

Xia, Yingcun; Tong, Howell; Li, W. K.; Zhu, Li-Xing

An adaptive estimation of dimension reduction space (with discussion). (English)

Zbl 1091.62028

J. R. Stat. Soc., Ser. B, Stat. Methodol. 64, No. 3, 363-410 (2002).

Summary: Searching for an effective dimension reduction space is an important problem in regression, especially for high dimensional data. We propose an adaptive approach based on semiparametric models, which we call the (conditional) minimum average variance estimation (MAVE) method, within quite a general setting. The MAVE method has the following advantages. Most existing methods must undersmooth the nonparametric link function estimator to achieve a faster rate of consistency for the estimator of the parameters (than for that of the nonparametric function). In contrast, a faster consistency rate can be achieved by the MAVE method even without undersmoothing the nonparametric link function estimator. The MAVE method is applicable to a wide range of models, with fewer restrictions on the distribution of the covariates, to the extent that even time series can be included.

Because of the faster rate of consistency for the parameter estimators, it is possible for us to estimate the dimension of the space consistently. The relationship of the MAVE method with other methods is also investigated. In particular, a simple outer product gradient estimator is proposed as an initial estimator. In addition to theoretical results, we demonstrate the efficacy of the MAVE method for high dimensional data sets through simulation. Two real data sets are analysed by using the MAVE approach.

MSC:

62G08 Nonparametric regression and quantile regression
62J12 Generalized linear models (logistic models)

Cited in **6** Reviews
Cited in **289** Documents

Keywords:

average derivative estimation; local linear smoother; multiple time series; nonlinear time series analysis; principal Hessian direction; projection pursuit; semiparametrics; sliced inverse regression estimation

Full Text: [DOI](#)

References:

- [1] Auestad, Identification of nonlinear time series: first order characterization and order determination, *Biometrika* 77 pp 669– (1990)
- [2] Breiman, L. , Friedman, J. H. , Olshen, R. A. and Stone, C. J. (1984) *Classification and Regression Trees*. Belmont: Wadsworth. · Zbl 0541.62042
- [3] Cai, Functional-coefficient regression models for nonlinear time series, *J. Am. Statist. Ass.* 95 pp 941– (2000) · Zbl 0996.62078
- [4] Carroll, Generalized partially linear single-index models, *J. Am. Statist. Ass.* 92 pp 477– (1997) · Zbl 0890.62053
- [5] Carroll, Binary regressors in dimension reduction models: a new look at treatment comparisons, *Statist. Sin.* 5 pp 667– (1995) · Zbl 0828.62033
- [6] Chaudhuri, Piecewise-polynomial regression trees, *Statist. Sin.* 4 pp 143– (1994) · Zbl 0824.62032
- [7] Chen, Can SIR be as popular as multiple linear regression, *Statist. Sin.* 8 pp 289– (1989) · Zbl 0897.62069
- [8] Chen, Estimation of a projection-pursuit type regression model, *Ann. Statist.* 19 pp 142– (1991) · Zbl 0736.62055
- [9] Cheng, On consistent nonparametric order determination and chaos (with discussion), *J. R. Statist. Soc. B* 54 pp 427– (1992)
- [10] Cook, On the interpretation of regression plots, *J. Am. Statist. Ass.* 89 pp 177– (1994) · Zbl 0791.62066
- [11] â (1998) Principal Hessian directions revisited (with discussions). *J. Am. Statist. Ass.*, 93, 85 â 100.
- [12] Fan, J. and Gijbels, I. (1996) *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall. · Zbl 0873.62037
- [13] Fan, Statistical estimation in varying coefficient models, *Ann. Statist.* 27 pp 1491– (1999) · Zbl 0977.62039
- [14] Friedman, Projection pursuit regression, *J. Am. Statist. Ass.* 76 pp 817– (1981)
- [15] Hall, On projection pursuit regression, *Ann. Statist.* 17 pp 573– (1989) · Zbl 0698.62041
- [16] Hannan, The estimation of mixed moving average autoregressive system, *Biometrika* 56 pp 579– (1969) · Zbl 0186.52802

- [17] Härdle, Optimal smoothing in single-index models, *Ann. Statist.* 21 pp 157– (1993)
- [18] Härdle, Investigating smooth multiple regression by method of average derivatives, *J. Am. Statist. Ass.* 84 pp 986– (1989)
- [19] Hastie, Generalized additive models (with discussion), *Statist. Sci.* 1 pp 297– (1986) · [Zbl 0955.62603](#)
- [20] Hotelling, The most predictable criterion, *J. Educ. Psychol.* 26 pp 139– (1935)
- [21] Huber, Projection pursuit (with discussion), *Ann. Statist.* 13 pp 435– (1985)
- [22] Ichimura, Semiparametric least squares estimation of multiple index models: single equation estimation, *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (1991) · [Zbl 0766.62065](#)
- [23] Li, Sliced inverse regression for dimension reduction (with discussion), *J. Am. Statist. Ass.* 86 pp 316– (1991)
- [24] Li (1992) On principal Hessian directions for data visualisation and dimension reduction: another application of Stein’s Lemma. *Ann. Statist.*, 87, 1025 à 1039. · [Zbl 0765.62003](#)
- [25] Li, Interactive tree-structured regression via principal Hessian directions, *J. Am. Statist. Ass.* 95 pp 547– (2000) · [Zbl 1013.62074](#)
- [26] Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*. New York: Wiley. · [Zbl 0256.62002](#)
- [27] Schott, Determining the dimensionality in sliced inverse regression, *J. Am. Statist. Ass.* 89 pp 141– (1994) · [Zbl 0791.62069](#)
- [28] Severini, Profile likelihood and conditionally parametric models, *Ann. Statist.* 20 pp 1768– (1992) · [Zbl 0768.62015](#)
- [29] Smith, *Assessing the human health risk of atmospheric particles*, *Environmental Statistics: Analysing Data for Environmental Policy* (1999)
- [30] Stone, Cross-validated choice and assessment of statistical predictions (with discussion), *J. R. Statist. Soc. B* 36 pp 111– (1974) · [Zbl 0308.62063](#)
- [31] Tong, H. (1990) *Nonlinear Time Series Analysis: a Dynamical System Approach*. Oxford: Oxford University Press.
- [32] Xia, Projection pursuit autoregression in time series. *J. Time Ser. Anal.* 20 pp 693– (1999) · [Zbl 0940.62083](#)
- [33] Xia, On single-index coefficient regression models, *J. Am. Statist. Ass.* 94 pp 1275– (1999) · [Zbl 1069.62548](#)
- [34] Xia, On extended partially linear single-index models, *Biometrika* 86 pp 831– (1999) · [Zbl 0942.62109](#)
- [35] Yang, Multivariate bandwidth selection for local linear regression, *J. R. Statist. Soc. B* 61 pp 793– (1999) · [Zbl 0952.62039](#)
- [36] Yao, On subset selection in nonparametric stochastic regression, *Statist. Sin.* 4 pp 51– (1994)
- [37] Zhu, Asymptotics for kernel estimate of sliced inverse regression, *Ann. Statist.* 24 pp 1053– (1996) · [Zbl 0864.62027](#)
- [38] Akaike, *Proc. 2nd Int. Symp. Information Theory* pp 267– (1973)
- [39] Atkinson, *Robust Diagnostic Regression Analysis* (2000) · [Zbl 0964.62063](#) · [doi:10.1007/978-1-4612-1160-0](#)
- [40] Basilevsky, *Statistical Factor Analysis and Related Methods* (1994)
- [41] Bickel, *Efficient and Adaptive Inference in Semiparametric Models* (1993)
- [42] Brillinger, *A Festschrift for Erich L. Lehmann* pp 97– (1983)
- [43] Li (1992) Nerve cell spike train data analysis: a progression of technique. *J. Am. Statist. Ass.*, 87, 260 à 271.
- [44] Brown, Bayesian wavelet regression on curves with application to a spectroscopic calibration problem, *J. Am. Statist. Ass.* 96 pp 398– (2001) · [Zbl 1022.62027](#)
- [45] Bura, Extending sliced inverse regression: the weighted chi-squared test, *J. Am. Statist. Ass.* 96 pp 990– (2001) · [Zbl 1047.62035](#)
- [46] Li (2001b) Estimating the structural dimension of regressions via parametric inverse regression. *J. R. Statist. Soc. B*, 63, 393 à 410. · [Zbl 0979.62041](#)
- [47] Carroll, Generalized partially linear single-index models, *J. Am. Statist. Ass.* 92 pp 477– (1997) · [Zbl 0890.62053](#)
- [48] Carroll, Second order effects in semiparametric weighted least squares regression, *Statistics* 2 pp 179– (1989) · [Zbl 0669.62020](#)
- [49] Cheng, On consistent nonparametric order determination and chaos (with discussion), *J. R. Statist. Soc.* 54 pp 451– (1992)
- [50] Chiaromonte, F. , Cook, R. D. and Li, B. (2002) Sufficient dimension reduction in regressions with categorical predictors. *Ann. Statist.*, 30, in the press. · [Zbl 1012.62036](#)
- [51] Cook, Graphics for regressions with a binary response, *J. Am. Statist. Ass.* 91 pp 983– (1996) · [Zbl 0882.62060](#)
- [52] Li (1996b) *Regression Graphics: Ideas for Studying Regressions through Graphics*. New York: Wiley.
- [53] Li (1998) *Regression Graphics*. New York: Wiley.
- [54] Cook, Identifying regression outliers and mixtures graphically, *J. Am. Statist. Ass.* 95 pp 781– (2000) · [Zbl 0999.62056](#)
- [55] Cook, R. D. and Li, B. (2002) Dimension reduction for conditional mean in regression. *Ann. Statist.*, 30, in the press. · [Zbl 1012.62035](#)
- [56] Cook, Re-weighting to achieve elliptically contoured covariates in regression, *J. Am. Statist. Ass.* 89 pp 592– (1994) · [Zbl 0799.62078](#)
- [57] Cook, Discussion on ‘Sliced inverse regression’ (by K. C. Li), *J. Am. Statist. Ass.* 86 pp 316– (1991)
- [58] Li (1994) *An Introduction to Regression Graphics*. New York: Wiley.
- [59] Cook, Dimension reduction and visualization in discriminant analysis (with discussion), *Aust. New Z. J. Statist.* 43 pp 147–

- (2001) · [Zbl 0992.62056](#)
- [60] Cui (2001) · [Zbl 1005.68529](#)
- [61] Dauxois, Un modèle semi-paramétrique pour variables hilbertiennes, C. R. Acad. Sci. 333 pp 947– (2001) · [Zbl 0996.62035](#) · [doi:10.1016/S0764-4442\(01\)02163-2](#)
- [62] Duan, A bias bound for least squares linear regression, Statist. Sin. 1 pp 127– (1991) · [Zbl 0824.62057](#)
- [63] Fan, Local Polynomial Modelling and Its Applications (1996)
- [64] Fan, Adaptive varying-coefficient linear models, J. R. Statist. Soc. (2001)
- [65] Ferraty, Dimension choice for Sliced Inverse Regression based on ranks, Student 2 pp 95– (1997)
- [66] (1998) Determination of the dimension in SIR and related methods. J. Am. Statist. Ass., 93, 132 à 140.
- [67] Fujikoshi, Selection of variables in two-group discriminant analysis by error rate and Akaike's information criteria, J. Multiv. Anal. 17 pp 27– (1985) · [Zbl 0591.62053](#)
- [68] Hall, On almost linearity of low dimensional projections from high dimensional data, Ann. Statist. 21 pp 867– (1993) · [Zbl 0782.62065](#)
- [69] Härdle, Optimal smoothing in single-index models, Ann. Statist. 21 pp 157– (1993)
- [70] Härdle, Investigating smooth multiple regression by method of average derivatives, J. Am. Statist. Ass. 84 pp 986– (1989)
- [71] Hristache, Ann. Statist. 29 pp 1537– (2001)
- [72] (2002) Structure adaptive approach for dimension reduction. Ann. Statist., to be published.
- [73] Hristache, Direct estimation of the index coefficients in a single-index model, Ann. Statist. 29 pp 595– (2001) · [Zbl 1012.62043](#)
- [74] Huber, Projection pursuit (with discussion), Ann. Statist. 13 pp 435– (1985)
- [75] Ichimura, Semiparametric least squares (SLS) and weighted SLS estimation of single index models, J. Econometr. 58 pp 71– (1993) · [Zbl 0816.62079](#)
- [76] Li, Some recent developments in projection pursuit in China, Statist. Sin. 3 pp 35– (1993) · [Zbl 0823.62056](#)
- [77] Li, Sliced inverse regression for dimension reduction (with discussion), J. Am. Statist. Ass. 86 pp 316– (1991)
- [78] (1997) Nonlinear confounding in high dimensional regression. Ann. Statist., 57, 577 à 612. · [Zbl 0873.62071](#)
- [79] Li, PhD Thesis (2000)
- [80] Li, Semiparametric reduced-rank regression, Technical Report 310 (2001)
- [81] Li, Regression analysis under link violation, Ann. Statist. 17 pp 1009– (1989) · [Zbl 0753.62041](#)
- [82] Linton, Second order approximation in the partially linear regression model, Econometrica 63 pp 1079– (1995) · [Zbl 0836.62050](#)
- [83] Mallows, Some comments on C_p , Technometrics 15 pp 661– (1973) · [Zbl 0269.62061](#)
- [84] Murphy, On profile likelihood (with discussion), J. Am. Statist. Ass. 95 pp 449– (2000)
- [85] Posse, Projection pursuit exploratory data analysis, Comput. Statist. Data Anal. 20 pp 669– (1995) · [Zbl 0875.62206](#)
- [86] Riani, A unified approach to outliers, influence, and transformations in discriminant analysis, J. Comput. Graph Statist. 10 pp 513– (2001)
- [87] Robinson, Root-N-consistent semiparametric regression, Econometrica 56 pp 931– (1988) · [Zbl 0647.62100](#)
- [88] Ruckstuhl, Reference band for nonparametrically estimated link function, J. Comput. Graph. Statist. 8 pp 699– (1999)
- [89] Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 pp 379– (1948) · [Zbl 1154.94303](#) · [doi:10.1002/j.1538-7305.1948.tb01338.x](#)
- [90] Shibata, Selection of the order of an autoregressive model by Akaike's information criterion, Biometrika 63 pp 117– (1976) · [Zbl 0358.62048](#)
- [91] Stenseth, Common dynamic structure of Canada lynx populations within three climatic regions, Science 285 pp 1017– (1999)
- [92] Stenseth, From patterns to processes: phases and density dependencies in the Canadian lynx cycle, Proc. Natn. Acad. Sci. USA 95 pp 15430– (1998)
- [93] Stone, Cross-validated choice and assessment of statistical predictions (with discussion), J. R. Statist. Soc. 36 pp 111– (1974) · [Zbl 0308.62063](#)
- [94] Stone, Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression (with discussion), J. R. Statist. Soc. 52 pp 237– (1990) · [Zbl 0708.62054](#)
- [95] Velilla, Assessing the number of linear components in a general regression problem, J. Am. Statist. Ass. 93 pp 1008– (1998) · [Zbl 1063.62553](#)
- [96] Weisberg, Adapting for the missing link, Ann. Statist. 22 pp 1674– (1994) · [Zbl 0828.62059](#)
- [97] Xia, On single-index coefficient regression models, J. Am. Statist. Ass. 94 pp 1275– (1999) · [Zbl 1069.62548](#)
- [98] Xia, Statist. Sin. (2002)
- [99] Zhu, Technical Report (2002)
- [100] Zhu, Asymptotics for kernel estimate of sliced inverse regression, Ann. Statist. 24 pp 1053– (1996) · [Zbl 0864.62027](#)

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.