Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Econometrics 117 (2003) 151–178
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables

Harley Frazis[a,*], Mark A. Loewenstein[b]

[a] *Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Suite 4945, Washington, DC 20212, USA*
[b] *BLS, 2 Massachusetts Ave. NE, Suite 4130, Washington, DC 20212, USA*

Accepted 6 February 2003

Abstract

This paper is concerned with mismeasured binary explanatory variables in a linear regression. Modification of a technique in Hausman et al. (J. Econometrics 87 (1998) 239) allows simple computation of bounds under relatively weak assumptions. When one has instruments, we show how to obtain consistent parameter estimates using GMM. We show how to incorporate the estimated measurement error bounds into the GMM estimates, and we develop a specification test based on the compatibility of the GMM estimates with the measurement error bounds. When the mismeasured variable is endogenous, the IV estimate and the measurement error bounds can be used to bound its coefficient.

JEL classification: C21; C31

Keywords: Measurement error; Binary variables; Instrumental variables

1. Introduction

This paper is concerned with mismeasured binary explanatory variables in a linear regression. We obtain results that allow us to improve on existing estimators under different assumptions about the extent of prior information. We examine three different cases: (1) the mismeasured variable is assumed exogenous, and no instruments are available; (2) the mismeasured variable is assumed exogenous, and one or more instruments are available; and (3) the mismeasured variable is not assumed exogenous, and

* Corresponding author. Fax: +1-202-691-7425.
E-mail addresses: frazis_h@bls.gov (H. Frazis), loewenstein_m@bls.gov (M.A. Loewenstein).

0304-4076/03/$ - see front matter
doi:10.1016/S0304-4076(03)00121-0

instruments are available. In the first two cases, we derive bounds or point estimates under weaker assumptions on prior information than in the previous literature, and in the third case—which has not to our knowledge been analyzed—we also derive bounds.

In case 1, the traditional approach to the measurement error problem is to use auxiliary information on the measurement error process. More generally, one may not have good point estimates of the measurement error parameters, but may nevertheless be able to bound them—using a validation study, for example. Bollinger (1996) shows how these bounds can in turn be used to bound the regression coefficient.

Recently Hausman et al. (1998) (hereafter HAS) have developed a technique that allows the analyst to bound the measurement error process by estimating the parameters from the data, without information from validation studies. We modify their technique to allow simple computation of bounds without functional form assumptions. Combining the estimated measurement error bounds with the OLS coefficient yields bounds on the true effect of the mismeasured explanatory variable.

When instruments are available, as in case 2, instrumental variable (IV) estimation is another common method of dealing with measurement error. However, recent research has shown that IV estimation is upwardly biased when the mismeasured variable is binary (Loewenstein and Spletzer, 1997; Berger et al., 2000; Kane et al., 1999) because measurement error in this case must be negatively correlated with the true value. Berger et al. (2000) (BBS hereafter) and Kane et al. (1999) (KRS hereafter) show that when one has two erroneous measures, one can obtain a consistent estimate using a generalized method of moments (GMM) technique.

Two distinct measures of the same variable are not commonly available. We thus extend the analysis in BBS and KRS to the case where the second measure is replaced by one or more instruments. We provide a closed-form solution for the GMM parameter estimates. We also show how to incorporate the estimated measurement error bounds into the GMM estimates, and we develop a specification test of the measurement error model based on the compatibility of the GMM estimates with the measurement error bounds.

Lastly, we show that the GMM technique is not easily extended to the case where the mismeasured variable is endogenous. However, the IV estimate and estimates of the measurement error bounds can be used to bound the effect of the mismeasured variable, analogous to the OLS case without endogeneity.

The paper is organized as follows. Section 2 outlines the model, summarizes the HAS technique, and introduces our extension of HAS. Section 3 then turns to the case where one or more instruments exist for the mismeasured variable. Section 4 considers the case where there are available instruments, but the mismeasured binary variable is endogenous. Section 5 presents an empirical example looking at the returns to on-the-job training and Section 6 concludes.

2. Bounding the effect of a mismeasured binary explanatory variable when no instrument is available

Our model is:

$$Y_i = c + X_i\gamma + \beta T_i^* + e_i \tag{1}$$