MR4265707 62G07 62G08 62G20
Zhang, Jun [Zhang, Jun[24]] (PRC-SZU-CMS); Lin, Bingqing (PRC-SZU-CMS);
Zhou, Yan [Zhou, Yan[11]] (PRC-SZU-CMS)

Kernel density estimation for partial linear multivariate responses models. (English summary)

J. Multivariate Anal. **185** (2021), *Paper No.* 104768, 17 *pp.*

This article proposes an estimation procedure for the q-dimensional partial linear multivariate response model

$$(1)\qquad \begin{bmatrix} Y_1 \\ \vdots \\ Y_q \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \boldsymbol{\beta}_0 \\ \vdots \\ \mathbf{X}^\top \boldsymbol{\beta}_0 \end{bmatrix} + \begin{bmatrix} g_1(U) \\ \vdots \\ g_q(U) \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_q \end{bmatrix}.$$

In this model, the p-dimensional covariate vector $\mathbf{X}$ and parameter vector of interest $\boldsymbol{\beta}_0$ in the parametric component are common across q equations, whereas unknown functions $g_1(U), \dots, g_q(U)$ with a common scalar covariate U in the nonparametric component vary across q equations. The errors $(\epsilon_1, \dots, \epsilon_q)^\top$ are assumed to be independent of $(\mathbf{X}, U)$.

The kth equation for $k = 1, \dots, q$ in (1) implies that

$$Y_k - E(Y_k|U) = \{\mathbf{X} - E(\mathbf{X}|U)\}^\top \boldsymbol{\beta}_0 + \epsilon_k.$$

The conditional expectations, namely, $s_{Y_k}(u) := E(Y_k|U = ut)$ for $k = 1, \dots, q$ and $s_{X_s}(u) := E(X_s|U = u)$ for $s = 1, \dots, p$, can be estimated by a local linear (LL) regression smoother using n *iid* observations, a univariate nonnegative symmetric kernel $K(\cdot)$, and a bandwidth h_1. Denote their local linear estimates by $\widehat{s}_{Y_k}(\cdot)$ and $\widehat{s}_{X_s}(\cdot)$. Then, the parameter of interest $\boldsymbol{\beta}_0$ can be estimated by the method of maximum likelihood based on a kernel density estimate as

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^n \ln\left\{\widehat{f}_{nh}\left(\tilde{Y}_{1i} - \tilde{\mathbf{X}}_i^\top \boldsymbol{\beta}, \dots, \tilde{Y}_{qi} - \tilde{\mathbf{X}}_i^\top \boldsymbol{\beta}\right)\right\},$$

where $\tilde{Y}_{ki} := Y_{ki} - \widehat{s}_{Y_k}(U_i)$, $\tilde{X}_{si} := X_{si} - \widehat{s}_{X_s}(U_i)$, $\tilde{\mathbf{X}}_i = \tilde{X}_{1i}, \dots, \tilde{X}_{pi})^\top$, and

$$\widehat{f}_{nh}(t_1, \dots, t_q) := \frac{1}{nh^q} \sum_{i=1}^n \prod_{k=1}^q K\left(\frac{\tilde{Y}_{ki} - \tilde{\mathbf{X}}_i^\top \boldsymbol{\beta} - t_k}{h}\right) + c_n$$

is the joint density estimator of $f(\epsilon_1, \dots, \epsilon_q)$ using the product kernel based on $K(\cdot)$, a bandwidth $h \neq h_1$ and a positive sequence $c_n \to 0$ as $n \to \infty$.

In practice, $\widehat{\boldsymbol{\beta}}$ can be computed via an iterative algorithm. The details for implementing the estimation procedure are as follows:

(1) The Gaussian kernel $K(t) = \exp(-t^2/2)/\sqrt{2\pi}$.
(2) The bandwidth $h_1 = \widehat{\sigma} U n^{-1/3}$ for LL estimation, where $\widehat{\sigma}_U$ is the sample standard deviation of U.
(3) The bandwidth

$$h = (\log n/n)^{-1/(q+6)} \left(\prod_{k=1}^q \widehat{\sigma}_{\widehat{\epsilon}_k}\right)^{1/q}$$

for the joint density estimation, where $\widehat{\sigma}_{\widehat{\epsilon}_k}$ is the square root of 10% trimmed variance of $\widehat{\epsilon}_k = \tilde{Y}_{ki} - \tilde{\mathbf{X}}^\top \widehat{\boldsymbol{\beta}}_*$ and $\widehat{\boldsymbol{\beta}}_*$ is the estimate obtained from the previous iteration.

(4) The sequence $c_n = n^{-1000}$.

It is demonstrated that $\widehat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent and asymptotically normal. A testing procedure for the linear hypothesis of $\boldsymbol{\beta}_0$ is also proposed. Simulation results indicate that the estimator performs well when the joint density $f(\epsilon_1, \ldots, \epsilon_q)$ is multimodal, asymmetric, or heavy-tailed. *Masayuki Hirukawa*

References

1. L. Breiman, J.H. Friedman, Predicting multivariate responses in multiple linear regression, J. R. Stat. Soc. Ser. B Stat. Methodol. 59 (1997) 3–54. MR1436554
2. Y.C. Chang, A note on the covariance matrix of the maximum likelihood estimator in constrained multivariate linear regression, J. Econometrics 28 (1985) 247–252. MR0799306
3. X. Chen, X. Ma, W. Zhou, Kernel density regression, J. Statist. Plann. Inference 205 (2020) 318–329. MR4011640
4. J. Fan, I. Gijbels, Local polynomial modelling and its applications, in: Monographs on Statistics and Applied Probability, vol. 66, Chapman & Hall, London, London, 1996. MR1383587
5. W. Härdle, H. Liang, J. Gao, Partially linear models, in: Contributions to Statistics, Physica-Verlag, Heidelberg, 2000. MR1787637
6. Z. Hu, N. Wang, R.J. Carroll, Profile-kernel versus backfitting in the partially linear models for longitudinalclustered data, Biometrika 91 (2004) 251–262. MR2081299
7. G. Li, S. Feng, H. Peng, A profile-type smoothed score function for a varying coefficient partially linear model, J. Multivariate Anal. 102 (2011) 372–385. MR2739122
8. H. Lian, Asymptotics of the non-parametric function for B-splines-based estimation in partially linear models, Int. Stat. Rev. 88 (2020) 142–154. MR4088014
9. H. Liang, W. Hardle, R.J. Carroll, Estimation in a semiparametric partially linear errors-in-variables model, Ann. Statist. 27 (1999) 1519–1535. MR1742498
10. H. Liang, Y. Qin, X. Zhang, D. Ruppert, Empirical likelihood-based inferences for generalized partially linear models, Scand. J. Stat. 36 (2009) 433–443. MR2549703
11. X. Lin, R.J. Carroll, Semiparametric estimation in general repeated measures problems, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (2006) 69–88. MR2212575
12. Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61 (1982) 405–415. MR0679685
13. C.F. Manski, Maximum score estimation of the stochastic utility model of choice, J. Econometrics 3 (1975) 205–228. MR0436905
14. C.F. Manski, Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator, J. Econometrics 27 (1985) 313–333. MR0788628
15. A.J. Rothman, E. Levina, J. Zhu, Sparse multivariate regression with covariance estimation., J. Comput. Graph. Statist. 19 (2010) 947–962. MR2791263
16. D.W. Scott, Multivariate density estimation, in: Wiley Series in Probability and Statistics, second ed., John Wiley & Sons, Inc., Hoboken, NJ, 2015, Theory, practice, and visualization. MR3329609
17. R.J. Serfling, Approximation theorems of mathematical statistics, in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1980. MR0595165
18. R. Varadhan, P. Gilbert, Bb: An r package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function, J.

Stat. Softw. 32 (2009) 1–26.
19. A.H. Welsh, T.W. Yee, Local regression for vector responses, J. Statist. Plann. Inference 136 (2006) 3007–3031. MR2281238
20. Y. Yang, G. Li, H. Lian, Nonconcave penalized estimation for partially linear models with longitudinal data, Statistics 50 (2016) 43–59. MR3435168
21. W. Yao, Z. Zhao, Kernel density based linear regression estimate, Comm. Statist. Theory Methods 42 (2013) 4499–4512. MR3171013
22. J. You, Y. Zhou, G. Chen, Statistical inference for multivariate partially linear regression models, Canad. J. Statist. 41 (2013) 1–22. MR3030782
23. A. Yuan, Semiparametric inference with kernel likelihood, J. Nonparametr. Stat. 21 (2009) 207–228. MR2488155
24. A. Yuan, J.G.D. Gooijer, Semiparametric regression with kernel error model, Scand. J. Stat. 34 (2007) 841–869. MR2396942
25. J. Zhang, Partial linear models with general distortion measurement errors, Electron. J. Stat. 13 (2019) 5360–5414. MR4047590
26. J. Zhang, Z. Feng, P. Xu, Estimating the conditional single-index error distribution with a partial linear mean regression, Test (2015) 61–83. MR3314574
27. J. Zhang, B. Lin, Z. Feng, Conditional absolute mean calibration for partial linear multiplicative distortion measurement errors models, Comput. Statist. Data Anal. 141 (2020) 77–93. MR3980098
28. Y. Zhang, L. Zhu, Semiparametric estimation of multivariate partially linear models, J. Stat. Comput. Simul. 87 (2017) 2115–2127. MR3656094
29. L. Zhou, H. Lin, K. Chen, H. Liang, Efficient estimation and computation of parameters and nonparametric functions in generalized semi/non-parametric regression models, J. Econometrics 213 (2019) 593–607. MR4023924

Note: This list reflects references listed in the original paper as accurately as possible with no attempt to correct errors.