Operations Research and Fuzziology 运筹与模糊学, 2023, 13(5), 5266-5274
Published Online October 2023 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2023.135528

部分线性模型的高斯径向基函数估计

胡怀青

贵州大学数学与统计学院，贵州 贵阳

收稿日期：2023年8月14日；录用日期：2023年10月12日；发布日期：2023年10月23日

摘　要

部分线性模型是一种常用的现代统计模型，其同时具备参数与非参数回归的优点。我们基于高斯径向基函数估计部分线性模型的非线性部分，并给出估计过程中的超参数选择方法。在模拟仿真与实证分析中将高斯径向基函数与B样条进行对比，发现高斯径向基函数在部分线性模型中可以成为B样条的一种替代方法。

关键词

高斯径向基函数，部分线性模型，B样条

Gaussian Radial Basis Function Estimation of Partially Linear Model

Huaiqing Hu

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Aug. 14th, 2023; accepted: Oct. 12th, 2023; published: Oct. 23rd, 2023

Abstract

The partially linear model is a commonly used modern statistical model with the advantages of both parametric and nonparametric regression. We estimate the nonlinear part of the partially linear model based on the Gaussian radial basis function, and give the hyperparameter selection method in the estimation process. The Gaussian radial basis function is compared with the B-spline in simulation and empirical analysis, and it is found that the Gaussian radial basis function can be an alternative method to the B-spline in the partially linear model.

Keywords

Gaussian Radial Basis Function, Partially Linear Models, B-Spline

文章引用: 胡怀青. 部分线性模型的高斯径向基函数估计[J]. 运筹与模糊学, 2023, 13(5): 5266-5274.
DOI: 10.12677/orf.2023.135528

1. 引言

部分线性模型是一种常用的统计模型，其将参数模型与非参数模型的优势相结合，用于解决具有复杂关系的数据分析问题。在部分线性模型中，解释变量可以包含线性部分和非线性部分，通过将线性和非线性部分相结合，部分线性模型能够更好地拟合数据并提高预测准确性。部分线性模型形式如下：

$$Y_i = X_i'\beta + g\left(Z_i\right) + \epsilon_i,\ i = 1,2,\cdots,n$$

其中 Y_i 是响应变量， X_i 是线性部分解释变量， β 是线性部分系数。 Z_i 是非线性部分解释变量， g 为未知函数。 ϵ_i 为误差项， ϵ_i 与 X_i ， Z_i 独立。令 $\epsilon = \left(\epsilon_1,\cdots,\epsilon_n\right)'$ ， ϵ 满足 $Var\left(\epsilon\right) = \rho^2 I_n$ ， I_n 为 n 阶单位阵。

由于部分线性模型性质良好，其相关应用与理论受到了众多关注。实际应用方面，1986 年，Engle 等[1]首次将部分线性模型应用于用电需求相关问题。1988 年，Speckman [2]运用该模型探究了漱口水的效果。1999 年，Schmalensee 和 Stoker [3]运用该模型研究了美国家庭汽油消费情况。理论性质方面，1986 年，Heckman [4]基于光滑样条估计部分线性模型非参数部分。1988 年，Robinson [5]研究了该模型非参数部分的核估计方法，并给出估计方法的大样本性质。2006 年，Ma 等[6]研究了异方差下该模型的加权估计方法，并证明了相合性。2021 年，Liu 和 Yin [7]研究了时间序列数据下部分线性模型的样条估计方法。Zhong 等[8]基于深度神经网络估计部分线性 Cox 模型中的非线性部分，并给出了估计方法的大样本性质。Rodríguez 等[9]研究了具有单调性约束部分线性模型的稳健估计方法，并通过蒙特卡洛随机模拟验证了方法的有效性。

经典的估计方法主要使用核与样条。我们考虑了一种新的基函数：高斯径向基函数。由于高斯径向基函数的良好性质，有许多学者对其进行了研究。2008 年，Ando 等[10]使用信息准则选取了正则化径向基函数神经网络的超参数。2014 年，Lei 等[11]利用局部 Rademacher 复杂度研究了径向基函数网络的泛化性能，获得了新的估计误差边界。2021 年，Krzyżak 和 Niemann [12]研究了归一化径向基函数网络的收敛性与收敛速率，讨论了径向基函数和网络参数的选择。2022 年，Sosa 与 Buitrago [13]运用径向基函数逼近时变系数模型的系数函数，并给出频率方法与贝叶斯估计方法。

本文的组织如下：第二章介绍基于高斯径向基函数的估计方法。第三章进行模拟仿真，并与 B 样条进行对比。第四章，将估计方法应用于实际数据集。第五章，对全文总结。

2. 估计方法

2.1. 模型估计方法

使用高斯径向基函数估计部分线性模型就是使用一组高斯径向基函数的线性组合去逼近未知函数 g。一组高斯径向基函数形式如下：

$$\exp\left(-\frac{\left(z-c_1\right)^2}{2\sigma^2}\right),\cdots,\exp\left(-\frac{\left(z-c_k\right)^2}{2\sigma^2}\right),$$

其中 $\left\{c_j\right\}_{j=1}^{k}$ 为高斯径向基函数的中心参数， σ 为形状参数。由此有

$$g(z) \approx \sum_{j=1}^{k}\gamma_j \exp\left(-\frac{\left(z-c_j\right)^2}{2\sigma^2}\right).$$

则有

$$Y_i \approx X_i'\beta + \sum_{j=1}^{k}\gamma_j \exp\left(-\frac{\left(Z_i - c_j\right)^2}{2\sigma^2}\right) + \epsilon_i,\ i = 1, 2, \cdots, n,$$

通过最小二乘求解参数，即

$$\min_{\beta,\gamma,\sigma}\left(Y_i - X_i'\beta - \sum_{j=1}^{k}\gamma_j \exp\left(-\frac{\left(Z_i - c_j\right)^2}{2\sigma^2}\right)\right)^2 \tag{1}$$

其中 $\gamma = \left(\gamma_1, \cdots, \gamma_k\right)$。以上的优化问题是一个非凸优化。记 $\mathbb{X}$ 为(1)的设计阵，$\mathbb{Y} = \left(Y_1, \cdots, Y_t\right)'$。

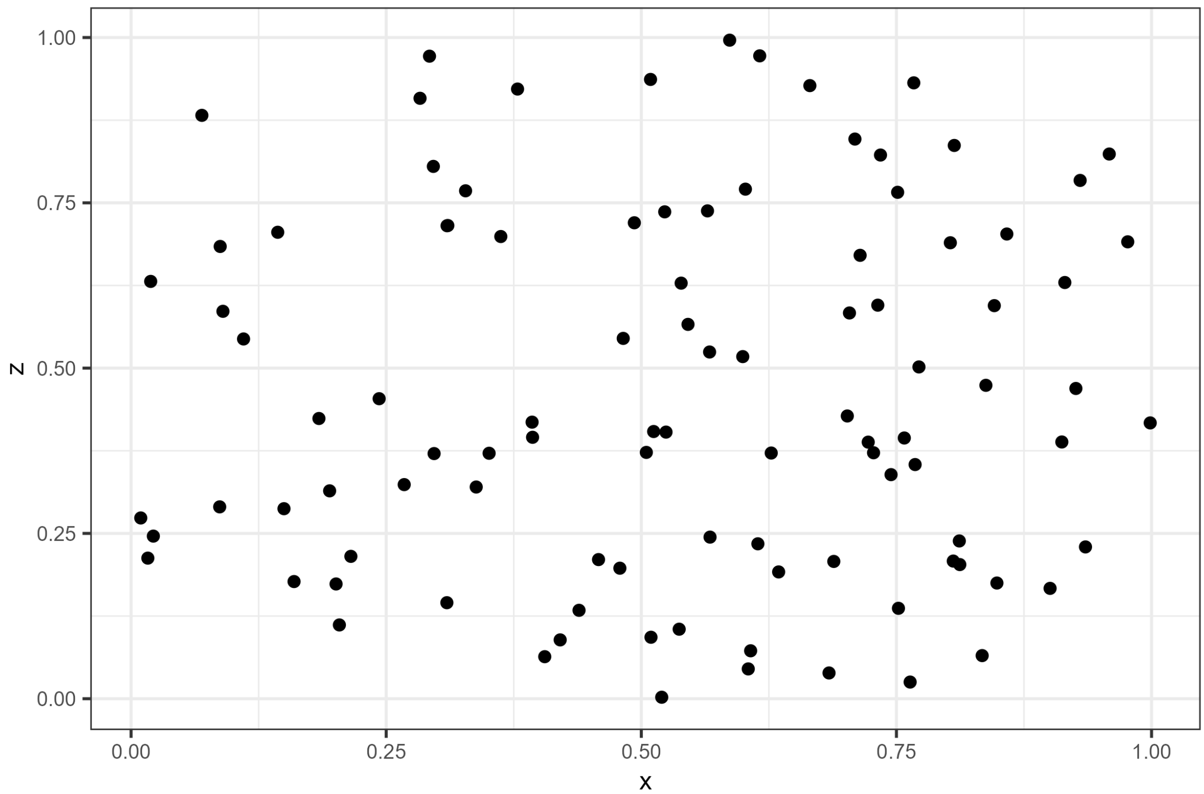

Figure 1. Diagram of non-sparse point
图 1. 数据点均匀分布图

2.2. 模型参数选择

目标函数(2)为一个非凸优化问题，无法直接应用优化算法求解。我们将该问题分三步求解，最终将转化为凸优化问题。

首先进行中心参数选择。中心参数选择分为两个部分：中心位置与中心个数。中心参数位置的选择方法主要有两类。第一类不依赖于 $\{Z_i\}_{i=1}^{n}$ 的分布，如等间距点与分位数点。这种方法是多项式样条的常用取点方式，高斯径向基函数在一维下也是适用的。这类取点方法有着很多优点，如计算成本低，理论性质好等。等间距点适用的 $\{Z_i\}_{i=1}^{n}$ 如图 1，在 z 轴均匀分布。

第二类方法依赖于 $\{Z_i\}_{i=1}^{n}$ 的分布，如 Kmeans 聚类以及在 $\{Z_i\}_{i=1}^{n}$ 中随机抽取。第二类方法显然比第

一类更加适合在样本点稀疏分布时取点。但其有着缺陷：随着中心参数的增加，中心参数之间的距离变化不易描述，甚至 Kmeans 聚类无法保证中心间的最小距离趋于 0，且 Kmeans 聚类计算量过大。等间距点适用的 $\{Z_i\}_{i=1}^{n}$ 如图 2，在 z 轴非均匀分布。本文主要考虑一维问题，所以采用等间距点。

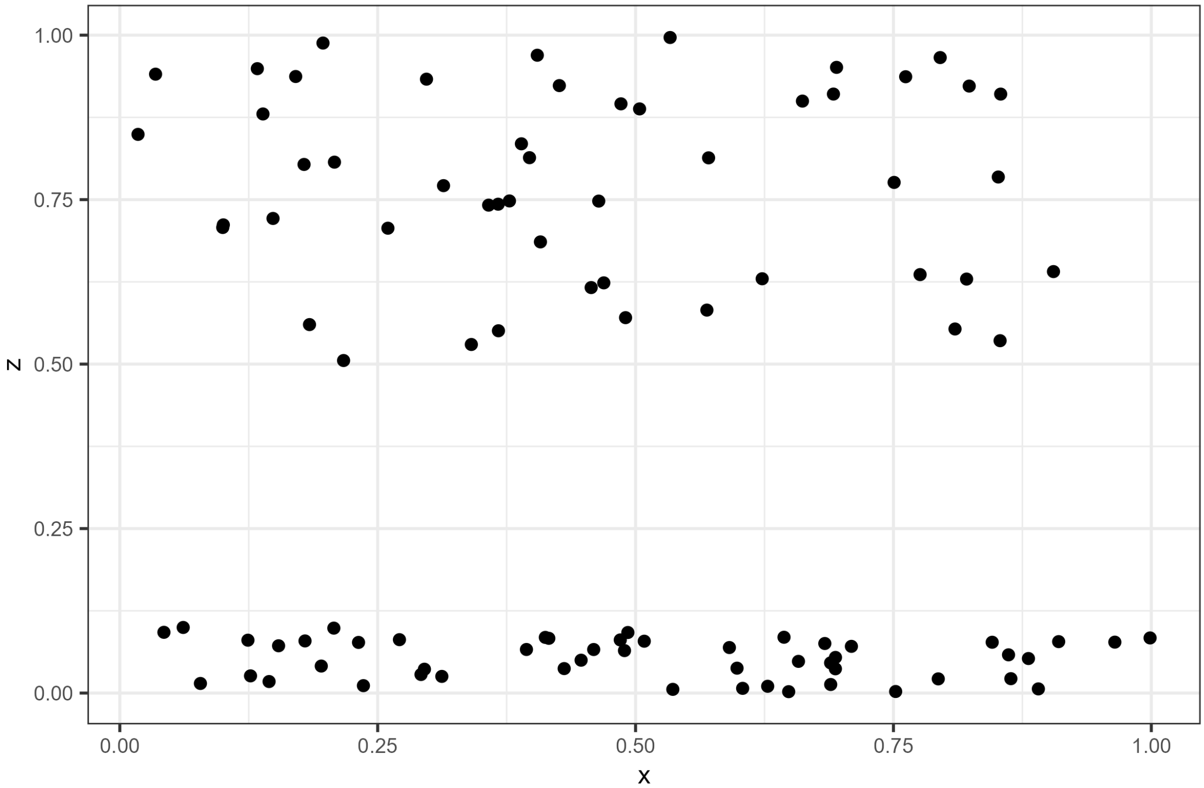

Figure 2. Diagram of sparse point
图 2. 数据点非均匀分布图

中心个数过多会产生过拟合，反之则欠拟合。样条类的方法根据大样本性质可以直接使用 $n^{1/5}$ 作为节点个数。高斯径向基函数无渐近理论参考，我们采用数据驱动的中心个数选择方法。中心个数选择通过如下信息准则

$$\mathrm{AIC} = \log\left(\frac{\mathrm{RSS}}{n}\right) + \frac{2K}{n};$$
$$\mathrm{AICc} = \log\left(\frac{\mathrm{RSS}}{n}\right) + \frac{2K}{n} + \frac{2(K+1)(K+2)}{n(n-K-2)};$$
$$\mathrm{BIC} = \log\left(\frac{\mathrm{RSS}}{n}\right) + \log(n)\frac{K}{n},$$

其中 n 为样本量，K 为(1)待估参数个数，RSS 为(1)的残差平方和最小值。使用了 AIC，AICc，BIC 等方法进行参数选择，其数值结果显示 AIC 是一种简单且效果良好的方法，因此采用 AIC 准则。信息准则是根据样本内的误差进行模型选择。若要使用样本外误差进行模型选择，需使用交叉验证以及广义交叉验证，可以参考[14]。

其次，位置参数选择。高斯径向基函数的逼近效果依赖于形状参数 σ。选择方法可主要分为三类，第一类取决于节点的位置，可参考[15]。第二类由数据驱动选择，如 AIC 与交叉验证等。第三类同时依赖于位置与个数，可参考[16]。

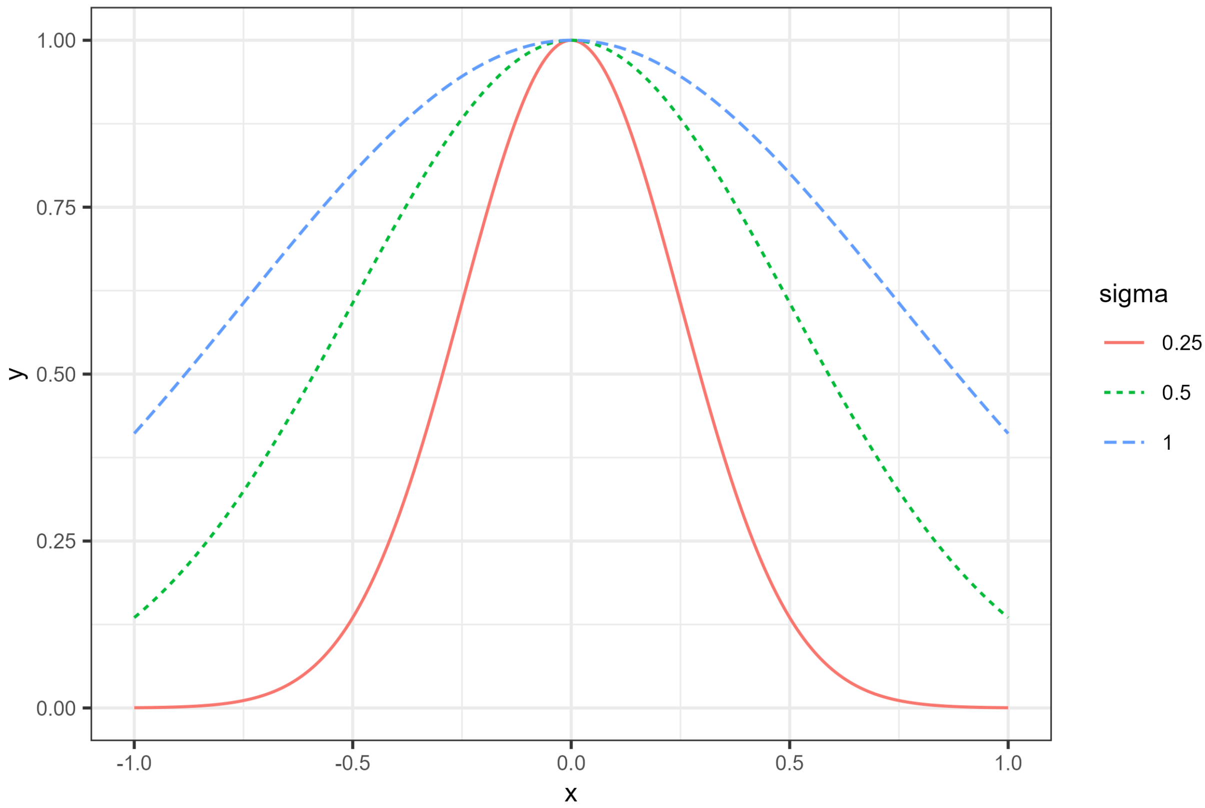

Figure 3. Diagram of Gaussian radial basis functions under different σ
图 3. 不同 σ 下的高斯径向基函数

以上三类方法有一种共同思想，当中心点稀疏时，σ 应越大，从而使得高斯径向基函数趋于平稳，影响范围变大，基函数不同 σ 下的变化趋势如图 3。结合以上思想，本文采取 $\sigma = M_0/\sqrt{k}, M_0 > 0$。对于 M_0 的选择可以考虑 AIC，BIC，交叉验证等方法。本文为减少计算量直接采用 $M_0 = 1$。

最后求解问题(1)的系数。在给定中心参数与未知参数后，问题(1)为普通最小二乘问题，即有显示解。

3. 模拟仿真

本节中，我们对比基于高斯径向基函数与基于 B 样条的估计方法。考虑如下模型：

$$Y_i = X_i\beta + g(Z_i) + \epsilon_i,\ i = 1,.2,\cdots,n \tag{2}$$

X_i 与 Z_i 服从 $U(0,1)$，ϵ_i 服从 $N(0,0.25), \beta = 2$。$g(z)$ 考虑如下 5 种函数：$x, x^2, \sin(x), \exp(-x^2)$ 以及 $\log(1+x)$，将其分别简记为 $f_1,\cdots,f_5$。

按照 X_i, Z_i, ϵ_i 的分布进行独立抽样，样本量 n 为 100，200，300，400，500。再利用模型(2)计算 Y_i，从而得到观测数据 (X_i, Z_i, Y_i)。基于该观测数据分别使用高斯径向函数与 B 样条估计未知函数 $g(z)$。取[0, 1]上等间距点 $\{Xtest_j\}_{j=1}^{100}, \{Ztest_j\}_{j=1}^{100}$，基于 $(Xtest_j, Ztest_j)_{j=1}^{100}$ 与模型(2)计算 $\{Ytest_j\}_{j=1}^{100}$，但计算过程中不加入扰动。使用 $(Xtest_j, Ztest_j, Ytest_j)_{j=1}^{100}$ 与 MAE 比较两种方法，MAE 公式如下：

$$\text{MAE} = \frac{\sum_{j=1}^{100}\left|\hat{Y}test_j - Ytest_j\right|}{100}$$

其中 $\hat{Y}test_j$ 为估计方法的预测值。

Table 1. Mean of 500 MAEs of Gaussian radial basis functions
表 1. 高斯径向基函数 500 次 MAE 均值

函数样本量	f_1	f_2	f_3	f_4	f_5
100	0.089	0.105	0.090	0.092	0.090
200	0.065	0.073	0.065	0.058	0.062
300	0.054	0.060	0.049	0.049	0.050
400	0.046	0.051	0.045	0.042	0.043
500	0.044	0.046	0.040	0.038	0.040

Table 2. Mean of 500 MAEs of Bspline
表 2. B 样条 500 次 MAE 均值

函数样本量	f_1	f_2	f_3	f_4	f_5
100	0.108	0.110	0.112	0.113	0.106
200	0.075	0.076	0.076	0.073	0.075
300	0.061	0.062	0.060	0.061	0.059
400	0.051	0.052	0.054	0.052	0.051
500	0.046	0.047	0.046	0.047	0.046

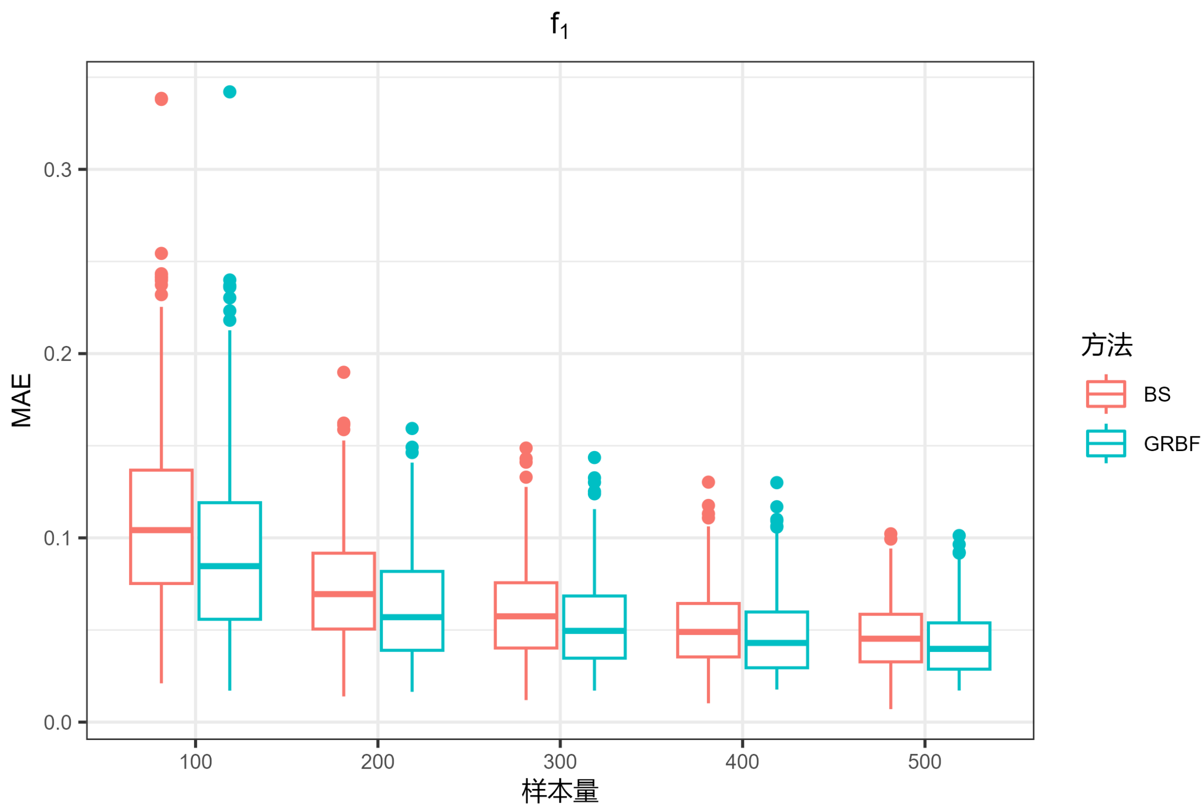

Figure 4. Box plot of 500 MAEs of f_1
图 4. f_1 500 次 MAE 的箱线图

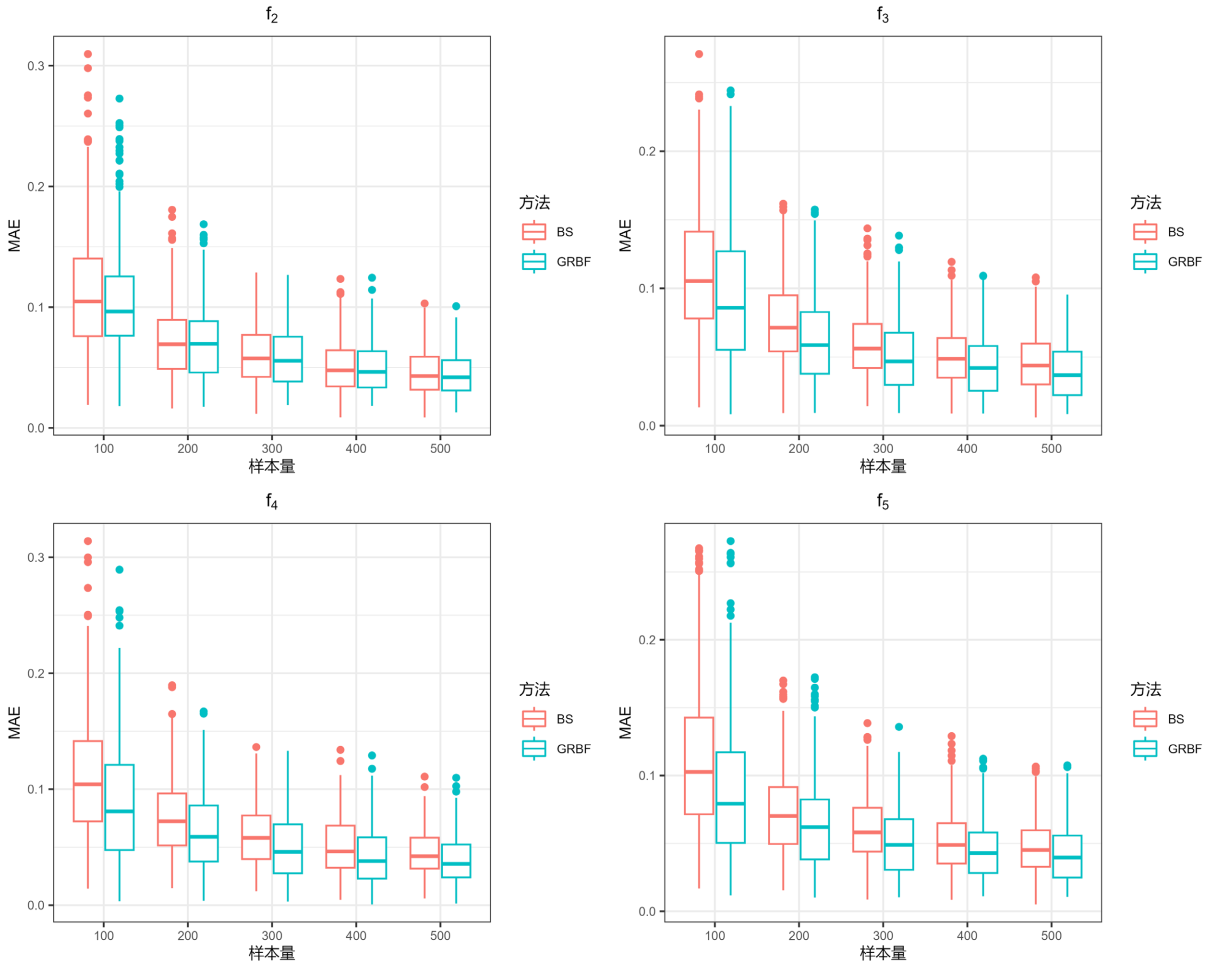

Figure 5. Box plot of 500 MAEs of $f_2,\cdots,f_5$

图 5. $f_2,\cdots,f_5$ 500 次 MAE 的箱线图

B 样条节点与高斯径向基函数中心都采用等间距点，个数范围为 2~10，使用 AIC 准则选取节点个数。为减少模拟随机性，重复模拟 500 次。使用 500 次试验 MAE 的均值衡量估计效果，试验结果如表 1 与表 2。我们发现高斯径向基函数在 5 个函数上都优于 B 样条，且在 f_4 上差距表现最为明显。同时基于高斯径向基函数的估计方法随着样本量的增大，MAE 逐渐减小。这说明估计方法具有一致性。图 4 与图 5 为 5 个函数下 500 次 MAE 的箱线图，高斯径向基函数在 5 个函数上 MAE 的稳定性都高于 B 样条，这说明未知函数估计上高斯径向基函数更为稳定。高斯径向基函数在 $f_1,\cdots,f_5$ 上表现也有所不同，f_1, f_2 的异常值更多更大，f_3, f_4, f_5 异常值相比则较少，f_5 在样本量少时稳定性最高。

4. 实证研究

实证研究使用波士顿房价数据集，该数据集可以从 R 包 MASS 中获取。刘志伟和夏志明[17]也使用该数据集研究了半线性模型。我们基于数据集中 medv，rm，lstat，ptratio，dis 等变量进行数据分析，其中 medv 为被解释变量，其余为解释变量。medv 为房价。

各变量间的散点图如图 6。rm 为每间住宅的平均房间数，容易发现 rm 与 medv 的散点图有线性的趋势，且呈现房间数上升房价上升的规律，这符合常理。lstat 为低收入群比例，lsta 与 medv 的散点图依然具有线性的趋势，且呈现低收入群比例上升房价降低的规律，数据是符合现实的。ptratio 为城镇中的教

师学生比例，ptratio 与 medv 的散点图任然呈现线性关系，但随着 ptratio 变化的上升下降趋势并不显著。dis 为距离 5 个波士顿的就业中心的加权距离，dis 与 medv 的散点图既具有线性趋势，但又可能具有更复杂的函数关系。同时具有距离就业中心越远，房价越高的倾向。

由于 dis 与 medv 不一定呈现线性关系，计算相关系数用于判断变量间线性相关程度。相关系数表如表 3。容易发现 medv 与 rm，lstat，ptratio，相关系数较大，结合散点图判断 medv 与这些变量之间呈线性关系。medv 与 dis 之间相关系数为 0.250，数值较小，我们结合散点图推断 medv 与 dis 之间为非线性关系。

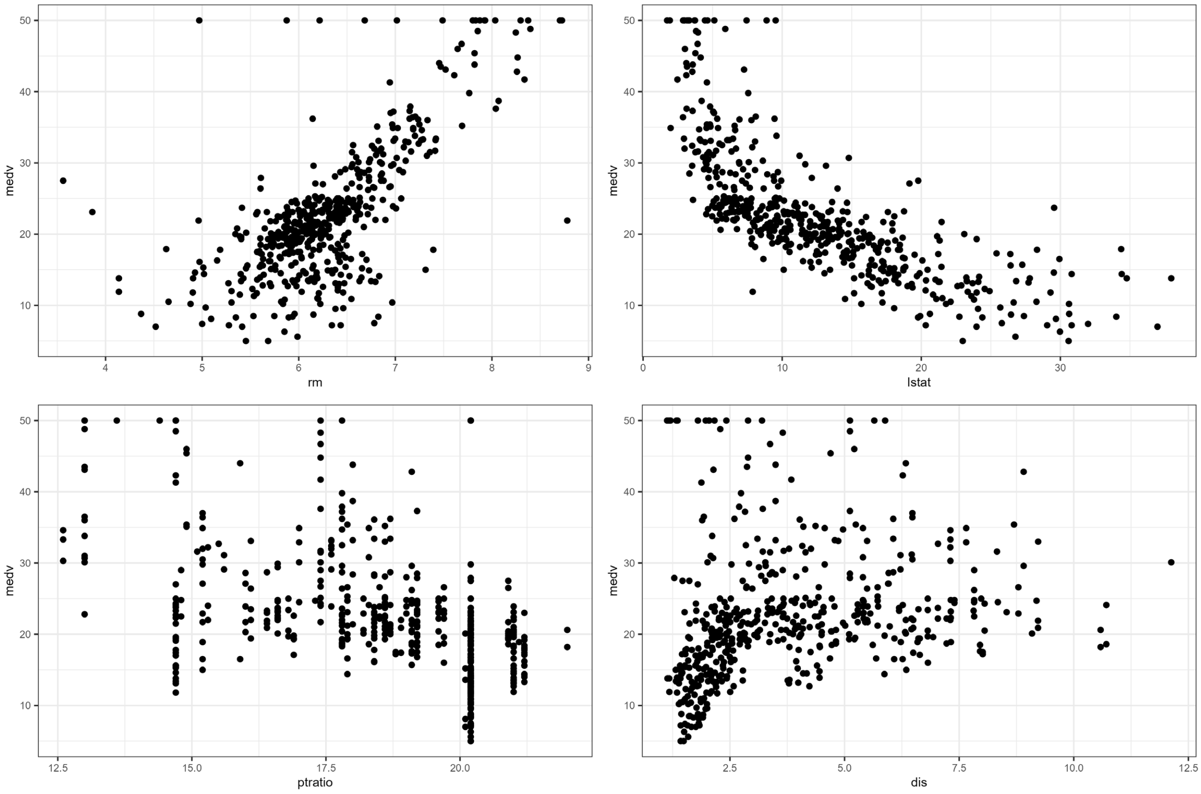

Figure 6. Scatterplot between variables
图 6. 变量间散点图

Table 3. Table of correlation coefficients
表 3. 相关系数表

	rm	lstat	ptratio	dis
medv	0.695	−0.738	−0.508	0.250

根据变量间的线性与非线性关系，我们考虑建立如下部分线性模型：

$$medv_i = \beta_0 rm_i + \beta_1 lstat_i + \beta_2 ptratio_i + g\left(dis_i\right) + \epsilon_i,\ i = 1, 2, \cdots, 506.$$

使用该模型对比高斯径向基函数与 B 样条，将数据 10 折，计算 MAE 用于对比两种方法的数值效果。高斯径向基函数与 B 样条的 10 折 MAE 如表 4，均值分别为 0.086 与 0.087，高斯径向基函数的表现优于 B 样条。这反应了高斯径向基函数在实证有着不错的效果，可以在实际数据中成为 B 样条的替代方法。

Table 4. 10 fold MAE
表 4. 10 折 MAE

	1	2	3	4	5	6	7	8	9	10
BS	0.074	0.063	0.080	0.084	0.096	0.082	0.078	0.083	0.079	0.156
GRBF	0.076	0.063	0.080	0.083	0.094	0.082	0.078	0.083	0.081	0.143

5. 结论

我们使用高斯径向基函数估计了部分线性模型中的非参数部分，分析了高斯径向基函数的超参数选择方法。同时，在 5 种不同的函数下对比了高斯径向基函数与 B 样条的数值效果，在波士顿房价数据集下对比了两种方法的实际表现。我们发现高斯径向基函数都优于 B 样条，由此认为高斯径向基函数可以成为 B 样条在部分线性模型中的一种替代方法。

参考文献

[1] Engle, R.F., Granger, C.W., Rice, J. and Weiss, A. (1986) Semiparametric Estimates of the Relation between Weather and Electricity Sales. *Journal of the American statistical Association*, **81**, 310-320. https://doi.org/10.1080/01621459.1986.10478274

[2] Speckman, P. (1988) Kernel Smoothing in Partial Linear Models. *Journal of the Royal Statistical Society Series B* (*Methodology*), **50**, 413-436. https://doi.org/10.1111/j.2517-6161.1988.tb01738.x

[3] Schmalensee, R. and Stoker, T.M. (1999) Household Gasoline Demand in the United States. *Econometrica*, **67**, 645-662. https://doi.org/10.1111/1468-0262.00041

[4] Heckman, N.E. (1986) Spline Smoothing in a Partly Linear Model. *Journal of the Royal Statistical Society Series B* (*Methodology*), **48**, 244-248. https://doi.org/10.1111/j.2517-6161.1986.tb01407.x

[5] Robinson, P.M. (1988) Root-N-Consistent Semiparametric Regression. *Econometrica*, **56**, 931-954. https://doi.org/10.2307/1912705

[6] Ma, Y. and Carroll, R.J. (2006) Locally Efficient Estimators for Semiparametric Models with Measurement Error. *Journal of the American Statistical Association*, **101**, 1465-1474. https://doi.org/10.1198/016214506000000519

[7] Liu, Y. and Yin, J. (2021) Spline Estimation of Partially Linear Regression Models for Time Series with Correlated Errors. *Communications in Statistics-Simulation and Computation*, 1-15. https://doi.org/10.1080/03610918.2021.1990328

[8] Zhong, Q., Mueller, J. and Wang, J.L. (2022) Deep Learning for the Partially Linear Cox Model. *The Annals of Statistics*, **50**, 1348-1375. https://doi.org/10.1214/21-AOS2153

[9] Rodríguez, D., Valdora, M. and Vena, P. (2022) Robust Estimation in Partially Linear Regression Models with Monotonicity Constraints. *Communications in Statistics-Simulation and Computation*, **51**, 2039-2052. https://doi.org/10.1080/03610918.2019.1691732

[10] Ando, T., Konishi, S. and Imoto, S. (2008) Nonlinear Regression Modeling via Regularized Radial Basis Function Networks. *Journal of Statistical Planning and Inference*, **138**, 3616-3633. https://doi.org/10.1016/j.jspi.2005.07.014

[11] Lei, Y., Ding, L. and Zhang, W. (2014) Generalization Performance of Radial Basis Function Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **26**, 551-564. https://doi.org/10.1109/TNNLS.2014.2320280

[12] Krzyżak, A. and Niemann, H. (2021) Convergence Properties of Radial Basis Functions Networks in Function Learning. *Procedia Computer Science*, **192**, 3761-3767. https://doi.org/10.1016/j.procs.2021.09.150

[13] Sosa, J. and Buitrago, L. (2022) Time-Varying Coefficient Model Estimation through Radial Basis Functions. *Journal of Applied Statistics*, **49**, 2510-2534. https://doi.org/10.1080/02664763.2021.1910938

[14] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning. Springer, New York. https://doi.org/10.1007/978-1-4614-7138-7

[15] Schwenker, F., Kestler, H.A. and Palm, G. (2001) Three Learning Phases for Radial-Basis-Function Networks. *Neural Networks*, **14**, 439-458. https://doi.org/10.1016/S0893-6080(01)00027-2

[16] Haykin, S. (2009) Neural Networks and Learning Machines. Prentice Hall/Pearson, New York.

[17] 刘志伟, 夏志明. 部分线性模型的半线性神经网络估计[J]. 应用概率统计, 2023(2): 218-238.