Errors-in-Variables Regression: Why Stata’s -eivreg- is Wrong and What To Do Instead

J.R. Lockwood, Daniel F. McCaffrey and Corey B. Savage
Eduational Testing Service
Errors-in-Variables Regression: Why Stata’s -eivreg- is Wrong and What To Do Instead

Background

Applied educational researchers often want to estimate statistical models involving quantities that are not directly observable. For example, it is often necessary to adjust for prior achievement differences across students with different educational experiences (e.g., curricula, teachers, programs) when using observational data to estimate the effects of those experiences on student outcomes. The data typically available for such adjustment are test scores from standardized assessments, which measure achievement with error (Lord, 1980). Ignoring measurement error can bias parameter estimators from statistical models. There is growing acknowledgment of the importance of this issue in the educational research community (Battauz, Bellio, & Gori, 2011; Battauz & Bellio, 2011; Lockwood & McCaffrey, 2014; Shang, Van Iwaarden, & Betebenner, 2015), and regression models with corrections for measurement error in standardized test scores are being applied in high-profile educational policy studies (e.g., Isenberg et al., 2013).

Statistical Framework

To make the ideas concrete, suppose that an analyst is interested in estimating the effect of some treatment using nonexperimental data, and assumes the model:

\[ Y_i = \alpha + \theta T_i + Z_i \beta_0 + X_i \beta_1 + \epsilon_i, \]

where \( i = 1, \ldots, n \) indexes units (e.g., students), \( Y_i \) is an outcome of interest, \( T_i \) is a treatment indicator with true effect \( \theta \), \( Z_i \) is a vector of observed covariates assumed to be measured without error, and \( X_i \) is a vector of unobserved (latent) covariates measured with error by observed proxies \( W_i \). For example, \( Y_i \) may be a current-year achievement test score, \( T_i \) may be an indicator of whether a student participated in a voluntary after-school tutoring program, \( Z_i \) may include student background characteristics such as demographic and economic status information, and \( X_i \) may be latent achievement attributes prior to program participation (e.g. from the previous school years) that are measured with prior test scores \( W_i \).

Because \( X_i \) is unobserved, the parameters of Model (1) cannot be estimated directly. Moreover, it is well-known that plugging \( W_i \) in place of \( X_i \) and running the usual regression leads to estimated regression coefficients, including \( \hat{\theta} \) for the treatment indicator, that are generally biased and inconsistent (see, e.g., Fuller 2006). A common method for consistent estimation is the so-called “errors-in-variables (EIV)” regression estimator, which uses information about the magnitude of the measurement error in \( W_i \) to adjust the matrices used in the computation of the estimated regression coefficients to produce estimators \( (\hat{\alpha}, \hat{\theta}, \hat{\beta}_0, \hat{\beta}_1) \) that consistently estimate \( (\alpha, \theta, \beta_0, \beta_1) \) under appropriate assumptions. The texts by Fuller (2006) and Carroll et al. (2006) provide additional details and historical context. The information about the magnitude of the measurement error in \( W_i \) may be available in a variety of forms, depending on the setting. For standardized test scores, common forms include either published test reliabilities, average standard errors of measurement, or conditional standard errors of measurement.
Objective of Presentation and Relevance to SREE Community

The EIV estimator is likely to be of increasing interest to both methodologists and applied researchers in the SREE community, given the large number of applications that rely on modeling with test scores and other error-prone measures (e.g., evaluations of classroom instruction), and the growing acknowledgement of the need to account for those errors to avoid bias. The paper supporting the SREE presentation, which is underway and will be available by the conference, will provide the following information:

1. The EIV estimator can be motivated from classical method-of-moments considerations as well as more modern structural equation modeling perspectives. As such, it can be computed in numerous routines available in R, Stata, SAS and Mplus, and these routines provide a variety of methods for estimating standard errors for the EIV estimator. The paper will provide a summary of the common estimation methods for both the regression parameters and standard errors.

2. The paper will provide syntax for implementing the EIV estimator in three packages in R, two functions in Stata, PROC CALIS in SAS, and Mplus.

3. The paper will discuss the different forms in which measurement error information may be available, and the implications of these differences for estimating standard errors of the EIV estimator.

4. The paper will provide simulation results demonstrating the performance of different software routines and their associated standard error estimators as a function of sample size, reliability, what information is known about the magnitude of the measurement error, and various distributional specifications.

The conference presentation will motivate and introduce the EIV estimator, and then focus on a few substantive results of key interest to the SREE community:

1. What information is known about the magnitude of the measurement error can affect the performance of standard error estimates provided by different software routines. The presentation will provide guidance about which combinations of routines and standard error estimation options are best suited to different circumstances.

2. The Stata -eivreg- routine would probably be used to implement EIV regression by many members of the SREE community. The presentation will demonstrate that -eivreg- generally provides estimated standard errors that underestimate true standard errors. The underestimation can be severe in some circumstances, leading to spuriously short confidence intervals with inadequate coverage properties. The presentation will argue that Stata users should instead use the -sem- routine, and will discuss options for doing so.

The presentation is appropriate for the conference theme of “Expanding the Toolkit: Maximizing Relevance, Effectiveness and Rigor in Education Research” given its focus on tools for improving inferences from observational data, and its message that a software routine in common usage should be passed over in favor of more effective alternatives.
References


