Most studies of persistence behavior use path analysis or ordinary least squares regression to estimate unknown coefficients. However, estimates produced by these techniques are biased if selectivity bias contaminates choices made by individuals in the data sample. We explain this problem, argue that it is present in data samples used in persistence studies, and discuss an alternative estimation technique that controls for it. The methodology and the differences in the interpretation of coefficient estimates are illustrated with a data sample of individual students at a single university.

Most of the recent studies of persistence decisions of undergraduate students are based on Tinto's (1975) theoretical approach, which postulates a temporal framework that flows from precollege characteristics to social and academic integration in higher education to the decision to continue one's education. Hence, researchers employing multivariate statistical techniques have considered persistence as a function of precollege characteristics such as ability, family income, parental education, sex, and characteristics of the high school attended, plus measures of academic and social integration that include grades earned, intellectual development, and participation in extracurricular activities. The academic and social integration variables are themselves functions of precollege characteristics in Tinto's model, and most researchers specify a recursive structural model that includes equations to explain these variables as well. Since both the direct and indirect effects of the exogenous variables may be of policy interest, Tinto suggested using path analysis to estimate the parameters. A large fraction of recent persistence studies have used path analysis, although a few have used ordinary least squares (OLS).

William C. Weiler, Office of Planning & Management, Stanford University, Building 310, Stanford, CA 94305. Daniel J. Pierro, Management Information Division, University of Minnesota.

This much is well known to those familiar with or engaged in persistence studies. However, estimation by either path analysis or OLS has two shortcomings that can be addressed by an alternative methodological approach. The first is that some of the dependent variables in a model's equations are dichotomous, the persistence variable itself being the most obvious example. Neither OLS nor path analysis is appropriate in this context, although this has been noted by researchers who have used those estimation techniques anyway, usually arguing they are suitable if the fraction of persisters in the data sample is neither too low nor too high.

The second, more important, factor is that persistence and the variables measuring the concepts of academic and social integration into higher education may both be functions of such inherently unmeasurable individual characteristics as the taste for college. In this case the integration variables can be correlated with the error term in the persistence regression, which means that both OLS and path analysis will produce biased estimates of the effects these and all other variables have on persistence.

In this paper we illustrate an alternative to path analysis or OLS regression that can solve both of these problems. To motivate this alternative, we first discuss in more detail the problem in interpreting path analysis or OLS estimates of the coefficients in the persistence equation. This is done in the next section in the context of a specific academic integration measure, the initial attendance status of a student (full- or part-time). Previous research suggests that this particular variable has a significant effect on persistence, but there are two conflicting explanations of how these effects could have been obtained, one of which implies that the significant effect does not actually reflect a true causal influence of this status on persistence.

The interpretation problem is the result of potential selection bias. While this issue has not been considered in previous persistence studies, it has been discussed by researchers studying other topics in education. In the second section we define the concept of selection bias, discuss two such topics, and show how the problem of selection bias faced in those studies applies to persistence analyses in general.

In the final section we illustrate how a variant of the estimation methodology used to control for selection bias in these other studies in education can be adapted to persistence studies. The example used is the one discussed in the first section, the impact of initial enrollment status on persistence. While the data do not include the full range of background variables employed in some other persistence studies, the results show how the alternative methodology can be applied to persistence studies and highlight the insights on the true causal relationships between academic integration and persistence that can be obtained from the technique.