

Li, Yujie; Li, Gaorong; Lian, Heng; Tong, Tiejun

Profile forward regression screening for ultra-high dimensional semiparametric varying coefficient partially linear models. (English) Zbl 1360.62180

J. Multivariate Anal. 155, 133-150 (2017).

Summary: In this paper, we consider semiparametric varying coefficient partially linear models when the predictor variables of the linear part are ultra-high dimensional where the dimensionality grows exponentially with the sample size. We propose a profile forward regression (PFR) method to perform variable screening for ultra-high dimensional linear predictor variables. The proposed PFR algorithm can not only identify all relevant predictors consistently even for ultra-high semiparametric models including both nonparametric and parametric parts, but also possesses the screening consistency property. To determine whether or not to include the candidate predictor in the model of selected ones, we adopt an extended Bayesian information criterion (EBIC) to select the “best” candidate model. Simulation studies and a real data example are also carried out to assess the performance of the proposed method and to compare it with existing screening methods.

MSC:

- 62G08 Nonparametric regression and quantile regression
- 62J02 General nonlinear regression
- 62H12 Estimation in multivariate analysis
- 62G20 Asymptotic properties of nonparametric inference

Cited in **3** Documents

Keywords:

varying coefficient partially linear model; profile forward regression; variable screening; screening consistency property; ultra-high dimension; EBIC

Full Text: [DOI](#)

References:

- [1] Ahmad, I.; Leelahanon, S.; Li, Q., Efficient estimation of a semiparametric partially linear varying coefficient model, *Ann. Statist.*, 33, 258-283, (2005) · [Zbl 1064.62043](#)
- [2] Chen, J. H.; Chen, Z. H., Extended Bayesian information criterion for model selection with large model spaces, *Biometrika*, 95, 759-771, (2008) · [Zbl 1437.62415](#)
- [3] M.Y. Cheng, S.Y. Feng, G.R. Li, H. Lian, Greedy forward regression for variable screening. Preprint, arXiv:1511.01124, 2015. · [Zbl 06867960](#)
- [4] Cheng, M. Y.; Honda, T.; Zhang, J. T., Forward variable selection for sparse ultra-high dimensional varying coefficient models, *J. Amer. Statist. Assoc.*, 111, 1209-1221, (2016)
- [5] Chernoff, H., A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statist.*, 23, 493-507, (1952) · [Zbl 0048.11804](#)
- [6] Cui, H. J.; Li, R. Z.; Zhong, W., Model-free feature screening for ultra-high dimensional discriminant analysis, *J. Amer. Statist. Assoc.*, 110, 630-641, (2015) · [Zbl 1373.62305](#)
- [7] Dudoit, S.; Fridlyand, J.; Speed, T. P., Comparison of discriminant methods for the classification of tumors using gene expression data, *J. Amer. Statist. Assoc.*, 97, 77-87, (2002) · [Zbl 1073.62576](#)
- [8] Fan, J. Q.; Feng, Y.; Song, R., Nonparametric independence screening in sparse ultra-high dimensional additive models, *J. Amer. Statist. Assoc.*, 106, 544-557, (2011) · [Zbl 1232.62064](#)
- [9] Fan, J. Q.; Huang, T., Profile likelihood inferences on semiparametric varying-coefficient partially linear models, *Bernoulli*, 11, 1031-1057, (2005) · [Zbl 1098.62077](#)
- [10] J.Q. Fan, R.Z. Li, Statistical challenges with high-dimensionality: feature selection in knowledge discovery, in: *Proceedings of International Congress of Mathematicians (M. Sanz-Solé, J. Soria, J.L. Varona, J. Verdera, eds.)*, Vol. III, 2006, pp. 595-622. · [Zbl 1117.62137](#)
- [11] Fan, J. Q.; Lv, J. C., Sure independence screening for ultra-high dimensional feature space (with discussion), *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 70, 849-911, (2008)
- [12] Fan, J. Q.; Ma, Y. B.; Dai, W., Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient

- models, *J. Amer. Statist. Assoc.*, 109, 1270-1284, (2014) · [Zbl 1368.62095](#)
- [13] Fan, J. Q.; Samworth, R.; Wu, Y. C., Ultra-high dimensional feature selection: beyond the linear model, *J. Mach. Learn. Res.*, 10, 2013-2038, (2009) · [Zbl 1235.62089](#)
- [14] Fan, J. Q.; Song, R., Sure independence screening in generalized linear models with NP-dimensionality, *Ann. Statist.*, 38, 3567-3604, (2010) · [Zbl 1206.68157](#)
- [15] Gilliam, M.; Rifas-Shiman, S.; Berkey, C.; Field, A.; Colditz, G., Maternal gestational diabetes, birth weight and adolescent obesity, *Pediatrics*, 111, 221-226, (2003)
- [16] Hong, Z. P.; Hu, Y.; Lian, H., Variable selection for high-dimensional varying coefficient partially linear models via nonconcave penalty, *Metrika*, 76, 887-908, (2013) · [Zbl 06224852](#)
- [17] Ishida, M.; Monk, D.; Duncan, A. J.; Abu-Amero, S.; Chong, J.; Ring, S. M.; Pembrey, M. E.; Hindmarsh, P. C.; Stanier, P.; Moore, G. E., Maternal inheritance of a promoter variant in the imprinted PHLDA2 gene significantly increases birth weight, *Am. J. Hum. Genet.*, 90, 715-719, (2012)
- [18] Kai, B.; Li, R. Z.; Zou, H., New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models, *Ann. Statist.*, 39, 305-332, (2011) · [Zbl 1209.62074](#)
- [19] Lam, C.; Fan, J. Q., Profile-kernel likelihood inference with diverging number of parameters, *Ann. Statist.*, 36, 2232-2260, (2008) · [Zbl 1274.62289](#)
- [20] Li, G. R.; Feng, S. Y.; Peng, H., Profile-type smoothed score function for a varying coefficient partially linear model, *J. Multivariate Anal.*, 102, 372-385, (2011) · [Zbl 1327.62263](#)
- [21] Li, G. R.; Zhang, J.; Feng, S. Y., *Modern measurement error models*, (2016), Science Press Beijing
- [22] Li, Q.; Huang, C. J.; Li, D.; Fu, T. T., Semiparametric smooth coefficient models, *J. Bus. Econom. Statist.*, 20, 412-422, (2002)
- [23] Li, R. Z.; Liang, H., Variable selection in semiparametric regression modeling, *Ann. Statist.*, 36, 261-286, (2008) · [Zbl 1132.62027](#)
- [24] Li, G. R.; Lin, L.; Zhu, L. X., Empirical likelihood for varying coefficient partially linear model with diverging number of parameters, *J. Multivariate Anal.*, 105, 85-111, (2012) · [Zbl 1236.62020](#)
- [25] Li, G. R.; Peng, H.; Zhang, J.; Zhu, L. X., Robust rank correlation based screening, *Ann. Statist.*, 40, 1846-1877, (2012) · [Zbl 1257.62067](#)
- [26] Li, G. R.; Peng, H.; Zhu, L. X., Nonconcave penalized M-estimation with a diverging number of parameters, *Statist. Sinica*, 21, 391-419, (2011) · [Zbl 1206.62036](#)
- [27] Li, G. R.; Xue, L. G.; Lian, H., Semi-varying coefficient models with a diverging number of components, *J. Multivariate Anal.*, 102, 1166-1174, (2011) · [Zbl 1216.62060](#)
- [28] Li, R. Z.; Zhong, W.; Zhu, L. P., Feature screening via distance correlation learning, *J. Amer. Statist. Assoc.*, 107, 1129-1139, (2012) · [Zbl 1443.62184](#)
- [29] Liang, H.; Wang, H. S.; Tsai, C. L., Profile forward regression for ultrahigh dimensional variable screening in semiparametric partially linear models, *Statist. Sinica*, 22, 531-554, (2012) · [Zbl 1238.62045](#)
- [30] Liu, J. Y.; Li, R. Z.; Wu, R. L., Feature selection for varying coefficient models with ultrahigh-dimensional covariates, *J. Amer. Statist. Assoc.*, 109, 266-274, (2014) · [Zbl 1367.62048](#)
- [31] Sherwood, B.; Wang, L., Partially linear additive quantile regression in ultrahigh dimension, *Ann. Statist.*, 44, 288-317, (2016) · [Zbl 1331.62264](#)
- [32] Votavová, H.; Dostálová Merkerová, M.; Fejglova, K.; Vašíková, A.; Krejčík, Z.; Pastorková, A.; Tabashidze, N.; Topinka, J.; Velemínský, M.; Šrám, R. J.; Brdička, R., Transcriptome alterations in maternal and fetal cells induced by tobacco smoke, *Placenta*, 32, 763-770, (2011)
- [33] Wang, H. S., Forward regression for ultra-high dimensional variable screening, *J. Amer. Statist. Assoc.*, 104, 1512-1524, (2009) · [Zbl 1205.62103](#)
- [34] Wang, L. F.; Li, H. Z.; Huang, J., Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements, *J. Amer. Statist. Assoc.*, 103, 1556-1569, (2008) · [Zbl 1286.62034](#)
- [35] Wu, C.; Cui, Y. H.; Ma, S. G., Integrative analysis of gene-environment interactions under a multi-response partially linear varying coefficient model, *Stat. Med.*, 33, 4988-4998, (2014)
- [36] Xia, Y. C.; Zhang, W. Y.; Tong, H., Efficient estimation for semivarying-coefficient models, *Biometrika*, 91, 661-681, (2004) · [Zbl 1108.62019](#)
- [37] Xue, L. G.; Zhu, L. X., Empirical likelihood for a varying coefficient model with longitudinal data, *J. Amer. Statist. Assoc.*, 102, 642-654, (2007) · [Zbl 1172.62306](#)
- [38] You, J. H.; Chen, G. M., Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model, *J. Multivariate Anal.*, 97, 324-341, (2006) · [Zbl 1085.62043](#)
- [39] You, J. H.; Zhou, Y., Empirical likelihood for semiparametric varying-coefficient partially linear regression models, *Statist. Probab. Lett.*, 76, 412-422, (2006) · [Zbl 1086.62057](#)
- [40] Zhang, W. W.; Li, G. R.; Xue, L. G., Profile inference on partially linear varying-coefficient errors-in-variables models under restricted condition, *Comput. Statist. Data Anal.*, 55, 3027-3040, (2011) · [Zbl 1218.62038](#)
- [41] Zhao, P. X.; Xue, L. G., Variable selection for semiparametric varying coefficient partially linear models, *Statist. Probab. Lett.*, 79, 2148-2157, (2009) · [Zbl 1171.62026](#)

- [42] Zhao, W. H.; Zhang, R. Q.; Liu, J.; Lv, Y. Z., Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression, *Ann. Inst. Statist. Math.*, 66, 165-191, (2014) · [Zbl 1281.62109](#)
- [43] Zhong, W. X.; Zhang, T. T.; Zhu, Y.; Liu, J. S., Correlation pursuit: forward stepwise variable selection for index models, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 74, 849-870, (2012)
- [44] Zhou, Y.; Liang, H., Statistical inference for semiparametric varying-coefficient partially linear models with generated regressors, *Ann. Statist.*, 37, 427-458, (2009) · [Zbl 1156.62036](#)
- [45] Zhu, L. P.; Li, L. X.; Li, R. Z.; Zhu, L. X., Model-free feature screening for ultrahigh-dimensional data, *J. Amer. Statist. Assoc.*, 106, 1464-1475, (2011) · [Zbl 1233.62195](#)

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.