Journal of Econometrics 185 (2015) 95–109

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Semiparametric estimation of models with conditional moment restrictions in the presence of nonclassical measurement errors ☆

Suyong Song

Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 870, 3210 N. Maryland Ave., Milwaukee, WI 53211, United States

ARTICLE INFO

Article history:
Received 3 February 2012
Received in revised form
20 February 2014
Accepted 20 October 2014
Available online 6 November 2014

JEL classification:
C14
C20

Keywords:
Semiparametric conditional moment restrictions
Method of sieve
Endogeneity
Nonclassical measurement error
Instrumental variable

ABSTRACT

This paper develops a framework for the analysis of semiparametric conditional moment models with endogenous and mismeasured causes, which is of empirical importance. We show that one set of valid instruments is sufficient to control for both endogeneity and measurement errors of the causes of interest, which has been observed in linear parametric models. Two-step consistent estimators of the parameters of interest are proposed. We also show that the proposed estimators are consistent with a rate faster than $n^{-1/4}$ under a certain metric, and the proposed estimators of the finite-dimensional unknown parameters obtain root-n asymptotic normality. Monte Carlo evidences show that the proposed estimators perform well under a variety of identification conditions. An application to instrumental variables estimation of Engel curves illustrates the usefulness of our method. It supports that correcting for both endogeneity and measurement errors on total expenditure is substantial in estimating economically meaningful Engel curves.

1. Introduction

Semiparametric conditional moment restrictions have been used for a variety of economic models such as models of strategic interaction, dynamic game, auctions, asset pricing, investment equation, and consumer demand systems. By the nature of economic decision making, many economic models are subject to endogeneity. Furthermore, it is common that economic variables contributing to the economic systems are often unobserved to empirical researchers. As a result, addressing these issues simultaneously is crucial to identify true economic relationship between economic variables. This paper develops a framework for the analysis of semiparametric conditional moment models with unobserved endogenous causes, which is of empirical importance.

We consider the following models defined by conditional moment restrictions,

$$E[\rho(Z, \theta_0, h_0(\cdot)) \mid X] = 0, \tag{1}$$

where $Z \equiv (Y', X_1')'$, $Y \equiv (Y_1, Y_2')'$ is a vector of endogenous (or dependent) variables, X_1 is a subset of conditioning variables $X \equiv (X_1', X_2')'$, $\rho(\cdot)$ is a vector of generalized residual functions whose functional forms are known up to the unknown vector of finite-dimensional parameters θ_0 and the unknown functions $(h_0 \equiv (h_{01}(\cdot), \ldots, h_{0q}(\cdot)))$, where the arguments of each function $h_{0\ell}(\cdot)$, $\ell = 1, \ldots, q$, may depend on different arguments, and, in particular, may depend on Y. $E[\rho(Z, \theta_0, h_0) \mid X]$ is the conditional expectation of $\rho(Z, \theta_0, h_0)$ given X. Classical model of conditional moment restrictions without the unknown functions h_0 has been exploited considerably in the literature on nonlinear parametric models (see e.g. Hansen, 1982; Chamberlain, 1987; Newey, 1990). There have also been a lot of work on more general frameworks including the unknown function h_0 in the literature on nonparametric and semiparametric models (see e.g. Robinson, 1988; Powell et al., 1989; Chamberlain, 1992; Ichimura, 1993). In the seminal papers, Newey and Powell (2003) and Ai and Chen (2003) study method of sieves when the unknown functions h_0 depend on the endogenous variables. To be specific, they approximate the unknown functions h_0 by sieves, and apply the method of minimum

☆ This is a revised version of part of the author's Ph.D. dissertation. He is deeply grateful to Halbert White for his encouragement and extensive support. The author would like to thank a co-editor, an associate editor, and two referees for helpful suggestions and comments that substantially improve the quality of the paper. He would also like to thank Don Andrews, Xiaohong Chen, Gordon Dahl, Antonio Galvao, Chuan Goh, Bruce Hansen, Joel Horowitz, Kyoo il Kim, Arthur Lewbel, Susanne Schennach, Yixiao Sun, Elie Tamer, seminar participants at various universities, the 2011 Econometric Society North American Summer Meeting, and the 2011 Asian Meeting of Econometric Society for valuable discussions.

E-mail address: songs@uwm.edu.

http://dx.doi.org/10.1016/j.jeconom.2014.10.007
0304-4076/

distance to estimate parameters of interest. Ai and Chen (2003) establish that estimators of the parameters of interest $\alpha_0 \equiv (\theta_0, h_0)$ are consistent with a rate faster than $n^{-1/4}$, and that the estimator of the parametric component θ_0 is $\sqrt{n}$ consistent, asymptotically normally distributed, and efficient, while Newey and Powell (2003) characterize sufficient identification conditions and propose consistent estimators for the parameters of interest.

The main deviation of our setup from the literature is that the models (1) encompass the case where the true Y_2, causes of interest, are unobserved. For these models, the conventional methods in Newey and Powell (2003) and Ai and Chen (2003) are not applicable since they assume that the true endogenous causes are observed to researchers. In this paper, we propose a two-step estimation procedure addressing both issues, endogeneity and unobservability of Y_2 which is inside the infinite-dimensional functions, h_0. Sources of unobservability of causes can be thought of as unobserved heterogeneity, missing data, rounding errors, or measurement errors in reporting. For illustration purposes, this paper primarily focuses on measurement errors as a source of unobservability, although they do not exhaust all possibilities. In the first step, consistent estimates of the true unobserved conditional densities of endogenous variables given conditioning variables, which are possibly masked by nonclassical measurement errors, are obtained. In order to uniquely identify the true conditional densities, we adopt a method proposed by Hu and Schennach (2008), which relies on the eigenvalue–eigenfunction decomposition of an integral operator associated with joint densities of observables, and extend their method to allow for the presence of a vector of additional observable regressors. We propose sieve maximum likelihood estimators of conditional densities associated with the unobserved endogenous causes of interest. We then propose sieve minimum distance estimators of parameters of interest α_0 in the second step.

There have been few works which simultaneously resolve both endogeneity and measurement errors imposed on the same variable of interest in nonparametric and semiparametric models, despite there being a number of empirical observations where endogenous variables are also mismeasured. In the returns-to-education literature, for instance, education, the cause of interest, is endogenous in that it is correlated with unobserved ability which is an unobservable driver of income, dependent variable. Moreover, there is often erroneous reporting on education due to the nature of survey data. In linear parametric models, the use of valid instruments could resolve issues of identification and estimation associated with both endogeneity and measurement errors (e.g., Geraci, 1977; Hausman, 1977). However, the existence of valid instruments is not sufficient for the identification and estimation of parameters in nonlinear models with measurement error, even in the absence of endogeneity, as demonstrated by Griliches and Ringstad (1970), Amemiya (1985), and Hsiao (1989). As a result, correcting for both endogeneity and measurement errors in nonparametric and semiparametric models is not straightforward and difficult to deal with.

Chen et al. (2005) consider the problem of nonclassical measurement errors[1] in unconditional moment restrictions which allow for nonlinearity and nonsmoothness. Their models are simpler than ours in the sense that they do not include the infinite-dimensional unknown functions, h_0, in the unconditional moment restrictions. As a result, variables associated with the finite-dimensional parameters, θ_0, are measured with errors in their models. Furthermore, they require the existence of an auxiliary data set containing correctly-measured observations for the identification of parameters. Song et al. (2011) examine a method to control for both endogeneity and measurement errors on regressors in more general nonseparable models. They do, however, require two different sets of variables from the outside of the structural models to get the identification of the parameters of interest; they control for endogeneity by using instruments and resolve issues associated with measurement errors by using repeated measurements of true regressors. Measurement errors in their models need to be classical in that the measurement errors must have zero mean conditional on the true regressors or must be independent of them. Our identification result for the conditional densities of endogenous variables given conditioning variables in the first step extends Hu and Schennach (2008). Nevertheless, we depart from their models by allowing Y_2, causes of interest, to be endogenous. Thus, their parameters of interest are the conditional densities in the first step of our models, but we are primarily interested in the second-step parameters α_0.

In this paper, we contribute in several main directions to the literature on nonparametric and semiparametric models with endogeneity or measurement errors. *First*, under some mild conditions, we show that one set of valid instruments are sufficient to simultaneously resolve endogeneity and measurement errors in semiparametric models. Major idea behind the result is similar to the case in linear parametric models which require one set of instruments to be uncorrelated with both unobserved causes of the dependent variable and measurement errors in endogenous regressors. This is especially useful in empirical works because there are lots of evidences that obtaining two different sets of valid instruments for endogeneity and measurement errors, respectively, is not an easy task. Since our models are nonlinear, nevertheless, we need to use methods in the literature on nonparametric measurement errors in order to control for measurement errors, instead of using the conventional instrumental variable approach. *Second*, we show that the sieve estimators of α_0 are consistent with a rate faster than $n^{-1/4}$ under a certain metric, and the sieve estimators of the finite-dimensional unknown parameters θ_0 are $\sqrt{n}$ consistent and asymptotically normally distributed. The findings are interesting because the asymptotic properties of the proposed estimators are the same as those in Ai and Chen (2003), even though our models additionally allow for unobservability of the true endogenous regressors. *Third*, we show in an empirical application that failure of controlling for both endogeneity and measurement errors in total expenditure makes it difficult to estimate economically meaningful Engel curves. There have been two directions in the analysis of Engel curves; one is endogeneity in total expenditure and the other is measurement errors on it. However, there have been no published works which consider both issues simultaneously in nonparametric or semiparametric models. We find that estimated Engel curves from our methods are different than those from Blundell et al. (2007) who assume total expenditure is endogenous but correctly measured in semiparametric models. Moreover, we allow for measurement errors in total expenditure to be nonclassical (so it does not need to have zero mean conditional on the true total expenditure). Since nonclassical measurement errors have been observed in lots of survey data (e.g. see Bound and Krueger, 1991; Bound et al., 1994; Bollinger, 1998, for income data), this type of flexibility is useful in empirical works.

The rest of the paper is organized as follows. We first motivate the question by revisiting two interesting economic models and discuss identification results in Section 2. Issues on the estimation are discussed and asymptotic properties of the proposed estimator are analyzed in Section 3. In Section 4, the finite-sample properties of the proposed estimator are investigated via Monte Carlo studies,

[1] In the absence of endogeneity, there is a large literature in econometrics and statistics studying measurement errors in nonlinear models. For instance, classical measurement errors models include Hausman et al. (1991), Chesher (1991), Hausman et al. (1995), Hsiao and Wang (2000), Taupin (2001), Li (2002), Hong and Tamer (2003), Schennach (2004, 2007), Hu and Ridder (2012). Nonclassical measurement errors models include Mahajan (2006), Lewbel (2007), Hu (2008), Hu and Schennach (2008). See Chen et al. (2011) for an excellent review.