Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Econometrics 133 (2006) 51–70
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Bounding parameters in a linear regression model with a mismeasured regressor using additional information

Yingyao Hu*

Department of Economics, University of Texas at Austin, 1 University Station C3100, BRB 1.116, Austin, TX 78712, USA

Accepted 21 March 2005
Available online 17 May 2005

Abstract

This paper discusses a linear regression model with a mismeasured regressor in which the measurement error is correlated with both the latent variable and the regression error. We use a linear structure to capture the correlation between the measurement error and the latent variable. This paper shows that the variance of the latent variable is very useful for revealing information on the parameters which otherwise cannot be obtained with such a nonclassical measurement error. The main result is that the finite bounds on the parameters can be found using the variance of the latent variable, regardless of how severely the measurement error and the regression error are correlated, if the mismeasured regressor contains enough information on the latent one. This paper also discusses the special but interesting case of the latent variable being dichotomous. In this case, the mean of the latent variable may even reveal information on the correlation between the measurement error and the regression error. All the bounds developed in the paper are tight.

JEL classification: C10; C20

Keywords: Nonclassical measurement error; Dichotomous latent regressor

*Tel.: +1 512 475 8556; fax: +1 512 471 3510.
E-mail address: hu@eco.utexas.edu.
URL: http://www.eco.utexas.edu/~hu/.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.03.009

1. Introduction

The measurement error model has increasingly been a topic of interest among researchers who want to estimate economic parameters such as the return to schooling and the union wage differential. When a regressor is mismeasured in a linear regression model, the least-squares estimator is generally not consistent, but at least some information can be inferred about the true parameters from the inconsistent estimators. These types of results are in the form of bounds on the parameters, which will hold asymptotically. Under the classical assumption that the measurement error is independent of the latent regressor and the regression error, it is well known that the regressions of x on y and y on x provide asymptotic bounds on the coefficient on x in the one-regressor case (Gini, 1921). However, the problem is more complicated in a multi-regressor context, and the existence of bounds is limited to certain cases. The classical result in the area is due to Koopmans (1937), who shows that such a generalization is possible only under very restrictive conditions. Patefield (1981) and Klepper and Leamer (1984) present a similar result. When further information on the measurement error distribution, such as bounds on the error variance, is available, narrower bounds on the parameters can be found (Bekker et al., 1984). Similar types of bounds are also discussed in Leamer (1982, 1987) and Klepper (1988b).

While the classical measurement error has been studied intensively, nonclassical measurement error has drawn more and more attention from researchers in recent decades. Bekker et al. (1987) discuss the case of errors in regressors and the regression error being correlated. Iwata (1992) and Krasker and Pratt (1986, 1987) show that bounds on these correlations may help find bounds on parameters of interest. Erickson (1993) provides a neat result when the measurement error is independent of the latent regressor but correlated with the regression error. As for empirical evidence of the nonclassical measurement error, Rodgers et al. (1993) suggest that the measurement error may be correlated with the latent variable. Bound et al. (2001) also find that the assumption that the measurement error is independent of the latent variable is strong and often implausible.

This paper discusses a linear measurement error model in which the measurement error is correlated with both the latent variable and the regression error. Let y denote the dependent variable, x^* denote the latent regressor and w denote the row vector of the other regressors (excluding the constant). Let α, β and γ be the intercept, the regression coefficients of x^* and w respectively, where γ is a column vector with the same dimension as w. Let u stand for the regression error. The linear regression model is as follows:

$$y = \alpha + \beta x^* + w\gamma + u \tag{1}$$

with $\mathrm{E}(u|x^*, w) = 0$. The researcher observes another variable x together with y and w as the proxy of the latent variable x^*. A critical assumption in this paper is that the conditional mean of the measurement error $v = x - x^*$ is linear in the latent