

Zhang, Jun; Feng, Zhenghui; Xu, Peirong; Liang, Hua

Generalized varying coefficient partially linear measurement errors models. (English)

Zbl 1398.62200

Ann. Inst. Stat. Math. 69, No. 1, 97-120 (2017).

Summary: We study generalized varying coefficient partially linear models when some linear covariates are error prone, but their ancillary variables are available. We first calibrate the error-prone covariates, then develop a quasi-likelihood-based estimation procedure. To select significant variables in the parametric part, we develop a penalized quasi-likelihood variable selection procedure, and the resulting penalized estimators are shown to be asymptotically normal and have the oracle property. Moreover, to select significant variables in the nonparametric component, we investigate asymptotic behavior of the semiparametric generalized likelihood ratio test. The limiting null distribution is shown to follow a Chi-square distribution, and a new Wilks phenomenon is unveiled in the context of error-prone semiparametric modeling. Simulation studies and a real data analysis are conducted to evaluate the performance of the proposed methods.

MSC:

- 62J12 Generalized linear models (logistic models)
- 62G08 Nonparametric regression and quantile regression
- 62G10 Nonparametric hypothesis testing
- 62G20 Asymptotic properties of nonparametric inference

Cited in 2 Documents

Keywords:

ancillary variables; errors-in-variable; error prone; Lasso; measurement errors; quasi-likelihood; penalized quasi-likelihood; SCAD; varying coefficient models; nonparametric component

Full Text: [DOI](#)

References:

- [1] Akaike, H, Maximum likelihood identification of Gaussian autoregressive moving average models, *Biometrika*, 60, 255-265, (1973) · [Zbl 0318.62075](#)
- [2] Cai, Z., Fan, J., Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95, 888-902. · [Zbl 0999.62052](#)
- [3] Cambien, F., Warnet, J., Eschwege, E., Jacqueson, A., Richard, J., Rosselin, G. (1987). Body mass, blood pressure, glucose, and lipids. Does plasma insulin explain their relationships? *Arteriosclerosis, Thrombosis, and Vascular Biology*, 7, 197-202.
- [4] Carroll, R. J., Wang, Y. (2008). Nonparametric variance estimation in the analysis of microarray data: A measurement error approach. *Biometrika*, 95(2), 437-449. · [Zbl 1437.62408](#)
- [5] Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438), 477-489. · [Zbl 0890.62053](#)
- [6] Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Nonlinear measurement error models, a modern perspective* (2nd ed.). New York: Chapman and Hall. · [Zbl 1119.62063](#)
- [7] Cook, J. R., Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314-1328. · [Zbl 0810.62028](#)
- [8] Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications* (Vol. 66). London: Chapman & Hall. · [Zbl 0873.62037](#)
- [9] Fan, J., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11, 1031-1057. · [Zbl 1098.62077](#)
- [10] Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360. · [Zbl 1073.62547](#)
- [11] Fan, J., Zhang, C. M., Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, 29, 153-193. · [Zbl 1029.62042](#)
- [12] Foster, D., George, E. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22,

1947-1975. · [Zbl 0829.62066](#)

- [13] Hall, P., Ma, Y. (2007). Semiparametric estimators of functional measurement error models with unknown error. *\textit{Journal of the Royal Statistical Society, Series B Statistical Methodology}*, *\textit{69}*(3), 429-446.
- [14] Han, T. S., van Leer, E. M., Seidell, J. C., Lean, M. E. (1995). Waist circumference action levels in the identification of cardiovascular risk factors: prevalence study in a random sample. *\textit{British Medical Journal (BMJ)}*, *\textit{311}*(7017), 1401-1405. · [Zbl 0671.62045](#)
- [15] Härdle, W., Liang, H., Gao, J. (2000). *\textit{Partially linear models}*. Heidelberg: Physica-Verlag. Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models (with discussion). *\textit{Journal of the Royal Statistical Society, Series B Statistical Methodology}*, *\textit{55}*, 757-796. · [Zbl 0812.62044](#)
- [16] Hastie, T., Tibshirani, R. (1993). Varying-coefficient models (with discussion). *\textit{Journal of the Royal Statistical Society, Series B Statistical Methodology}*, *\textit{55}*, 757-796. · [Zbl 0796.62060](#)
- [17] Hunsberger, S, Semiparametric regression in likelihood-based models, *Journal of the American Statistical Association*, 89, 1354-1365, (1994) · [Zbl 0812.62044](#)
- [18] Hunsberger, S., Albert, P. S., Follmann, D. A., Suh, E. (2002). Parametric and semiparametric approaches to testing for seasonal trend in serial count data. *\textit{Biostatistics}*, *\textit{33}*, 289-298. · [Zbl 1133.62319](#)
- [19] Li, G., Xue, L., Lian, H. (2011). Semi-varying coefficient models with a diverging number of components. *\textit{Journal of Multivariate Analysis}*, *\textit{102}*(7), 1166-1174. · [Zbl 1216.62060](#)
- [20] Li, R., Liang, H. (2008). Variable selection in semiparametric regression modeling. *\textit{The Annals of Statistics}*, *\textit{36}*, 261-286. · [Zbl 1132.62027](#)
- [21] Liang, H., Li, R. (2009). Variable selection for partially linear models with measurement errors. *\textit{Journal of the American Statistical Association}*, *\textit{104}*(485), 234-248. · [Zbl 1388.62208](#)
- [22] Lin, X., Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *\textit{Journal of the American Statistical Association}*, *\textit{96}*, 1045-1056. · [Zbl 1072.62566](#)
- [23] Lobach, I., Carroll, R. J., Spinka, C., Gail, M., Chatterjee, N. (2008). Haplotype-based regression analysis and inference of case-control studies with unphased genotypes and measurement errors in environmental exposures. *\textit{Biometrics}*, *\textit{64}*, 673-684. · [Zbl 1274.62829](#)
- [24] Lobach, I., Fan, R., Carroll, R. J. (2010). Genotype-based association mapping of complex diseases: gene-environment interactions with multiple genetic markers and measurement error in environmental exposures. *\textit{Genetic Epidemiology}*, *\textit{34}*, 792-802.
- [25] Ma, Y., Carroll, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *\textit{Journal of the American Statistical Association}*, *\textit{101}*(476), 1465-1474. · [Zbl 1171.62324](#)
- [26] Ma, Y., Li, R. (2010). Variable selection in measurement error models. *\textit{Bernoulli}*, *\textit{16}*(1), 274-300. · [Zbl 1200.62071](#)
- [27] Ma, Y., Tsiatis, A. A. (2006). On closed form semiparametric estimators for measurement error models. *\textit{Statistica Sinica}*, *\textit{16}*(1), 183-193. · [Zbl 1087.62047](#)
- [28] Robinson, PM, Root n-consistent semiparametric regression, *Econometrica*, 56, 931-954, (1988) · [Zbl 0647.62100](#)
- [29] Schwarz, G, Estimating the dimension of a model, *The Annals of Statistics*, 6, 461-464, (1978) · [Zbl 0379.62005](#)
- [30] Severini, T. A., Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *\textit{Journal of the American Statistical Association}*, *\textit{89}*, 501-511. · [Zbl 0798.62046](#)
- [31] Silverman, B. W. (1986). *\textit{Density estimation for statistics and data analysis}*, Vol. 26 of Monographs on statistics and applied probability. London: Chapman and Hall. · [Zbl 0617.62042](#)
- [32] Sinha, S., Mallick, B. K., Kipnis, V., Carroll, R. J. (2010). Semiparametric Bayesian analysis of nutritional epidemiology data in the presence of measurement error. *\textit{Biometrics}*, *\textit{66}*(2), 444-454. · [Zbl 1192.62092](#)
- [33] Speckman, PE, Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society, Series B Statistical Methodology*, 50, 413-436, (1988) · [Zbl 0671.62045](#)
- [34] Stefanski, LA, Unbiased estimation of a nonlinear function of a normal Mean with application to measurement error models, *Communications in Statistics. Theory and Methods*, 18, 4335-4358, (1989) · [Zbl 0707.62058](#)
- [35] Stefanski, L. A., Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *\textit{Biometrika}*, *\textit{74}*(4), 703-716. · [Zbl 0632.62052](#)
- [36] Tsiatis, A. A., Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *\textit{Biometrika}*, *\textit{91}*(4), 835-848. · [Zbl 1064.62034](#)
- [37] Wang, H., Xia, Y. (2009). Shrinkage estimation of the varying coefficient model. *\textit{Journal of the American Statistical Association}*, *\textit{104}*, 747-757. · [Zbl 1388.62213](#)
- [38] Wang, L., Liu, X., Liang, H., Carroll, R. J. (2011). Estimation and variable selection for generalized additive partial linear models. *\textit{The Annals of Statistics}*, *\textit{39}*, 1827-1851. · [Zbl 1227.62053](#)
- [39] Wei, F., Huang, J., Li, H. (2011). Variable selection and estimation in high-dimensional varyingcoefficient models. *\textit{Statistica Sinica}*, *\textit{21}*(4), 1515-1540. · [Zbl 1225.62056](#)
- [40] Xia, Y., Zhang, W., Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *\textit{Biometrika}*, *\textit{91}*, 661-681. · [Zbl 1108.62019](#)
- [41] Yi, G. Y., Ma, Y. Y., Carroll, R. J. (2012). A functional generalized method of moments approach for longitudinal studies

- with missing responses and covariate measurement error. *Biometrika*, *99*(1), 151-165. · [Zbl 1234.62131](#)
- [42] Yuan, M., Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B Statistical Methodology*, *68*, 49-67. · [Zbl 1141.62030](#)
- [43] Zhang, C, Calibrating the degrees of freedom for automatic data smoothing and effective curve checking, *Journal of the American Statistical Association*, *98*, 609-628, (2003) · [Zbl 1040.62027](#)
- [44] Zhang, W., Lee, S.-Y., Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, *82*, 166-188. · [Zbl 0995.62038](#)
- [45] Zhou, Y., Liang, H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *The Annals of Statistics*, *37*, 427-458. · [Zbl 1156.62036](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.