

Published in final edited form as:

*Res Soc Stratif Mobil.* 2011 September 1; 29(3): 343–349. doi:10.1016/j.rssm.2011.04.001.

## VALUES AND LIMITATIONS OF STATISTICAL MODELS

Yu Xie\*

University of Michigan

### Abstract

Methodological consequences of population heterogeneity for the sequential logit model in studies of education transitions are now well understood. There are two main mechanisms by which heterogeneity may cause biases to parameter estimates in sequential logit models: outcome incommensurability and population incommensurability. These methodological problems are intrinsic to the substantive research question and thus are not easily remediable with better statistical models. All statistical solutions require extra information in the form of additional data or additional assumptions. In some settings, the researcher may explicitly introduce a form of heterogeneity into the sequential logit model and then evaluate the model. In other settings, the researcher may wish to stay with the conventional sequential logit model and interpret the results descriptively.

---

In the little thinking I do these days about the old battles I fought, it has increasingly seemed to me that one of two or three cardinal problems that social science has not yet come to grips with is precisely this issue of heterogeneity ... The ubiquity of heterogeneity means that for the most part we substitute actuarial probabilities for the true individual probabilities, and therefore we generate mainly descriptively accurate but theoretically empty and prognostically useless statistics.

--Otis Dudley Duncan

The above statement is from a letter Duncan wrote to me back in 1996 (Xie 2007, p.154). Otis Dudley Duncan was a population thinker and was keenly aware of individual-level heterogeneity in a human population.<sup>1</sup> To him, even the best statistical models were no more than crude devices that summarize aggregate patterns and regularities in observed data. The presence of population heterogeneity meant that the social researcher would never understand theoretically meaningful causal mechanisms at the individual level. For this reason, despite his unrivalled achievements in and contributions to quantitative sociology in his prime years, Duncan became rather pessimistic toward the end of his career, around the mid-1980s, regarding the true value of statistical models in social science (Xie 2007).

Duncan was way ahead of his time. Much has changed since the mid-1980s. Indeed, most important advances in social science methodology since then have been in response to the real concern underlying Duncan's pessimism: the problem of heterogeneity. Extending Duncan's own interest in the Rasch model (Duncan 1982), sociological and psychological

---

© 2011 Elsevier Ltd. All rights reserved

\*Direct all correspondence to Yu Xie (yuxie@umich.edu) at Population Studies Center, Institute for Social Research, 426 Thompson Street, University of Michigan, Ann Arbor, MI 48106.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

<sup>1</sup>For discussions on population thinking, see Duncan (1984), Lieberman and Lynn (2002), Mayr (1982, 2001), and Xie (2000, 2007).

methodologists have developed various statistical tools to uncover group-level heterogeneity, including multilevel models (Raudenbush and Bryk 1986), growth-curve models (Muthén and Muthén 2000), and latent class models (D'Unger et al. 1998) – all now under the rubric of mixed models (Demidenko 2004). In the literature on causal inference, a literature Duncan first helped pioneer but later became skeptical of (Xie 2007), the new emphasis is now explicitly on population heterogeneity (see, for example, Brand and Xie 2010; Heckman 2005; Holland 1986; Morgan and Winship 2007; Winship and Morgan 1999). Furthermore, and perhaps most significantly, micro economists, led by James Heckman, have revealed the real consequences to us of population heterogeneity in conventional statistical models and provided various solutions to this problem (e.g., Heckman 1979; 2001).

James Heckman, along with coauthor Stephen V. Cameron (1998), brought to light the potential harm of population heterogeneity in Mare's (1980) classic sequential logit model in studying schooling transitions at different school levels. Cameron and Heckman's critique has somewhat shaken sociologists' confidence in their substantive findings in a very large social stratification literature on educational outcomes that is based on Mare's model (e.g., Shavit and Blossfeld 1993). However, sociologists of social stratification have not remained idle. Instead, they have worked hard to understand the potential problems caused by heterogeneity in such models and sought solutions that might help them overcome these problems. The articles published in this special issue of *Research in Social Stratification and Mobility* represent the heroic efforts on behalf of the social stratification community in response to Cameron and Heckman's challenge.

In the remainder of this essay, I will discuss: (1) the pragmatic values of Mare's model and statistical models in general, (2) the epistemological limitations of Mare's model and statistical models in general, and (3) some key themes and insights in the papers published in this special issue.

## Values of Statistical Models

There are three broad reasons for using statistical models: causation, prediction/smoothing, and description (Powers and Xie 2008, p.12). Of the three, causation is of the highest scientific value but the most difficult to accomplish. In the past three decades or so, much criticism has been launched against statistical models as tools for uncovering causality (see Freedman, 1987; Morgan and Winship 2007). We will discuss this in more detail in the next section.

The other two reasons are less lofty but more practical. Researchers sometimes use statistical models simply for prediction/smoothing even though they do not understand the causal mechanisms in the observed data. All that is required is that observed data exhibit some type of regular pattern, for whatever reason, that seems to hold true statistically. For example, statistical models have been used extensively in the literature dealing with missing data (Little and Rubin 2002).

It is also legitimate to resort to statistical models as descriptive summaries of statistical information. Some researchers may feel that this third reason is too basic to have scholarly merit. I disagree. To me, descriptive statistical models are both necessary and desirable.

I am of the opinion that social scientists should focus their attention on understanding the empirical social world. At least, knowing the empirical facts about the social world cannot be a bad thing for social scientists. In Lieberman's (1992, p.2) words,

One of the great contributions of sociology is its ability to bring information to bear on topics of interest and basic concern to society... [e.g., statistics about income differences between blacks and whites, rape, poverty, homelessness, and intergenerational mobility] I select these questions because their answers provide useful information about society even when they sometimes are not couched in theoretical terms ... (By the way, who said that knowing some facts is such a bad idea?)

In this context, we may view results from statistical models as descriptions of the empirical world. Ironically, population heterogeneity, which makes statistical models unreliable as devices for discovering causality (a topic to be discussed in more depth in the next section), necessitates statistical modeling as a descriptive device (Xie 2005), because population heterogeneity makes it fruitless to describe individual phenomena at the individual level. Rather, we need statistical tools to summarize diverse social phenomena as they are observed. In education research, for example, it is not enough to note that some people have completed college educations while others have not. Rather, it is more useful to know what proportion of a given population has completed college – a summary statistic.

When we break such summary statistics down by social attributes, such as cohort, gender, family background, we essentially build a descriptive statistical model. Although in practice, we may add some modeling constraints such as additivity or linearity, these constraints are mainly introduced for convenience and ease of interpretation. The results in descriptive models are merely summaries of statistical information in observed data. As such, they are no more -- and no less -- than observed facts, to be interpreted, not discredited, by the researcher.

The descriptive interpretation of statistical models is also desirable because such models are closely tied to the empirical world and thus are less subject to debates as to their truthfulness. As I will discuss in the next section, all causal interpretations of statistical results involve unverifiable assumptions. Observed data alone never give us enough information to allow us to draw causal inferences. Results from descriptive models are, in fact, observed information, albeit disguised in mathematical form. They cannot, by definition, be wrong in themselves (except in situations of bad data or bad algorithms) and do not require any assumptions for their validity. Therefore, descriptive models are desirable because, having no unverifiable assumptions, they are indeed the only thing we can study as social scientists. At a minimum, they provide the foundation from which we can advance to higher levels of knowledge. Here, I wish to draw a distinction between an “observable” (meaning calculable) descriptive statistic that may be subject to alternative interpretations versus a quantity of interest (or estimand) that is intended to quantify a causal relationship (and thus incalculable without assumptions).

However, the call for a descriptive interpretation of the Mare model should not be taken too far. While any well-defined summary statistic, as a logit coefficient in the Mare education transition model, is always “correct” descriptively, its interpretation does not need to be, and often is not. An example is the Pearson correlation coefficient. As a preliminary step of data analysis, we often compute such coefficients as simple, summary statistics about pairwise linear relationships among variables, but we seldom jump to conclusions about causal relationships on the basis of such observed correlations. An appropriate interpretation of “descriptively accurate” (to quote Duncan) summary statistics is far more difficult than most researchers realize. A great deal depends on the concrete research setting, particularly the research question.

Let us now return to the educational transition question studied by the Mare model. As an illustration, I borrow the example of two transitions in the U.S. from Lucas, Fucella, and

Berends (2011): (1) whether a person completes high school or not ( $H = 1$  if yes,  $H = 0$  if no), and (2) whether a person enters college or not ( $C = 1$  if yes,  $C = 0$  if no). The unconditional probability of college completion can be decomposed into the product of the following two component probabilities:

$$P(C=1) = P(H=1) P(C=1|H=1). \quad (1)$$

We can also add pre-existing covariates  $X$  to equation (1) so that the decomposition is at the subgroup level (defined by the values of  $X$ ):

$$P(C=1|X=x) = P(H=1|X=x) P(C=1|H=1, X=x). \quad (2)$$

As a general rule, it is always possible to decompose the marginal probability at a later point in the life-cycle to be the product of earlier transition probabilities, separately for sub-populations (Xie 1996). We can always do this as a descriptive device, without any assumptions.

Thus, one fruitful interpretation of Mare's model is simply descriptive. Note that logit coefficients from the Mare model, for example, are essentially log-odds-ratios in a multivariate context and as such can be interpreted as descriptive measures of associations even when their causal interpretations may be ambiguous. Furthermore, Mare's transition model allows researchers to focus on separate components along the path to the destination probability of the highest level of education. For this interpretation, parameter estimates of Mare's consequential logit model should be converted back to transition probabilities (see also Lucas, Fucella, and Berends 2011). However, this interpretation is most defensible when  $X$  is preexisting, or time-invariant, rather than time-varying.

## Limitations of Statistical Models

Description cannot be the only research goal. As Heckman (2005) and Pearl (2009) argue, understanding causality should be the ultimate goal of social science, as in other branches of science. To satisfy this intellectual yearning for a higher level of understanding, a large "new" literature on causal inference using statistical methods has emerged in recent decades (Heckman 2005; Holland 1986; Manski 1995; Pearl 2009; Rubin 1974; Winship and Morgan 1999). A key lesson from the literature, however, is the extreme difficulty of reaching this grand goal with observational data that are typically at the disposal of social scientists.

Suppose that a population,  $U$ , is being studied. Let  $Y$  (either  $H$  or  $C$  in equation (1)) denote an outcome variable of interest (its realized value being  $y$ ) that is a function for each member in  $U$ . Let us define treatment as an externally induced intervention that can, at least in principle, be given to or withheld from a unit under study. For simplicity, we consider only dichotomous treatments and use  $D$  to denote the treatment status (its realized value being  $d$ ), with  $D=1$  if a member is treated and  $D=0$  if a member is not treated. For studies of educational attainment, one possible treatment variable of this kind is whether or not parents were married when the respondent was a child. Let subscript  $i$  represent the  $i$ th member in  $U$ . We further denote  $y_i^{d=1}$  as the  $i$ th member's potential outcome if treated (i.e., when  $d_i=1$ ), and  $y_i^{d=0}$  as the  $i$ th member's potential outcome if untreated (i.e., when  $d_i=0$ ). With this counterfactual framework, we can conceptualize a treatment effect as the difference in potential outcomes associated with different treatment states for the *same* member in  $U$ :

$$\delta_i = y_i^{d=1} - y_i^{d=0}, \quad (3)$$

where  $\delta_i$  represents the hypothetical treatment effect for the  $i$ th member. The fundamental problem of causal inference (Holland 1986) is that, for a given unit  $i$ , we observe either  $y_i^{d=1}$  (if  $d_i=1$ ) or  $y_i^{d=0}$  (if  $d_i=0$ ), but not both.

Given population heterogeneity, different members in  $U$  have different values of  $y^{d=1}$ ,  $y^{d=0}$  and thus  $\delta$ . Causal inference at the individual level is therefore impossible. Causal inference at the group level is possible, but only under strong assumptions. For example, we may assume that the potential outcomes  $(y^{d=1}, y^{d=0})$  are independent of treatment, i.e.,

$$(y^{d=1}, y^{d=0}) \perp\!\!\!\perp D, \quad (4)$$

where  $\perp\!\!\!\perp$  denotes independence. Under this assumption, we can compare the averages in outcome between treated units and untreated units to derive the population average treatment effect.

The assumption of equation (4) is satisfied in experiments when the researcher randomly assigns subjects to treatment and control conditions. In general, however, this is an implausible assumption for observational data. To overcome this difficulty, researchers have relied on the collection of additional covariates so as to hope that the potential outcomes  $(y^{d=1}, y^{d=0})$  are independent of treatment conditional on a set of observed preexisting covariates ( $X$ ), changing equation (4) to:

$$(y^{d=1}, y^{d=0}) \perp\!\!\!\perp D|X. \quad (5)$$

Equation (5) is called the “ignorability” assumption. Various statistical methods for estimating causal effects with observation data relies on the ignorability assumption (e.g., Brand and Xie 2010; Holland 1986; Morgan and Winship 2007). The other main approach makes use of instrumental variables that affect the outcome *only* indirectly through the treatment variable (e.g., Angrist and Pischke 2009; Heckman, Urzua, and Vytlacil 2006). Both approaches impose strong and unverifiable assumptions.

Note that in the standard setup for causal inference described above, the researcher is interested in estimating an average effect of treatment on a given outcome in a given population. For this simple and well-defined problem, causal inference is already intractable without a strong assumption, the situation is much worse for the objective in a typical research setting that employs the Mare model: the comparison of treatment effects across successive transitions.

In such a research setting, the researcher wishes to compare the treatment effects on probabilities of two transitions, which can be defined as

$$\delta_H = E \left[ P(H=1)^{d=1} - P(H=1)^{d=0} \right], \quad (6a)$$

$$\delta_{cH} = E \left[ P(C=1|H=1)^{d=1} - P(C=1|H=1)^{d=0} \right]. \quad (6b)$$

Equations (6a) and (6b) are written in terms of a linear probability model. In the Mare model, they are written in terms of log-odd ratios, expressed as logit coefficients. A main point, however, is that the research objective is much less tractable here, because  $\delta_H$  are  $\delta_{cH}$  not strictly comparable, for two reasons: (1) the outcomes being compared are different, and (2) the populations being compared are different. Let me call the first problem “outcome incommensurability,” and the second problem “population incommensurability.” Both problems are discussed by Cameron and Heckman (1998).

Outcome incommensurability means that the outcome of interest changes across transitions so that it is not possible to compare effect sizes without additional assumptions. In earlier equations, we use  $P(H = 1)$  to denote the probability of high school graduation and  $P(C = 1 | H = 1)$  the probability of entering college given high school completion. Substantively, they are different outcomes. Comparing the magnitude of group differences in these outcomes requires their conversion into a common scale. A conversion scale imposes an assumption about how they should be compared. In previous work that employs the Mare model, the logit scale is assumed. As pointed out by Cameron and Heckman (1998), it is not clear why the logit scale is preferred over other possible scales (such as the linear probability scale in equations (6a) and (6b)). Also, it is well known that logit coefficients (or probit coefficients) are implicitly normalized within an equation and thus are identifiable up to an arbitrary scale only within an equation (Powers and Xie 2008, p.44). In brief, comparison of logit coefficients across equations with different outcome variables is fraught with potential difficulties.

Population incommensurability means that over successive transitions, the population at risk for making the next transition is no longer comparable to the population at risk making an earlier transition, as it changes in systematic ways that that may cause biases. Education transition models are essentially survival models. In the presence of unobserved population heterogeneity, even when heterogeneity is initially uncorrelated with observed covariates, the survival dynamics generate over time the correlation between unobserved heterogeneity and observed covariates (Cameron and Heckman 1998). This is because observations that leave the system are truncated, with the truncation likelihood driven in part by unobserved heterogeneity (Vaupel and Yashin 1985). This induced correlation, in turn, causes biases in the estimates for later transitions.

Given these two forms of incommensurability, outcome incommensurability and population incommensurability, it is no wonder that the sequential logit model fails badly in the presence of heterogeneity. Instead of asking why the model fails, perhaps we should rephrase the question and ask why we even ask the model to do what we wish it to do: to provide a comparison of a causal effect over successive transitions. The data in a typical research setting are simply too thin to allow us to entertain the question with any confidence. Any sharp conclusion would then require the use of additional information, i.e., assumptions.

## Contributions of the Papers

It is a relief, and indeed reassuring, that sophisticated researchers in sociology today understand these methodological issues in the Mare model that were raised by Cameron and Heckman in 1998. This point is well illustrated in the collection of papers being published in this special issue. The collection indicates that quantitative sociologists in general, and



social stratification researchers in particular, have followed methodological advances in other fields and are attending to methodological issues in their work. Taken together, the papers make a valuable contribution to the methodological discussion on Mare's school transition model. Lessons learned from them are many and varied, ranging from substantive/empirical to methodological.

Of the five papers published here, the paper by Lucas, Fucella, and Berends (2011) is the most substantive and empirical. In a sense, Lucas and his coauthors are true followers in the empiricist tradition espoused by Lieberman (1992). Their main message is loud and simple: potential methodological difficulties posed by heterogeneity should not stand in the way of analyzing educational transitions. Their main remedy to the methodological difficulties is empirical: rich longitudinal data with time-varying covariates that allow the researcher to use parametric models for modeling stage-specific selective transitions. Substantively, their study of multiple cohorts in the U.S. data does not show a waning pattern as reported in many earlier studies.

The parametric solution of Lucas, Fucella, and Berends (2011) is echoed as the major theme in the paper by Holm and Jaeger (2011). The basic idea is to conceptualize the two transition outcomes (as in equation (1)) in two separate probit regression models with correlated errors, with the likelihood of the second transition conditional on making the first transition. Two additional assumptions are needed for identification. First, a parametric (bivariate normality) assumption is needed for the error terms. Second, instrumental variables that only affect the first transition but not the second transitions directly need to be assumed. The result is the bivariate probit selection model. Holm and Jaeger demonstrate the usefulness of the model under a wide range of conditions, with both simulated and real data.

The paper by Karlson (2011) is an ambitious undertaking. It attempts to add a heterogeneity component to multinomial logit models for multiple-outcome sequential transitions. Instead of assuming a heterogeneity term with a parametric assumption, Karlson introduces latent classes representing heterogeneity. One nice byproduct is that this approach allows for the violation of the IIA (independence of irrelevant alternatives) assumption that is implicit in the standard multinomial logit model. In practice, this approach is implemented in Stata `gllamm` ("generalized linear latent and mixed model"). Karlson illustrates the usefulness of the approach in a study of the educational careers of a Danish cohort.

The paper by Buis (2011) provides an excellent discussion of the methodological issues caused by population heterogeneity. Instead of finding a concrete fix, which would inevitably commit the researcher to a particular assumption about the data, Buis suggests instead a more general approach -- sensitivity analysis - that is modeled after Rosenbaum (1992). When applied to a study of educational attainment in the Netherlands, the approach yields an interesting substantive finding: whereas the waning pattern in the effect of father's education over transitions is sensitive to potential biases caused by the presence of heterogeneity, the temporal decline in the effect of father's education across cohorts is not.

Of all the authors for this special issue, Tam (2011) appears to be the most faithful follower of Cameron and Heckman. In a series of extensive and carefully designed simulations, Tam provides evidence for preferring the Cameron-Heckman model over the Mare model. Closely following Cameron and Heckman, Tam resorts to a simple latent class specification for population heterogeneity. Going beyond the latent class approach, however, Tam also asks whether the presence of either good or poor indicators of latent classes would help, for both the Cameron-Heckman model and the Mare model. An interesting result of Tam's study is that whereas a strong indicator appears to be similarly helpful in both models, a poor

indicator is helpful only in the Cameron-Heckman model. Based on these results, Tam favors the Cameron-Heckman model over the traditional Mare model.

## Conclusion

One essential feature common to all social and behavioral phenomena is population heterogeneity across different units of analysis. Inspired by the work of Cameron and Heckman (1998), we now have a better understanding of the methodological consequences of population heterogeneity for Mare's (1980) sequential logit model for the study of education transitions. The papers published in this issue all make important contributions towards this understanding.

There are two main mechanisms by which heterogeneity causes biases in sequential logit models. First, across successive transitions, the researcher essentially deals with different outcomes and thus encounters the problem of outcome incommensurability. Second, across successive transitions, units with higher latent likelihood of continuation are selectively included with a higher proportion than for a previous transition in the population, creating a problem of population incommensurability. Both mechanisms can cause serious biases to results from the conventional sequential logit model ignoring heterogeneity.

While not all statistical models are useful, all statistical models have limitations. A better knowledge of the methodological pitfalls of the sequential logit model does not by itself negate the use of such models in empirical settings. Rather, it pushes us to think harder about its use, or the use of alternative models. The reality is that there is no fool-proof method to the heterogeneity problem. Even though we could speculate on potential biases caused by unobserved heterogeneity, it is difficult to account for it empirically. By definition, unobserved heterogeneity is unobserved and thus always remains elusive. All statistical solutions require extra information in the form of either additional data or additional assumptions. What the researcher actually should do depends on the research setting. In some settings, the researcher may explicitly introduce a form of heterogeneity into a statistical model and evaluate such a model. In other settings, the researcher may still wish to stay with the conventional sequential logit model and interpret the results descriptively.

## Acknowledgments

Financial support for this research was provided by the National Institutes of Health, Grant 1 R21 NR010856-01. I am grateful to Kristian Karlson and Tony Tam for his helpful comments on an earlier draft of this paper.

## References

- Angrist, Joshua D.; Pischke, Jorn-Steffen. *Mostly Harmless Econometrics*. Princeton University Press; Princeton, NJ: 2009.
- Brand, Jennie E.; Xie, Yu. Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education. *American Sociological Review*. 2010; 75(2):273–302. [PubMed: 20454549]
- Buis, Maarten. The consequences of unobserved heterogeneity in a sequential logit model. 2011 This issue.
- Cameron, Stephen V.; Heckman, James J. Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*. 1998; 106:262–333.
- Demidenko, Eugene. *Mixed Models: Theory and Applications*. Wiley; 2004.
- Duncan, Otis Dudley. *Rasch Measurement and Sociological Theory*. Hollingshead Lecture, Yale University; 1982.



- Duncan, Otis Dudley. *Notes on Social Measurement, Historical and Critical*. Russell Sage Foundation; New York: 1984.
- D'Unger AV, Land Kenneth C. McCall PL, Nagin Daniel S. How Many Latent Classes of Delinquent/criminal Careers? Results from Mixed Poisson Regression Analyses. *American Journal of Sociology*. 1998; 103:1593–1630.
- Freedman, David .A. As Others See Us: A Case Study in Path Analysis. *Journal of Educational Statistics*. 1987; 12:101–128.
- Heckman, James J. Sample Selection Bias as a Specification Error. *Econometrica*. 1979; 47:153–61.
- Heckman, James J. Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy*. 2001; 109:673–748.
- Heckman, James J. *The Scientific Model of Causality*. University of Chicago; 2005. Unpublished manuscript
- Heckman, James; Urzua, Sergio; Vytlačil, Edward. Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*. 2006; 88:389–432.
- Holland, Paul W. Statistics and Causal Inference. *Journal of American Statistical Association*. 1986; 81:945–70. with discussion.
- Holm, Anders; Jæger, Mads Meier. Dealing with Selection Bias in Educational Transition Models: The Bivariate Probit Selection Model. 2011 This issue.
- Karlson, Kristian Bernt. Multiple paths in educational transitions: A multinomial transition model with unobserved heterogeneity. 2011 This issue.
- Lieberman S, Lynn FB. Barking Up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science. *Annual Review of Sociology*. 2002; 28:1–19.
- Little, RJ.; Rubin, DB. *Statistical Analysis with Missing Data*. 2nd Edition. John Wiley; New York: 2002.
- Lucas, Samuel R.; Fucella, Phillip N.; Berends, Mark. A Neo-Classical Education Transitions Approach: A Corrected Tale for Three Cohorts. 2011 This issue.
- Manski, Charles. *Identification Problems in the Social Sciences*. Harvard University Press; Boston, MA: 1995.
- Mare, Robert D. Social Background and School Continuation Decisions. *Journal of the American Statistical Association*. 1980; 75:295–305.
- Mayr, Ernst. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Harvard University Press; Cambridge, MA: 1982.
- Mayr, Ernst. The Philosophical Foundations of Darwinism. *Proceedings of the American Philosophical Society*. 2001; 145(4):488–495. [PubMed: 11894859]
- Morgan, Stephen L.; Winship, Christopher. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press; New York: 2007.
- Muthén B, Muthén LK. Integrating Person-centered and Variable-centered Analyses: Growth Mixture Modeling with Latent Trajectory Classes. *Alcoholism-Clinical and Experimental Research*. 2000; 24(6):882–891.
- Pearl, Judea. *Causality: Models, Reasoning, and Inference*. Second Edition. Cambridge University Press; New York: 2009.
- Powers; Xie. *Statistical Methods for Categorical Data Analysis*. Second Edition. Emerald; Howard House, UK: 2008.
- Raudenbush S, Bryk AS. A Hierarchical Model for Studying School Effects. *Sociology of Education*. 1986; 59(1):1–17.
- Rosenbaum, Paul. R. *Observational Studies*. 2nd Edition. Springer; New York: 2002.
- Rubin, Donald B. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*. 1974; 66:688–701.
- Shavit, Yossi; Blossfeld, Hans-Peter, editors. *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Westview Press; Boulder, CO: 1993.
- Tam, Tony. Accounting for Dynamic Selection Bias in Educational Transitions: The Cameron-Heckman Latent Class Estimator and Its Generalizations. 2011 This issue.

- Vaupel, James W.; Yashin, Anatoli I. Heterogeneity ruses - Some surprising effects of selection on population dynamics. *American Statistician*. 1985; 39:176–185. [PubMed: 12267300]
- Winship, Christopher; Morgan, Stephen L. The Estimation of Causal Effects From Observational Data. *Annual Review of Sociology*. 1999; 25:659–707.
- Xie, Yu. A Demographic Approach to Studying the Process of Becoming a Scientist/Engineer. In: National Research Council. , editor. *Careers in Science and Technology: International Perspective*. National Academy Press; Washington, D.C: 1996. p. 43-57.
- Xie, Yu. Demography: Past, Present, and Future. *Journal of the American Statistical Association*. 2000; 95:670–673.
- Xie, Yu. Methodological Contradictions of Contemporary Sociology. *Michigan Quarterly Review*. 2005; 44:506–511.
- Xie, Yu. Otis Dudley Duncan's Legacy: the Demographic Approach to Quantitative Reasoning in Social Science. *Research in Social Stratification and Mobility*. 2007; 25:141–156.