Journal of Econometrics 200 (2017) 151–153

Contents lists available at ScienceDirect

Journal of Econometrics

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

Editorial

Measurement error models: Editors' introduction

1. Introduction

Early studies on measurement errors in the econometric literature started with the so-called classical measurement error, where the errors are usually assumed to be independent of the true values, arguably because the measurement error models were borrowed from the relevant statistical literature, where the independence assumption is quite reasonable when the measurement error is caused by using an instrument to measure a certain property of an object. The additivity and independence in the classical measurement error models lead to many important and fruitful results. In the econometric literature, the classical measurement error framework is adopted mainly for the parsimony of the measurement error part of the model and for the convenience of using existing results. In empirical macroeconomics and some applied microeconomic research, the classical measurement error framework is usually embedded into linear models, such as factor models, linear dynamic models, and linear panel data models. In microeconometrics, identification and estimation of nonlinear models, such as nonlinear regressions and limited dependent models, with classical measurement errors, had been a difficult problem for many years.

In recent years, econometricians have been leading the studies on the nonclassical measurement error model because of the need of handling measurement errors in economic survey data, where the measurement errors are usually caused by self-reporting behaviors. Such a need exists in most disciplines in social sciences. Instead of measuring certain properties of an object, many economic data are from surveys, where interviewees self-report their information. The classical measurement error assumption is unlikely to hold in these scenarios. Econometricians are, therefore, on the frontier of identification and estimation of the so-called nonclassical measurement errors models, where the errors may be correlated with the latent true values. In particular, the presence of nonclassical measurement errors makes the identification of nonlinear models containing the latent true values extremely difficult, that is, whether the models can be uniquely determined from the joint distribution of observed variables.

Based on conditional independence assumptions, which widely exist in economic theories, a recent breakthrough in the measurement error models literature has been the realization that the joint distribution of three observables may uniquely determine the joint distribution of four variables including the three observables and the latent variable. Hu (2008) uses a matrix eigenvalue–eigenvector decomposition to show this pathbreaking result for the case where the latent variable is a general discrete variable. The Hu–Schennach Theorem in Hu and Schennach (2008) nontrivially extends this result to the general continuous case using a unique representation of bounded linear operators. In addition, one of the three observables may contain as few information as a binary indicator. Such an identification result is nonparametric and global and leads to a closed-form estimation procedure in the discrete case. The flexibility of these results greatly extend applications of measurement error models to various areas in empirical economic research. In his review paper in this special issue, Hu (2017) provides a simple definition of measurement and organizes the existing technical results in terms of the number of measurements. In addition, he shows that these technical results may not only apply to measurement error models, but also many economic models with latent variables.

This special issue on measurement errors intends to present recent development in the measurement error literature in terms of methodologies, for both classical and nonclassical models, and their applications. Below we discuss the papers in this issue in three loosely-defined groups: technical results for classical measurement error models, technical results for nonclassical models, and applications of measurement error models. There must be many topics which we may not cover in this extensive literature. We refer to more complete review works, including Wansbeek and Meijer (2000), Bound et al. (2001), Chen et al. (2011), Carroll et al. (2012), and Schennach (2016).

In the classical measurement error setting, Meijer et al. (2017) present consistent estimation of linear panel data model using three sources of moment conditions, such as restrictions on the covariance matrix of the errors in the equations, third moments of the regressors, and heteroskedasticity and nonlinearity in the relation between the error-ridden regressor and other error-free regressors. Gospodinov et al. (2017) consider another class of parametric linear dynamic models, so-called the autoregressive distributed lag models, with measurement errors. They exploit the relation between the parameters of the model and the least squares biases and propose a minimum distance estimator when external instruments may not be available or are weak. Garcia and Ma (2017) develop a root-n consistent and efficient estimator for parametric nonlinear regression models with mismeasured covariates, where the classical measurement errors follow a parametric distribution with unknown variances and covariances. They form the estimator under possibly incorrect working distribution models for the model error, error-prone covariate, or both.

Ben-Moshe et al. (2017) show nonparametric point identification of a classical measurement error model, where the nonlinear regression function is additively separable in terms of observed

http://dx.doi.org/10.1016/j.jeconom.2017.06.001

and mismeasured regressors. Their results can be extended to the case where the observed and latent regressors interact through a polynomial. This allows for rich interactions between the variables, at the expense of introducing a parametric restriction. Their identification proofs are constructive, and so can be used to form estimators. Chesher (2017) investigates the effect of classical measurement errors on quantile regressions. Using a small-variance approximation, Chesher shows how the error-contaminated and error-free quantile regression functions are related. The approximation shows how the two quantile regression functions are related, a key factor being the distribution of the error-free covariate. The results can be extended to the case where the measurement errors are non-additive and, therefore, nonclassical.

In empirical research using self-reported data, the measurement errors are usually nonclassical, i.e., correlated with the true values. Hahn and Ridder (2017) use control variables to estimate nonlinear parametric models with nonclassical measurement errors. Their estimator also applies to the classical case and is consistent even if the true value is endogenous. They also derive the influence function of the semi-parametric estimator that accounts for the estimation of the control variable in the first stage. Lee et al. (2017) investigate a dynamic linear panel regression model with a nonclassical error, where a proxy variable equals a linear function of the latent variable of interest plus an independent error. They consider the panel data estimation whose time dimension (T) is not small and comparable to the cross-sectional dimension (N), characterize its asymptotic bias due to many IVs, and derive its limiting distribution under the alternative asymptotics. Davezies and Barbanchon (2017) consider regression discontinuity when the running variable is observed with continuous measurement error. They propose a consistent nonparametric estimator of the local average treatment effects (LATE) using an auxiliary validation sample where the joint distribution of the noisy and the true running variables are observed. They apply the method to estimate the effect of receiving unemployment benefits.

In the discrete case where the latent true value and its measurement share the same discrete support, the measurement error is inherently nonclassical and is called misclassification error. Bollinger and van Hasselt (2017) present a Bayesian analysis of a regression model with a binary covariate that may have misclassification error. The Bayesian approach adds assumptions in the form of priors on the unknown misclassification probabilities, and may be considered as intermediate between the frequentist bounds of previous literature and strong assumptions which achieve point identification. They focus on how varying amounts of information contained in a prior distribution on the misclassification probabilities change the posterior of the parameters of interest. While the priors add information to the model, they do not necessarily tighten the identified set, but are sufficient to tighten Bayesian inferences. Meyer and Mittag (2017) study the bias caused by the misclassification of a binary dependent variable in a regression model, where the misclassification errors can be correlated with both observables and unobservables. They use validation data to show that the proposed bias formulas are accurate in finite samples and imply a tendency to attenuation. In addition, they also examine the bias from misclassification in empirical studies of food stamp take-up using two validation datasets and several methods to account for misclassification.

Besides the development in methodology, this special issue also contains several important applications of measurement error models. Chen et al. (2017) consider measurement errors in extended Roll models in finance and propose new methods for identifying the bid–ask spread and the distribution of the latent true fundamental price increments from observed transaction prices alone. Their methods are based on the characteristic function approach and hence do not assume the existence of any finite moments of the observed price increments. They establish constructive identification results in the basic Roll model, and then in various extended Roll models, including unbalanced order flow, general asymmetric supported trade direction indicators, serially correlated (Hidden Markov) trade direction indicators, adverse selection parameters and the random spread.

An (2017) studies identification and estimation of first-price auction models with unobserved heterogeneity, which can be either bidders' beliefs about their opponents' bidding behavior or asymmetrically distributed bidders' values. The paper shows that both cases can be identified by a unified methodology, where the unobserved heterogeneity can be considered as the latent true value in a measurement error model and the observed bids are the measurements of the true value. The empirical application in this paper suggests that bidders hold heterogeneous and non-equilibrium beliefs. Battistin et al. (2017) derive bounds on the distribution of math and language scores of elementary school students in Italy correcting for pervasive manipulation. They use a natural experiment that randomly assigns external monitors to schools to deal with endogeneity of manipulation as well as possible misclassification of the manipulation status. Their results show that score distributions are heavily affected by manipulating behavior, with regional rankings by academic performance being reversed once manipulation is taken into account.

A common problem in household surveys is that reported values exhibit a significant degree of heaping. Arulamplam et al. (2017) study the effects of a cash-incentive program in India on neonatal mortality using district-level household survey data. They model mortality using survival analysis, paying special attention to substantial heaping present in the data. Their empirical findings do not suggest a significant reduction of mortality in treated districts, but they do indicate that accounting for heaping matters for the estimation of the hazard parameters. Drerup et al. (2017) consider measurement error in subjective-expectations data and argue that the individual-level precision of such subjective data may reflect the structure of the underlying decision process and may provide useful information to uncover heterogeneity in choice behavior. To explore their conjecture, they estimate a semiparametric double index model on a dataset with subjective stock market expectations and investment decisions specifically collected for this purpose. Their results show that investment decisions exhibit little variation in economic model primitives when individuals provide error-ridden belief statements, and that, in contrast, beliefs and risk preferences predict strong variation in investment decisions for individuals who report precise expectations measures.

Besides the methodologies and the applications of measurement error models presented above, we expect this literature to advance further, with more important results. For example, the flexible nonclassical measurement error models may also provide new and convincing solutions to the endogeneity problem, a fundamental problem in econometrics. Presumably, a complete economic model should explain the causality among all the variables in the model. Endogeneity then occurs when some of the variables in the model are unobserved by the researcher. Nonclassical measurement error models may then be used to handle the unobservables, and therefore, solve the endogeneity problem under certain assumptions.

With more and more data available for researchers, we look forward to more extensive applications of the measurement error models. Given the nonparametric identification, nonparametric or semiparametric estimation of the models with latent variables may become easier than before. On the one hand, sample sizes will become much larger than before with the abundance of observations; on the other hand, researchers may observe more measurements of the latent variables. Therefore, we expect that the literature of measurement error models and their applications will keep thriving.

References

An, Yonghong, 2017. Identification of first-price auctions with non-equilibrium beliefs: A measurement error approach. J. Econometrics 200 (2), 326–343.

Arulamplam, Wiji, Corradi, Valentina, Gutknecht, Daniel, 2017. Modeling heaped duration data: An application to neonatal mortality. J. Econometrics 200 (2), 363–377.

Battistin, Erich, Vuri, Daniela, Nadai, Michele De, 2017. Counting rotten apples: Student achievement and score manipulation in Italian elementary schools. J. Econometrics 200 (2), 344–362.

Ben-Moshe, Dan, D'Haultfœuille, Xavier, Lewbel, Arthur, 2017. Identification of additive and polynomial models of mismeasured regressors without instruments. J. Econometrics 200 (2), 207–222.

Bollinger, Chris, van Hasselt, Martijn, 2017. Bayesian moment-based inference in a regression model with misclassification error. J. Econometrics 200 (2), 282–294.

Bound, John, Brown, Charles, Mathiowetz, Nancy, 2001. Measurement error in survey data. Handb. Econom. 5, 3705–3843.

Carroll, Raymond J., Ruppert, David, Stefanski, Leonard A., Crainiceanu, Ciprian M., 2012. Measurement Error in Nonlinear Models: A Modern Perspective. CRC press.

Chen, Xiaohong, Hong, Han, Nekipelov, Denis, 2011. Nonlinear models of measurement errors. J. Econom. Lit. 49 (4), 901–937.

Chen, Xiaohong, Linton, Oliver, Yi, Yanping, 2017. Semiparametric identification of the bid-ask spread in extended roll models. J. Econometrics 200 (2), 312–325.

Chesher, Andrew, 2017. Understanding the effect of measurement error on quantile regressions. J. Econometrics 200 (2), 223–237.

Davezies, Laurent, Barbanchon, Thomas Le, 2017. Regression discontinuity with continuous measurement error in the running variable. J. Econometrics 200 (2), 260–281.

Drerup, Tilman, Enke, Benjamin, von Gaudecker, Hans-Martin, 2017. The precision of subjective data and the explanatory power of economic models. J. Econometrics 200 (2), 378–389.

Garcia, Tanya, Ma, Yanyuan, 2017. Simultaneous treatment of unspecified heteroskedastic model error distribution and mismeasured covariates for restricted moment models. J. Econometrics 200 (2), 194–206.

Gospodinov, Nikolay, Komunjer, Ivana, Serena, Ng, 2017. Simulated minimum distance estimation of dynamic models with errors-in-variables. J. Econometrics 200 (2), 181–193.

Hahn, Jinyong, Ridder, Geert, 2017. Instrumental variable estimation of nonlinear models with nonclassical measurement error using control variates. J. Econometrics 200 (2), 238–250.

Hu, Yingyao, 2008. Identification and estimation of nonlinear models with misclassification error using instrumental variables: a general solution. J. Econometrics 144, 27–61.

Hu, Yingyao, 2017. The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics. J. Econometrics 200 (2), 154–168.

Hu, Yingyao, Schennach, Susanne, 2008. Instrumental variable treatment of nonclassical measurement error models. Econometrica 76, 195–216.

Lee, Nayoung, Moon, Hyungsik Roger, Zhou, Qiankun, 2017. Many IVs estimation of dynamic panel regression models with measurement error. J. Econometrics 200 (2), 251–259.

Meijer, Erik, Spierdijk, Laura, Wansbeek, Tom, 2017. Consistent estimation of linear panel data models with measurement error. J. Econometrics 200 (2), 169–180.

Meyer, Bruce, Mittag, Nikolas, 2017. Misclassification in binary choice models. J. Econometrics 200 (2), 295–311.

Schennach, Susanne M., 2016. Recent advances in the measurement error literature. Ann. Rev. Econom. 8, 341–377.

Wansbeek, Thomas J., Meijer, Erik, 2000. Measurement error and latent variables in econometrics. In: Advanced Textbooks in Economics. Elsevier.

Yingyao Hu
Johns Hopkins University, Department of Economics, 3400 N. Charles Street, 21218 Baltimore, MD, United States
E-mail address: yhu@jhu.edu.

Tom Wansbeek
University of Groningen, Faculty of Economics and Business, Nettelbosje 2, 9747 AE Groningen, The Netherlands
E-mail address: t.j.wansbeek@rug.nl.

Available online 24 June 2017