

**Wang, Zhaoliang; Xue, Liugen; Li, Gaorong; Lu, Fei**

**Spline estimator for ultra-high dimensional partially linear varying coefficient models.** (English) [Zbl 1419.62095](#)

*Ann. Inst. Stat. Math.* 71, No. 3, 657-677 (2019).

**Summary:** In this paper, we simultaneously study variable selection and estimation problems for sparse ultra-high dimensional partially linear varying coefficient models, where the number of variables in linear part can grow much faster than the sample size while many coefficients are zeros and the dimension of nonparametric part is fixed. We apply the B-spline basis to approximate each coefficient function. First, we demonstrate the convergence rates as well as asymptotic normality of the linear coefficients for the oracle estimator when the nonzero components are known in advance. Then, we propose a nonconvex penalized estimator and derive its oracle property under mild conditions. Furthermore, we address issues of numerical implementation and of data adaptive choice of the tuning parameters. Some Monte Carlo simulations and an application to a breast cancer data set are provided to corroborate our theoretical findings in finite samples.

**MSC:**

[62G08](#) Nonparametric regression and quantile regression

[62H12](#) Estimation in multivariate analysis

[62G20](#) Asymptotic properties of nonparametric inference

**Keywords:**

[high dimensionality](#); [partially linear varying coefficient model](#); [variable selection](#); [nonconvex penalty](#); [oracle property](#)

**Full Text:** [DOI](#)

**References:**

- [1] Ahmad, I., Leelahanon, S., Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *\textit{The Annals of Statistics}*, *\textit{33}*, 258-283. · [Zbl 1064.62043](#)
- [2] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., Wellner, J. A. (1998). *\textit{Efficient and adaptive estimation for semiparametric models}*. New York: Springer. · [Zbl 0894.62005](#)
- [3] Bühlmann, P., Van de Geer, S. (2011). *\textit{Statistics for high dimensional data}*. Berlin: Springer.
- [4] Chen, J. H., Chen, Z. H. (2008). Extended bayesian information criteria for model selection with large model spaces. *\textit{Biometrika}*, *\textit{95}*, 759-771. · [Zbl 1437.62415](#)
- [5] Cheng, M. Y., Honda, T., Zhang, J. T. (2016). Forward variable selection for sparse ultra-high dimensional varying coefficient models. *\textit{Journal of the American Statistical Association}*, *\textit{111}*, 1209-1221.
- [6] de Boor, C. (2001). *\textit{A practical guide to splines}*. New York: Springer. · [Zbl 0987.65015](#)
- [7] Fan, J. Q., Huang, T. (2005). Profile likelihood inferences on semiparametric varying coefficient partially linear models. *\textit{Bernoulli}*, *\textit{11}*, 1031-1057. · [Zbl 1098.62077](#)
- [8] Fan, J. Q., Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *\textit{Journal of the American Statistical Association}*, *\textit{96}*, 1348-1360. · [Zbl 1073.62547](#)
- [9] Fan, J. Q., Lv, J. C. (2010). A selective overview of variable selection in high dimensional feature space. *\textit{Statistica Sinica}*, *\textit{20}*, 101-148. · [Zbl 1180.62080](#)
- [10] Fan, J. Q., Lv, J. C. (2011). Non-concave penalized likelihood with NP-dimensionality. *\textit{IEEE Transactions on Information Theory}*, *\textit{57}*, 5467-5484. · [Zbl 1365.62277](#)
- [11] Feng, S. Y., Xue, L. G. (2014). Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition. *\textit{Annals of the Institute of Statistical Mathematics}*, *\textit{66}*, 121-140. · [Zbl 1281.62098](#)
- [12] Huang, J., Horowitz, J. L., Wei, F. R. (2010). Variable selection in nonparametric additive models. *\textit{The Annals of Statistics}*, *\textit{38}*, 2282-2313. · [Zbl 1202.62051](#)
- [13] Huang, Z. S., Zhang, R. Q. (2009). Empirical likelihood for nonparametric parts in semiparametric varying coefficient partially linear models. *\textit{Statistics and Probability Letters}*, *\textit{79}*, 1798-1808. · [Zbl 1169.62028](#)

- [14] Kai, B., Li, R. Z., Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying coefficient partially linear models. *\textit{The Annals of Statistics}*, *\textit{39}*, 305-332. · [Zbl 1209.62074](#)
- [15] Knight, W. A., Livingston, R. B., Gregory, E. J., Mc Guire, W. L. (1977). Estrogen receptor as an independent prognostic factor for early recurrence in breast cancer. *\textit{Cancer Research}*, *\textit{37}*, 4669-4671.
- [16] Koren, Y., Bell, R., Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *\textit{Computer}*, *\textit{42}*(8), 30-37.
- [17] Li, G. R., Feng, S. Y., Peng, H. (2011a). A profile type smoothed score function for a varying coefficient partially linear model. *\textit{Journal of Multivariate Analysis}*, *\textit{102}*, 372-385. · [Zbl 1327.62263](#)
- [18] Li, G. R., Xue, L. G., Lian, H. (2011b). Semi-varying coefficient models with a diverging number of components. *\textit{Journal of Multivariate Analysis}*, *\textit{102}*, 1166-1174. · [Zbl 1216.62060](#)
- [19] Li, G. R., Lin, L., Zhu, L. X. (2012). Empirical likelihood for varying coefficient partially linear model with diverging number of parameters. *\textit{Journal of Multivariate Analysis}*, *\textit{105}*, 85-111. · [Zbl 1236.62020](#)
- [20] Li, R. Z., Liang, H. (2008). Variable selection in semiparametric regression modeling. *\textit{The Annals of Statistics}*, *\textit{36}*(1), 261-286.
- [21] Li, Y. J., Li, G. R., Lian, H., Tong, T. J. (2017). Profile forward regression screening for ultra-high dimensional semiparametric varying coefficient partially linear models. *\textit{Journal of Multivariate Analysis}*, *\textit{155}*, 133-150. · [Zbl 1360.62180](#)
- [22] Lustig, M., Donoho, D. L., Santos, J. M., Pauly, J. M. (2008). Compressed sensing MRI. *\textit{IEEE Signal Processing Magazine}*, *\textit{25}*, 72-82.
- [23] Stone, C.J. Additive regression and other nonparametric models, *The Annals of Statistics*, 13, 689-705, (1985) · [Zbl 0605.62065](#)
- [24] Sun, J., Lin, L. (2014). Local rank estimation and related test for varying coefficient partially linear models. *\textit{Journal of Nonparametric Statistics}*, *\textit{26}*, 187-206. · [Zbl 1359.62111](#)
- [25] Tibshirani, R., Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, 58, 267-288, (1996) · [Zbl 0850.62538](#)
- [26] van't Veer, L. J., Dai, H. Y., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., Friend, S. H., (2002). Gene expression profiling predicts clinical outcome of breast cancer. *\textit{Nature}*, *\textit{415}*, 530-536.
- [27] Wei, FR, Group selection in high dimensional partially linear additive models, *Brazilian Journal of Probability and Statistics*, 26, 219-243, (2012) · [Zbl 1239.62048](#)
- [28] Wei, F. R., Huang, J., Li, H. Z. (2011). Variable selection and estimation in high dimensional varying coefficient models. *\textit{Statistica Sinica}*, *\textit{21}*, 1515-1540. · [Zbl 1225.62056](#)
- [29] Xie, H. L., Huang, J. (2009). SCAD penalized regression in high dimensional partially linear models. *\textit{The Annals of Statistics}*, *\textit{37}*, 673-696. · [Zbl 1162.62037](#)
- [30] You, J. H., Chen, G. M. (2006a). Estimation of a semiparametric varying coefficient partially linear errors-in-variables model. *\textit{Journal of Multivariate Analysis}*, *\textit{97}*, 324-341. · [Zbl 1085.62043](#)
- [31] You, J. H., Zhou, Y. (2006b). Empirical likelihood for semiparametric varying coefficient partially linear model. *\textit{Statistics and Probability Letters}*, *\textit{76}*, 412-422. · [Zbl 1086.62057](#)
- [32] Yu, T., Li, J. L., Ma, S. G. (2012). Adjusting confounders in ranking biomarkers: A model-based ROC approach. *\textit{Briefings in Bioinformatics}*, *\textit{13}*, 513-523.
- [33] Zhang, CH, Nearly unbiased variable selection under minimax concave penalty, *The Annals of Statistics*, 38, 894-942, (2010) · [Zbl 1183.62120](#)
- [34] Zhao, P. X., Xue, L. G. (2009). Variable selection for semiparametric varying coefficient partially linear models. *\textit{Statistics and Probability Letters}*, *\textit{79}*, 2148-2157. · [Zbl 1171.62026](#)
- [35] Zhao, W. H., Zhang, R. Q., Liu, J. C., Lv, Y. Z. (2014). Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression. *\textit{Annals of the Institute of Statistical Mathematics}*, *\textit{66}*, 165-191. · [Zbl 1281.62109](#)
- [36] Zhou, S., Shen, X., Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *\textit{The Annals of Statistics}*, *\textit{26}*, 1760-1782. · [Zbl 0929.62052](#)
- [37] Zhou, Y., Liang, H. (2009). Statistical inference for semiparametric varying coefficient partially linear models with error-prone linear covariates. *\textit{The Annals of Statistics}*, *\textit{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.