Electronic Journal of Statistics > Vol. 3 (2009) open journal systems

SIMEX and standard error estimation in semiparametric measurement error models

Tatiyana Apanasovich, *Thomas Jefferson University*
Raymond J. Carroll, *Texas A&M University*
Arnab Maity, *Harvard School of Public Health*

Abstract

SIMEX is a general-purpose technique for measurement error correction. There is a substantial literature on the application and theory of SIMEX for purely parametric problems, as well as for purely nonparametric regression problems, but there is neither application nor theory for semiparametric problems. Motivated by an example involving radiation dosimetry, we develop the basic theory for SIMEX in semiparametric problems using kernel-based estimation methods. This includes situations that the mismeasured variable is modeled purely parametrically, purely nonparametrically, or that the mismeasured variable has components that are modeled both parametrically and nonparametrically. Using our asymptotic expansions, easily computed standard error formulae are derived, as are the bias properties of the nonparametric estimator. The standard error method represents a new method for estimating variability of nonparametric estimators in semiparametric problems, and we show in both simulations and in our example that it improves dramatically on first order methods.
We find that for estimating the parametric part of the model, standard bandwidth choices of order $O(n^{-1/5})$ are sufficient to ensure asymptotic normality, and undersmoothing is not required. SIMEX has the property that it fits misspecified models, namely ones that ignore the measurement error. Our work thus also more generally describes the behavior of kernel-based methods in misspecified semiparametric problems.

AMS 2000 subject classifications: 62G05, 62H12, 62F12, 62J05.

Keywords: Berkson measurement errors, measurement error, misspecified models, nonparametric regression, radiation epidemiology, semiparametric models, SIMEX, simulation-extrapolation, standard error estimation, uniform expansions.

Full Text: PDF

Apanasovich, Tatiyana, Carroll, Raymond J., Maity, Arnab, SIMEX and standard error estimation in semiparametric measurement error models, Electronic Journal of Statistics, 3, (2009), 318-348 (electronic). DOI: 10.1214/08-EJS341.

References

[1] Ai, C. and Chen, X. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. Journal of Econometrics, 141, 5–43. MR2411735

[2] Carroll, R.J. and Hall, P. (2004). Low order approximations in deconvolution and regression with errors in variables. Journal of the Royal Statistical Society, Series B, 66, 31–46. MR2035757

[3] Carroll, R.J., Maca, J.D. and Ruppert, D. (1999). Nonparametric regression with errors in covariates. Biometrika, 86, 541–554. MR1723777

[4] Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. (2006).

Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. Chapman and Hall CRC Press. MR2243417

[5] Carroll, R.J., Ruppert, D. and Welsh, A. (1998). Local estimating equations. Journal of the American Statistical Association, 93, 214–227. MR1614624

[6] Claeskens, G. and Carroll, R.J. (2007). Post-model selection inference in semiparametric models. Biometri ka, 94,249–265. MR2331485

[7] Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. Annals of Statistics, 31, 1852–1884. MR2036392

[8] Cook, J.R. and Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association, 89, 1314–1328.

[9] Delaigle, A. and Hall, P. (2008). Using SIMEX for smoothing parameter choice in errors-in-variables problems. Journal of the American Statistical Association, 130, 280 – 287. MR2394636

[10] Greene, W.F. and Cai, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. Biometrics, 60, 987–996. MR2133551

[11] Gould, W.R., Stefanski, L.A. and Pollock, K.H. (1997). Effects of measurement error on catch-effort estimation. Canadian Journal of Fisheries and Aquatic Science, 54, 898–906.

[12] Hwang, W.H. and Huang, S.Y.H. (2003). Estimation in capture-recapture models when covariates are subject to measurement errors. Biometrics, 59, 1113–1122. MR2025137

[13] Jeong, M. and Kim, C. (2003). Some properties of SIMEX estimator in partially linear measurement error model. Journal of the Korean Statistical Society, 32, 85–92. MR1984043

[14] Kangas, A.S. (1998). Effect of errors-in-variables on coefficients of a growth model and on prediction of growth. Forest Ecology And Management, 102, 203–212.

[15] Kerber, R.L., Till, J.E., Simon, S.L., Lyon, J.L. Thomas, D.C., Preston-Martin, S., Rollison, M.L., Lloyd, R.D. and Stevens, W. (1993). A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. Journal of the American Medical Association, 270, 2076–2083.

[16] Lechner, S. and Pohlmeier, W. (2004). To blank or not to blank? A comparison of the effects of disclosure limitation methods on nonlinear regression estimates. Annals of the New York Academy of Sciences, 3050, 187–200.

[17] Liang H., Hardle, W. and Carroll, R.J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. Annals of Statistics. 27, 1519–1535. MR1742498

[18] Liang, H. and Ren, H.B. (2005). Generalized partially linear measurement error models. Journal of Computational and Graphical Statistics, 14, 237-250. MR2137900

[19] Lin, X. and Carroll, R.J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association, 95, 520–534. MR1803170

[20] Lin, X. and Carroll, R.J. (2006). Semiparametric estimation in general repeated measures problems. Journal of the Royal Statistical Society, Series B, 68, 68–88. MR2212575

[21] Lubin, J.H., Schafer, D.W. Ron, E., Stovall, M. and Carroll, R.J. (2004). A reanalysis of thyroid neoplasms in the Israeli tinea capitis study accounting for dose uncertainties. Radiation Research, 161, 359–368.

[22] Lyon, J.L., Alder, S.C., Stone, M.B., Scholl, A., Reading, J.C. Holubkov, R., Sheng, X. White, G.L., Hegmann, K.T., Anspaugh, L., Hoffman, F.O., Simon, S.L., Thomas, B., Carroll, R.J. and Meikle, A.W. (2006). Thyroid disease associated with exposure to the Nevada Test Site radiation: a reevaluation based on corrected dosimetry and examination data. Epidemiology, 17, 604–614.

[23] Maity, A., Ma, Y. and Carroll, R.J. (2007). Efficient estimation of population-Level summaries in general semiparametric regression models with missing response. Journal of the American Statistical Association, 102(477), 123–139. MR2293305

[24] Mallick, B., Hoffman, F.O. and Carroll, R.J. (2002). Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada Test Site. Biometrics, 58, 13–20. MR1891038

[25] Marcus, A.H. and Elias, R.W. (1998) Some useful statistical methods for model validation. Environmental Health Perspectives, 106, 1541–1550.

[26] Marschner, I.C., Emberson, J., Irwig, L. and Walter, S.D. (2004). The number needed to treat (NNT) can be adjusted for bias when the outcome is measured with error. Journal of Clinical Epidemiology, 57, 1244–1252.

[27] Pierce, D.A. and Kellerer, A. (2004). Adjusting for covariate errors with nonparametric assessment of the true covariate distribution. Biometrika, 91, 863–876. MR2126038

[28] Reeves, G.K., Cox, D.R., Darby, S.C. and Whitley, E. (1998). Some aspects of measurement error in explanatory variables for continuous and binary regression models. Statistics in Medicine, 17, 2157–2177.

[29] Ron, E. and Hoffman, F.O. (1999). Uncertainties in Radiation Dosimetry and Their Impact on Dose response Analysis. National Cancer Institute Press.

[30] Ruppert, D. (1997) Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. Journal of the American Statistical Association, 92, 1049-1062. MR1482136

[31] Ruppert, D., Sheather, S.J. and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression (Corr: 96V91 p1380). Journal of the American Statistical Association, 90, 1257–1270. MR1379468

[32] Schafer, D.W. and Gilbert, E.S. (2006). Some statistical implications of does uncertainty in radiation dose-response analyses. Radiation Research, 166, 303–312.

[33] Schafer, D.W., Lubin, J.H., Ron, E., Stovall, M. and Carroll, R.J. (2001). Thyroid cancer following scalp irradiation: a reanalysis accounting for uncertainty in dosimetry. Biometrics, 57, 689–697. MR1859805

[34] Simon, S.L., Till, J.E., Lloyd, R.D., Kerber, R.L., Thomas, D.C., Preston–Martin, S., Lyon, J.L. and Stevens, W. (1995). The Utah Leukemia case–control study: dosimetry methodology and results. Health Physics, 68, 460–471.

[35] Simon, S.L., Anspaugh, L.R., Hoffman, F.O., et al. (2006). 2004 update of dosimetry for the Utah Thyroid Cohort Study. Radiation Research, 165, 208–222.

[36] Solow, A.R. (1998). On fitting a population model in the presence of observation error. Ecology, 79, 1463–1466.

[37] Stefanski, L.A. and Cook, J. (1995). Simulation extrapolation: the measurement error jackknife. Journal of the American Statistical Association, 90, 1247–1256. MR1379467

[38] Staudenmeyer, J. and Ruppert, D. (2004). Local polynomial regression and simulation-extrapolation. Journal of the Royal Statistical Society, Series B, 66, 17–30. MR2035756

[39] Stevens, W., Till, J.E., Thomas, D.C., et al. (1992). Assessment of leukemia and

thyroid disease in relation to fallout in Utah: report of a cohort study of thyroid disease and radioactive fallout from the Nevada test site. University of Utah.

[40] Stram, D.O. and Kopecky, K.J. (2003). Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: some observations. Radiation Research, 160, 408–417.

[41] Wang, N., Lin, X., Gutierrez, R.G. and Carroll, R.J. (1998). Generalized linear mixed measurement error models. Journal of the American Statistical Association, 93, 249–261. MR1614636

[42] Wang, C.Y., Wang, N. and Wang, S. (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. Biometrics, 56, 487–495.

[43] Zhu, L. and Cui, H. (2003). A semiparametric regression model with errors in variables. Scandinavian Journal of Statistics. 30, 429-442. MR1983135