one cannot even identify the relative sizes of the coefficients for the effects of a single variable.¹ By contrast, in typical binary response models, the effects of the covariates on the binary response, as measured either by derivatives (when they exist) or the differences in predicted probabilities, are identified and do not depend on a normalizing constraint (Long 1997; Allison 1999). Given the nonlinearity of the effects of the covariates on the response probability in logit, probit, and other models, however, these effects, unlike the coefficients, depend upon the specific point in the distribution of the covariates where they are evaluated.
This problem of identification arises both in unrestricted versions of binary response models for school transitions and also in constrained versions such as Hauser and Andrews' LRPC and LRPPC models. Hauser and Andrew find that my informal description of the unconstrained school transition model coefficients – that is, that the coefficients for socioeconomic background factors decline across transitions -- is consistent with the restrictions of their LRPPC model. Unfortunately, although the LRPPC and the unrestricted models of school transitions yield similar estimates, the coefficients of neither model confirm or disconfirm the conclusion that the effects of socioeconomic background decline across school transitions. Because the coefficients in the logistic response model are identified only up to a constant of proportionality within any given equation, true differences in effects of socioeconomic background across school transitions are empirically indistinguishable from heteroskedasticity of the conditional variances of the latent dependent variables.

The indeterminacy of the logistic regression coefficients raises the question of whether the validity of the proportionality restrictions in the LRPC or LRPPC models depends on the normalizing restrictions used to estimate the logistic coefficients. Fortunately, it is easy to show that this is not the case. That is, even though the logistic coefficients can only be estimated with an (arbitrary) normalizing constraint, the proportionality restrictions are independent of any specific constraint. For simplicity, in the following discussion I focus on the LRPC model and do not explicitly discuss the less restrictive LRPPC model. However, my conclusions about what one can learn from models that impose proportionality restrictions across equations in the school transition models apply equally to the LRPC and LRPPC models.2

If \( y_j \) denotes a binary variable that equals one if an individual makes the \( j \)th transition and zero otherwise, the \( J \) equation logit model for the effects of \( K \) fixed social background variables
is, omitting subscripts for individuals,

\[
\log \left[ \frac{p(y_j = 1 \mid y_{j-1} = 1, X_1, ..., X_k)}{1 - p(y_j = 1 \mid y_{j-1} = 1, X_1, ..., X_k)} \right] = b_{j0} + \sum_{k=1}^{K} b_{jk} X_k \quad (j = 1, ..., J; y_0 = 1) \quad (1)
\]

For this model, the LRPC hypothesis is that:

\[
\frac{b_{jk}}{b_{j-1,k}} = \frac{b_{jk'}}{b_{j-1,k'}} = \lambda_j \quad . \quad (j = 2, ..., J) \quad (2)
\]

Because the true variances of the underlying dependent variables in equations (1) are not identified, however, the true coefficients for the effects of the \(X_k\) are not identified. Equivalently stated, the constants of proportionality, \(\lambda_j\), are estimable, but their values incorporate both differences across equations in the effects of the regressors and also differences in the variances of the underlying dependent variables.

One can rewrite the logit model as a set of equations for latent continuous variables, say \(y_j^*\), as

\[
\frac{y_j^*}{\sigma_j} = \beta_{j0} + \sum_{k=1}^{K} \left( \frac{\beta_{jk}}{\sigma_j} \right) X_k + \frac{\epsilon_j}{\sigma_j} \quad (3)
\]

where \(\beta_{jk}\) denotes the true coefficient for the effects of the \(k\)th independent variable on the log odds of making the \(j\)th school transition and \(\sigma_j\) is the standard deviation of the random disturbance \(\epsilon_j\) for the equation for the \(j\)th transition. Neither the \(\beta_{jk}\) nor the \(\sigma_j\) are estimable from the data without additional information. They are, however, linked to the estimable quantities in equation (1) because

\[
b_{jk} = \frac{\beta_{jk}}{\sigma_j} \quad (4)
\]
If the LRPC constraints (2) hold for equation (1), then the LRPC model implies that
\[
\frac{\left( \frac{\beta_{j,k}}{\sigma_{j}} \right)}{\left( \frac{\beta_{j-1,k}}{\sigma_{j-1}} \right)} = \left( \frac{\beta_{j,k'}}{\sigma_{j}} \right) = \lambda_{j}, \quad (j = 2, ..., J). \tag{5}
\]

But this expression simplifies to
\[
\frac{\beta_{j,k}}{\beta_{j-1,k'}} = \frac{\beta_{j,k'}}{\beta_{j-1,k'}} = \lambda_{j} \left( \frac{\sigma_{j}}{\sigma_{j-1}} \right) = \lambda'_{j}, \tag{6}
\]

which shows that if the LRPC model holds for the estimable quantities \( b_{jk} \), it also holds for the true underlying parameters \( \beta_{jk} \). Although the constant of proportionality that links the true parameters for the different transitions for a given regressor (6) is not, in general, the same as for the estimable parameters (2), the LRPC hypothesis for the true parameters holds if and only if it holds for the estimable parameters as well.

Although this is an appealing result inasmuch as it implies that the LRPC hypothesis is not bedeviled by the identification problem that plagues the logistic response model itself, it may also be construed as a hollow one because the coefficients of proportionality \( \lambda'_{j} \) that link the equations for the different transitions are not identified. Fortunately, however, it is possible to draw stronger and more useful conclusions when the LRPC or LRPPC model holds. As noted above, although the coefficients of binary response models are not identified without an arbitrary scale restriction, other functions of the data are estimable, including functions of the predicted probabilities under the models. In particular, the effects of the regressors as represented by the partial derivatives of the probabilities with respect to the regressors (when they exist) or differences in adjusted predicted probabilities between different values of the regressors are
estimable. This raises the question of whether the restrictions of the LRPC model on the
coefficients of the logistic response model imply similar restrictions on the estimable functions
of the predicted probabilities. In fact, this is the case. To see this, consider the logistic response
model in equation (1) and assume, without loss of generality, that the derivatives of the
probabilities with respect to the $X_k$ exist. These derivatives are

$$\frac{\partial}{\partial X_k} \left[ p(y_j = 1 | y_{j-1} = 1, X_1, \ldots, X_K) \right] = b_{jk} \left[ 1 - p(y_j = 1 | y_{j-1} = 1, X_1, \ldots, X_K) \right]$$

(7)

where $f_j$, the density function of the standard logistic distribution, depends on the estimated
parameters for the $j$th transition and a selected vector of values for $X_1, \ldots, X_K$ but not on the
specific regressor $X_k$ under consideration. Thus, the ratio of effects of a given regressor across
two adjacent transitions evaluated at a common vector of values of the regressors is

simply $b_{jk} f_j / b_{j-1,k} f_{j-1}$. If the LRPC restriction (2) holds for the estimated coefficients, then it
holds for the derivatives too. That is,

$$\frac{b_{jk} f_j}{b_{j-1,k} f_{j-1}} = \frac{b_{jk} f_j}{b_{j-1,k} f_{j-1}} = \lambda_j \left( \frac{f_j}{f_{j-1}} \right)$$

(8)

Conversely, if (8) holds, (2) holds as well.

Thus the LRPC model implies a common hypothesis about the estimable coefficients, the
unidentified true coefficients, and the estimable effects of the independent variables evaluated at
a chosen vector of values of the regressors. This is an especially attractive property of the model,
which may not be shared by other restricted versions of the model for school transitions. When a
LRPC or LRPPC model fits the data, therefore, it is not only a pleasingly parsimonious model as Hauser and Andrew emphasize, but also deeply and robustly informative about the structure of association in school transition models.

Despite the great potential of the LRPC and LRPPC models, Hauser and Andrew overstate the case when they say that one ought to estimate the LRPC before proposing more complex explanations of change in the effects of socioeconomic background. Whether or not the LRPC model fits the data may or may not rule out specific explanations and, depending on which covariates obey the proportionality assumptions, the model may or may not yield an interpretable result. The LRPC and LRPPC models are elegant and sometimes adequate summaries of multivariate relationships. But explanations and interpretations of variation in the effects of social background may occur to an investigator before, during, or after fitting less or more parsimonious specifications of transition-specific effects and may be tested by a variety of alternative models. My own early efforts were influenced by the OCG educational stratification studies of Featherman and Hauser (Hauser and Featherman 1976; Featherman and Hauser 1978), which used relatively unparsimonious specifications of trends in social backgrounds effects. Within the constraints of additive, linear models, these studies described changes in the effects of specific background variables and, in yielding key descriptive statistics about trends in stratification, set the stage for subsequent thinking about the causes of these trends. When one turns to more parsimonious models, LRPC and LRPPC are just one of many possible classes of models. Another, as Hauser and Andrew indicate, is the model with no variation across transitions in the effects of some or all social background variables. The family of probability models for cumulative educational attainment is another (e.g., Breen 2005; Cameron and Heckman 1998). Yet another is models in which any or all background effects vary
systematically with the known values of macro-level covariates, such as the proportion of a cohort at risk to a given transition or characteristics of school systems and the economy (Mare 1977; 1981b; Rijken 1999). The LRPC and LRPPC models are very useful devices for studying educational stratification. But in a model-rich world, the investigator needs a flexible toolkit and some guiding substantive hypotheses to see what is going on. No single class of models or sequence of model tests is likely to do the trick.
ENDNOTES

1 When the error variance is fixed it is also inappropriate to make within sample comparisons among the coefficients for a given covariate across equations with varying subsets of covariates. In this case, the total variance of the latent dependent variable and thus the scale of the estimated coefficients vary from model to model as a function of the different regressors that are included. Fixing the variance of the latent dependent variable avoids this problem. It does not, however, avoid the problems of comparison across samples and across dependent variables.

2 My discussion rules out the case discussed by Allison (1999), who shows that one can compare the coefficients of binary response models across samples when one assumes that the coefficient for at least one regressor is invariant across samples. This identifying restriction may be defensible in many substantive investigations, but, without additional information, is hard to maintain a priori in studies of school transitions where one of the main analytic objectives is to describe variation in social background effects across school transitions.

3 A similar result to the one presented here for the effects of continuous $X_k$ holds for the adjusted predicted probabilities for different values of discrete regressors.
REFERENCES


