Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Statistical estimation in partially nonlinear models with random effects

Ye Que, Zhensheng Huang & Riquan Zhang

STATISTICAL THEORY AND RELATED FIELDS, 2017
VOL. 1, NO. 2, 227–233
https://doi.org/10.1080/24754269.2017.1396425

Statistical estimation in partially nonlinear models with random effects

Ye Que[a,b], Zhensheng Huang[a] and Riquan Zhang[c]

[a]School of Science, Nanjing University of Science and Technology, Nanjing, P. R. China; [b]School of Finance, Huainan Normal University, Huainan, P. R. China; [c]School of Statistics, East China Normal University, Shanghai, P. R. China

ABSTRACT
In this article, a partially nonlinear model with random effects is proposed and its new estimation procession is provided. In order to estimate the link function, we propose generalised least square estimate and B-splines estimate methods. Further, we also use the Gauss–Newton method to construct the estimates of unknown parameters. Finally, we also consider the estimation for the variance components. The consistency and the asymptotic normality of the estimator will be proved. Simulated and real examples are given to illustrate our proposed methodology, which shows that our methods give effective estimation.

ARTICLE HISTORY
Received 15 June 2017
Accepted 21 October 2017

KEYWORDS
Asymptotic properties; B-splines method; Gauss–Newton method; mixed-effects models; partially nonlinear models

1. Introduction

We consider a partially nonlinear mixed-effects model:

$$Y_i = g(U_i) + f(X_i, \beta) + Z_i^T b_i + \varepsilon_i,\ i = 1, \ldots, n, \qquad (1)$$

where $g(\cdot)$ is a smooth function, $f(\cdot, \cdot)$ is a pre-specified function, β is a P-dimensional parameter vector, b_i is a $n \times 1$ random effect vector with mean 0 and $\text{var}(b_i) = D$, here D is a $n \times 1$ vector. ε_i is an independent random variable with $E(\varepsilon_i) = 0$ and $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2 > 0$. X_i, U_i and Y_i are random variables, which can be observed. Z_i is an $n \times 1$ vector. We will show the estimation progress of the function $g(\cdot)$, the parameter β , the variance σ_ε^2 and D.

When $g(\cdot) = 0$ and $f(X_i, \beta) = X_i^T \beta$, model (1) is the linear mixed-effects model. Harville (1977) proposed maximum-likelihood approaches to study the variance component estimation of linear models with random effects. Zhong, Fung, and Wei (2002) studied linear mixed models with errors-in-variables and they presented a unified method for the estimation. When $f(X_i, \beta) = 0$, model (1) turns into the non-parametric mixed-effects model; for example, Gu and Ma (2005) proposed penalised least squares method to estimate the non-parametric mixed-effects model. When $g(\cdot) = 0$ and $f(X_i, \beta) = f(X_i^T \beta)$, model (1) turns into the single-index model with random effects; for this, Pang and Xue (2012) studied the single-index models with random effects. In order to estimate the index coefficients and the link function, they proposed a new set of estimating equations modified for the boundary effects and the local linear smoother method. Other literatures about the random effects are, see, for example, Jiang (2007); Ke and Wang (2001) and Lindstrom and Bates (2010). When random variable $b_i = 0$ and $f(X_i, \beta) = X_i^T \beta$, model (1) turns into the partially linear model. Müller and Vial (2009) studied partially linear model selection by the bootstrap, and they proposed a new approach to the selection of partially linear models based on the conditional expected prediction squares loss function. When random variable $b_i = 0$, model (1) turns into the partially nonlinear model. Ai and Mcfadden (1997) studied the estimation of some partially specified nonlinear models. They presented a procedure for analysing a partially specified nonlinear regression model is which the nuisance parameter in an unrestricted function of a subset of regressors.

Li and Nie (2008) studied the efficient statistical inference procedures for partially nonlinear models and their applications. They presented two new estimation procedures to estimate the parameters in the parametric component. They further presented an estimation procedure and a generalised F-test procedure for the non-parametric component in the partially nonlinear models. Song, Zhao, and Wang (2010) studied the sieve least squares estimation for partially nonlinear models and they proposed a sieve least squares method to estimate the parameters of the parametric part and the non-parametric part. Li and Mei (2013) studied estimation and inference for varying coefficient partially nonlinear models. They proposed a profile nonlinear least squares estimation procedure for the parameter vector and the coefficient function vector of the varying coefficient partially nonlinear model, and they further presented the generalised-likelihood ratio test to verify whether the varying coefficients are a constant. Xiao, Tian, and Li (2013) studied empirical likelihood-based inference for parametric and non-parametric functions in partially nonlinear models. They obtained the maximum empirical likelihood estimation of parameter in

CONTACT Zhensheng Huang stahzs@126.com, zshuang@njnust.edu.cn

nonlinear function and non-parametric function by empirical likelihood ratio function. Other relevant literatures are, see, for example, Bates and Watts (1988) and Fan (1993). However, the method to deal with partially nonlinear model with random effects is still lacking. In this article, such a model will be studied.

In addition, our another motivation is mainly from analysing the famous Boston housing price data (Harrison & Rubinfeld, 1978), in which the relationship between the variable AGE (the proportion of owner-occupied units built prior to 1940) and response variable MEDV (the median value of owner-occupied home) may be nonlinear and be described by our proposed function $f(\cdot, \cdot)$. More details can be seen in the real data example.

In this article, our purpose is to study the estimation progress in model (1). We use B-spline to estimate the function $g(\cdot)$ and use Gauss–Newton method to construct the parameter estimation in the part of nonlinear function. We also construct the estimation for σ_ε^2 and σ_b^2. Our algorithm is stable in numerical. Simulation and real data examples will show the performance of the methodology.

We organise the rest of the paper as follows: In Section 2, we demonstrate the estimation methodology and the asymptotic properties of the proposed estimation. In Section 3, we present the result of simulation study and real data example. The conditions and the proofs of theorem are shown in the Appendix.

2. Estimation methodology

In this section, we focus on the estimation of the link function $g(\cdot)$, the parameter β, the variance σ_ε^2 and σ_b^2 in this section. By using B-spline, the function $g(\cdot)$ can be approximated as $g(u) = \sum_{j=1}^{q} c_j B_j(u) = B(u)c$. So the estimation can be expressed as $\hat{g}(u) = \sum_{j=1}^{q} \hat{c}_j B_j(u)$, where $\{B_j(u)\}_{j=1}^{q}$ is a B-spline basis, $B(u) = (B_1(u), \ldots, B_q(u))$, $c = (c_1, \ldots, c_q)^T$. For our method, we need the initial value β^0 which can be obtained by fitting the linear model. Now, we will show the estimation process of $g(\cdot)$, β, σ_ε^2 and σ_b^2.

Step 1. Obtain the estimation of $g(\cdot)$.

First, we let $Y = (Y_1, \ldots, Y_n)^T$, $B = (B(U_1), \ldots, B(U_n))^T$, $f(X, \beta) = (f(X_1, \beta), \ldots, f(X_n, \beta))^T$, $b = (b_1^T, \ldots, b_n^T)^T$, $b_i = (b_{i1}, \ldots, b_{in})^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ and $Z = \mathrm{diag}(Z_1, \ldots, Z_n)$, $Z_i = (Z_{i1}, \ldots, Z_{in})^T$, where diag denotes diagonal matrix. Then model (1) can be expressed in the following form:

$$Y = Bc + f(X, \beta) + Z^T b + \varepsilon. \tag{2}$$

For the initial value β^0, we let $Y^* = Y - f(X, \beta^0)$, then (2) can be turned into $Y^* = Bc + Z^T b + \varepsilon$. We can obtain $\hat{c}$ by minimising the following problem: $(Y^* - Bc)^T V^{-1} (Y^* - Bc)$, where $V = \mathrm{var}(Z^T b + \varepsilon) = ZDZ^T + \sigma_\varepsilon^2 I_n$. Then we will get the estimation of c: $\hat{c} = (B^T V^{-1} B)^{-1} B^T V^{-1} Y^*$. Let $S = B(B^T V^{-1} B)^{-1} B^T V^{-1}$; so the estimation of $g(\cdot)$ is $\hat{g}_0(U) = B\hat{c} = B(B^T V^{-1} B)^{-1} B^T V^{-1} Y^* = SY^*$. We will give the estimations of D and σ_ε^2 in Step 3, so we treat V as known when we construct the estimation of $g(\cdot)$.

Step 2. Obtain the estimation of β.

Next, we discuss the estimation of β. For convenience, we let $\eta(\beta) = \{f(X_1, \beta), \ldots, f(X_n, \beta)\}^T$, $Y_* = Y - \hat{g}_0(U)$, $\beta = (\beta_1, \ldots, \beta_P)^T$ and $\beta^0 = (\beta_1^0, \ldots, \beta_P^0)^T$, so $Y_* = f(X, \beta) + \epsilon = \eta(\beta) + \epsilon$, where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$, $\epsilon_i = Z_i^T b_i + \varepsilon_i$, $E(\epsilon) = 0$, $\mathrm{var}(\epsilon) = V$. In a neighbourhood of β^0, we make the first-order Taylor expansion for $f(X_i, \beta)$: $f(X_i, \beta) \approx f(X_i, \beta^0) + V_{i1}(\beta_1 - \beta_1^0) + V_{i2}(\beta_2 - \beta_2^0) + \cdots + V_{iP}(\beta_P - \beta_P^0)$, where $V_{ip} = \frac{\partial f(X_i, \beta)}{\partial \beta_p} |_{\beta^0}$, $p = 1, \ldots, P$. We merge all the observations and get

$$\eta(\beta) \approx \eta(\beta^0) + V_0(\beta - \beta^0). \tag{3}$$

Here V_0 is an $n \times P$ order derivative matrix, in which the element is $\{V_{ip}\}$. Or equivalent, the residuals $\epsilon = Y_* - \eta(\beta)$ can be approximated as $\epsilon \approx Y_* - [\eta(\beta^0) + V_0\delta] = Z^0 - V_0\delta$, where $Z^0 = Y_* - \eta(\beta^0)$, $\delta = \beta - \beta^0$. To minimise the following problem: $(Z^0 - V_0\delta)^T V^{-1} (Z^0 - V_0\delta)$, we can get the estimation of β: $\hat{\beta}^* = (V_0^T V^{-1} V_0)^{-1} (V_0^T V^{-1} Z^0 + V_0^T V^{-1} V_0 \beta^0)$, where the initial value β^0 can be obtained by fitting the linear model. We will give the estimations of D and σ_ε^2 in Step 3, so we treat V as known when we construct the estimation of β. Let $\beta^0 = \hat{\beta}^* / \|\hat{\beta}^*\|$ and iterative until convergence, then we can get the final estimate $\hat{\beta}$ of β. With $\hat{\beta}$, we can get the final estimators of g: $\hat{g}(U) = B(B^T V^{-1} B)^{-1} B^T V^{-1} [Y - f(X, \hat{\beta})]$.

Step 3. Obtain the estimation of σ_ε^2 and σ_b^2.

Finally, we describe the estimation of variance components. This estimation method is similar to Pang and Xue (2012). We assume the covariance matrix of model (1) is $V = \sigma_b^2 J_n J_n^T + \sigma_\varepsilon^2 I_n$, where $J_n = (1, \ldots, 1)^T$ is an $n \times 1$ vector about ones and I_n is a $n \times 1$ identity matrix. If the random effects $b_i \sim N(0, D)$ and the error term $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, then the observation $Y \sim N(f(X, \beta) + g(U), V)$. We replace β and $g(\cdot)$ with their final estimators $\hat{\beta}$ and $\hat{g}(\cdot)$, respectively. We write the Gaussian likelihood about σ_b^2 and σ_ε^2 as

$$-n(n-1)\log(\sigma_\varepsilon^2) - n\log\left(\sigma_\varepsilon^2 + n\sigma_b^2\right) - \frac{n}{\sigma_\varepsilon^2 + n\sigma_b^2}(\bar{Y} - \bar{\hat{f}} - \bar{\hat{g}})^2$$

$$-\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\{Y_i - f(X_i, \hat{\beta}) - \hat{g}(U_i) - (\bar{Y} - \bar{\hat{f}} - \bar{\hat{g}})\}^2,$$

where $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$, $\bar{\hat{g}} = n^{-1}\sum_{i=1}^{n} \hat{g}(U_i)$, $\bar{\hat{f}} = n^{-1}\sum_{i=1}^{n} f(X_i, \hat{\beta})$. When $\hat{\sigma}_\varepsilon^2 > 0$, we can obtain the maximum of the likelihood function in the following points: $\hat{\sigma}_\varepsilon^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \{Y_i - f(X_i, \hat{\beta}) - \hat{g}(U_i) - (\bar{Y} - \bar{\hat{f}} - \bar{\hat{g}})\}^2$, $\hat{\sigma}_b^2 = \frac{1}{n}(\bar{Y} - \bar{\hat{f}} - \bar{\hat{g}})^2 - \frac{1}{n}\hat{\sigma}_\varepsilon^2$. When $\hat{\sigma}_b^2 = 0, \hat{\sigma}_\varepsilon^2 = \frac{1}{n^2} \sum_{i=1}^{n} \{Y_i - f(X_i, \hat{\beta}) - \hat{g}(U_i)\}^2$.

Theorem 2.1: *We suppose M is a finite and positive definite matrix. Let $f'(X, \beta) = \frac{\partial f(X,\beta)}{\partial \beta}$ and β_0 is the true value of β. Under the condition of $(C_1) \sim (C_5)$ in the Appendix and $\|\beta^0 - \beta_0\| = O_P(n^{-\frac{1}{2}})$, then we have*

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, VM^{-1}),$$

where $M = E[f'(X, \beta_0) - E\{f'(X, \beta_0)|U\}][f'(X, \beta_0) - E\{f'(X, \beta_0)|U\}]^T$.

Theorem 2.2: *We suppose $\|\hat{\beta} - \beta_0\| = O_P(n^{-\frac{1}{2}})$. Under the condition of $(C_1) \sim (C_5)$ in the Appendix and for any $u_0 \in \Re$, we have*

$$\sqrt{n}\{\hat{g}(u_0) - \hat{g}(u_0)\} \xrightarrow{D} N(0, \gamma^2(u_0)),$$

where $\gamma^2(u_0) = \sigma^2\{\varphi(u_0)\}^{-1}$, $\sigma^2 = \sigma_b^2 + \sigma_\varepsilon^2 > 0$ and $\varphi(\cdot)$ is the density function of U.

Theorem 2.3: *Under the condition of $(C_1) \sim (C_5)$ in the Appendix, if the distribution of X_i is compactly supported set, then $\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2 = O_P(n^{-\frac{1}{2}})$ and $\hat{\sigma}_b^2 - \sigma_b^2 = O_P(n^{-\frac{1}{2}})$.*

3. Numerical simulation studies

We will introduce the performance of the estimation methods for model (1) in this section. In our simulation, the root-mean-squared errors (RMSEs) are used to assess the precision of $\hat{\beta}$ and $\hat{g}(\cdot)$. We define the RMSE as RMSE$= \{n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} [\hat{g}(u_k) - g(u_k)]^2\}^{\frac{1}{2}}$ to evaluate $\hat{g}(\cdot)$, where $\{u_k, k = 1, \ldots, n_{\text{grid}}\}$ are equidistant grid points, and n_{grid} is the number of grid points.

Table 1. The result of simulation for Example 3.1. The mean (Bias), standard deviation (SD) and root-mean-squared error (RMSE) for the estimates of β.

Parameter	n	Mean (Bias)	SD	RMSE
	50	1.50054 (0.00054)	0.00648	0.00650
β	100	1.50051 (0.00051)	0.00485	0.00487
	150	1.50002 (0.00002)	0.00403	0.00402

3.1. Simulation

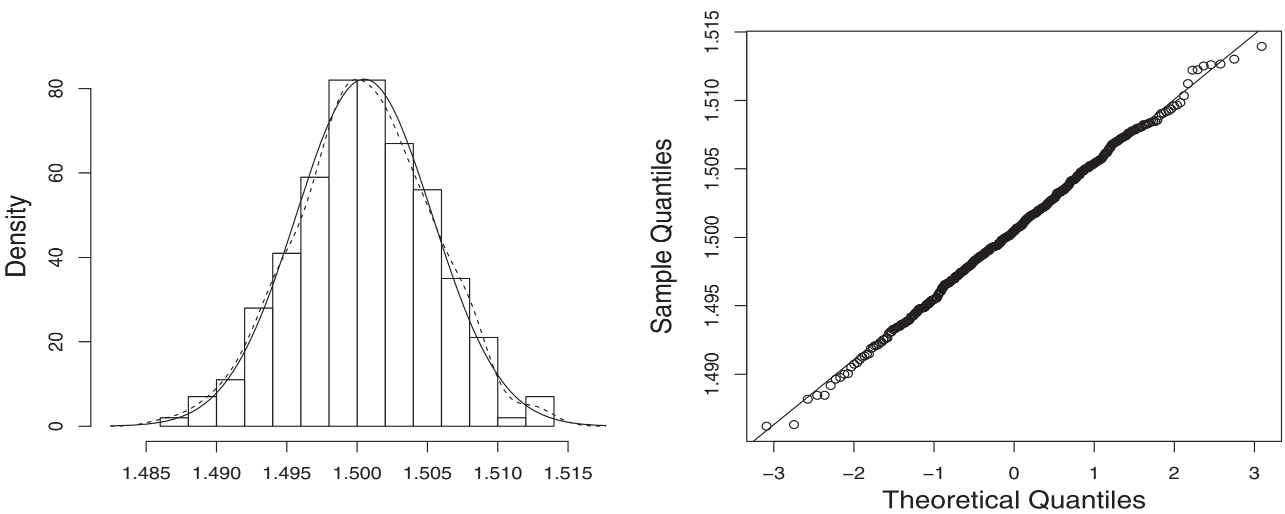
Example 3.1: We consider model (1) as

$$Y_i = 12(U_i - 0.5)^3 + X_i\beta^2 + b_i + \varepsilon_i,\ i = 1, \ldots, n$$

where $\beta = 1.5$, $U_i \sim U(0, 1)$, $X_i \sim N(0, 1)$, $b_i \sim N(0, 0.36)$ and $\varepsilon_i \sim N(0, 0.04)$. Here $g(u) = 12(u - 0.5)^3$ and $f(x, \beta) = x\beta^2$. The number of subjects $n = 50$, 100, 150. In our simulation, the initial value β^0 can be obtained by fitting the linear model. We compute the mean (Bias), standard deviation (SD) and RMSE by repeating the simulation 500 times. From Table 1, we can obtain the simulation results.

Table 1 tells us the small sample size will lead to larger SD and RMSE. It also tells us the Bias, SD and RMSE decrease as n increases, thus the improvement is significant. When $n = 100$, the histogram of the 500 estimates of β is shown in Figure 1(a). The Q–Q plot of the 500 estimates of β is shown in Figure 1(b).

As for the estimation of $g(u)$, we also considered. When $n = 100$, the real link function curve and the estimated link function curve are shown in Figure 2(a); it shows the estimated curve is close to the real link function curve, so it tells us that the estimation methods for data fitting is ideal. When $n = 100$, the boxplots of the 500 RMSEs for link function are shown in



(a) Histogram (b) Normal Q-Q plot.

Figure 1. When $n = 100$, the results of simulation for Example 3.1: (a) shows the histogram of the 500 estimates of β. The dash curve shows the estimated curve of density and the solid curve shows the curve of normal density. (b) shows the Q–Q plot of the 500 estimates of β.

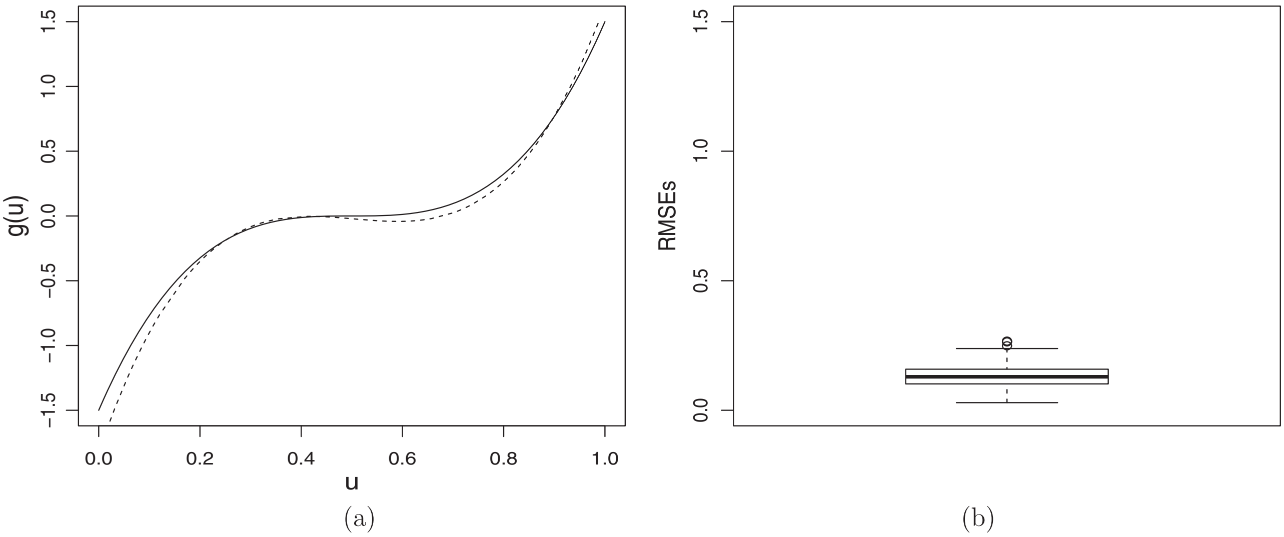


Figure 2. When $n = 100$, the result of simulation for Example 3.1. (a) solid curve shows the real curve of $g(\cdot)$ and dashed curve shows the estimated curve of $g(\cdot)$; (b) shows the boxplots of the 500 RMSEs of the estimates $\hat{g}(\cdot)$.

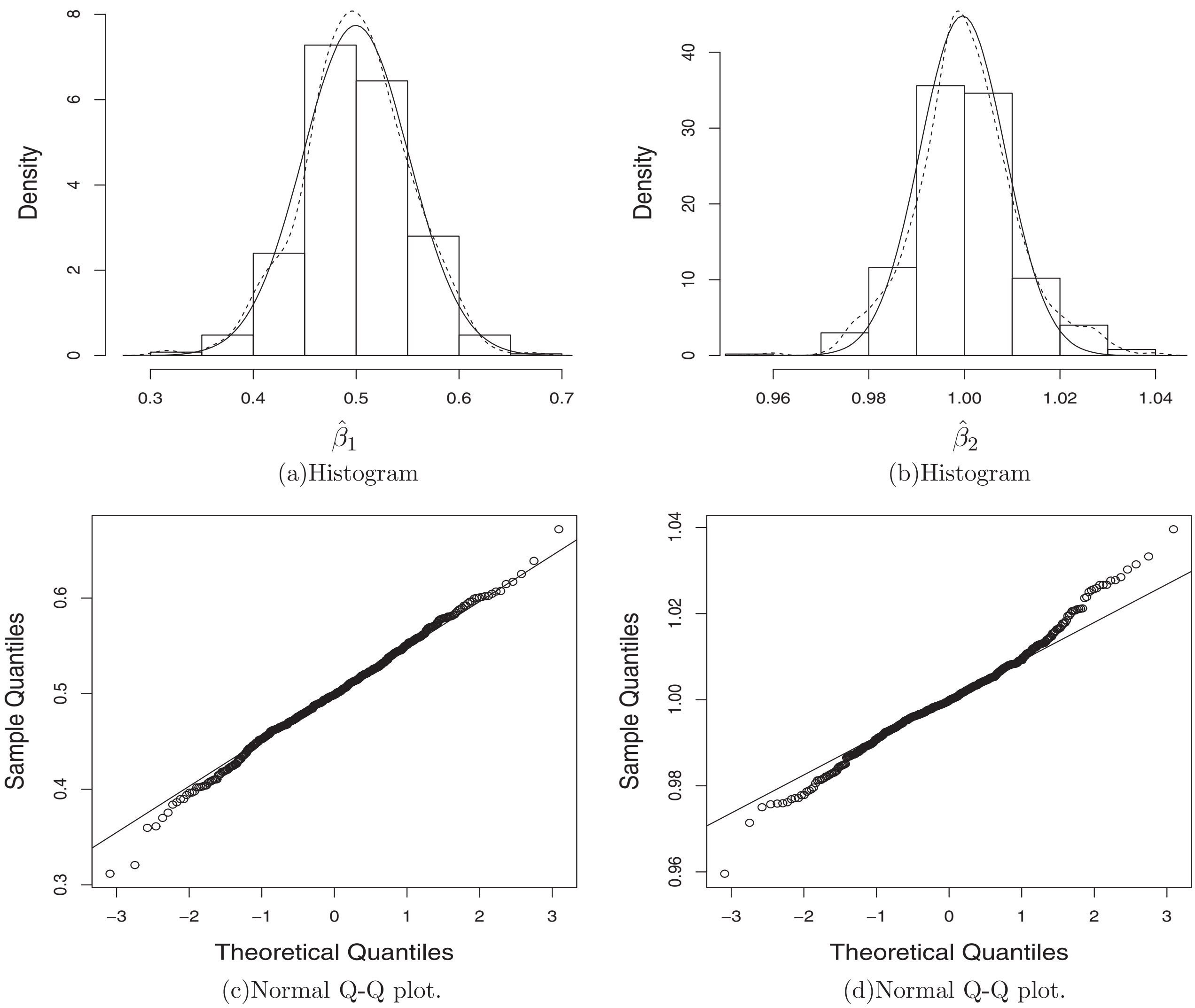


Figure 3. When $n = 100$, the result of simulation for Example 3.2: (a) and (b) shows the histogram of the 500 estimates of β_1 and β_2, respectively. The dash curve shows the estimated curve of density and the solid curve shows the curve of normal density. The Q–Q plot of the 500 estimates of β_1 and β_2 are shown in (c) and (d), respectively.

Table 2. The result of simulation for Example 3.2. The mean (Biases), standard deviations (SD) and root-mean-squared error (RMSE) for the estimates of β_1 and β_2.

β	n	Mean (Bias)	SD	RMSE
β_1	50	0.50148 (0.00148)	0.08003	0.07997
	100	0.50119 (0.00119)	0.05334	0.05329
	150	0.50117 (0.00117)	0.04218	0.04215
β_2	50	1.00096 (0.00096)	0.01770	0.01771
	100	1.00022 (0.00022)	0.01111	0.01110
	150	1.00007 (0.00007)	0.00831	0.00830

Figure 2(b), it shows the RMSEs of the estimates for link function are small.

Example 3.2: We consider model (1) as

$$Y_i = \cos(2\pi U_i) + X_i\beta_1 + \exp(X_i\beta_2) + b_i + \varepsilon_i,\ i = 1, \ldots, n$$

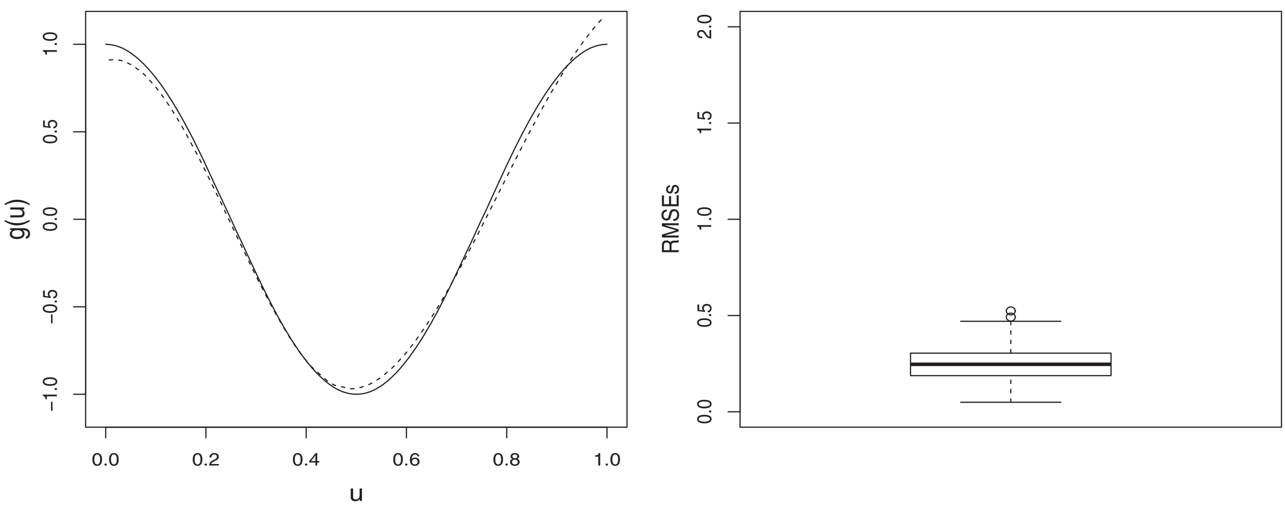
where $\beta_1 = 0.5$, $\beta_2 = 1$, $U_i \sim U(0, 1)$, $X_i \sim N(0, 1)$, $b_i \sim N(0, 1)$ and $\varepsilon_i \sim N(0, 0.36)$. Here $g(u) = \cos(2\pi u)$ and $f(x; \beta_1, \beta_2) = x\beta_1 + \exp(x\beta_2)$. The number of subjects $n = 50, 100, 150$. In our simulation, the initial value β^0 can be obtained by fitting the linear model. We compute the mean (Biases), standard deviation (SD) and RMSE by repeating the simulation 500 times. From Table 2, we can obtain the simulation results.

Table 2 tells us the small sample size will lead to larger SD and RMSE. It also tells us the Bias, SD and RMSE decrease as n increases.When n=100, the histograms of the 500 estimates of β_1 and β_2 are shown in Figure 3(a,b), respectively. The Q–Q plots of the 500 estimates of β_1 and β_2 are shown in Figure 3(c,d), respectively.

As for the estimation of $g(u)$, we also considered. When $n = 100$, the real link function curve and the estimated link function curve are shown in Figure 4(a); it shows the estimated curve is close to the real link function curve, so it tells us that the estimation methods for data fitting is ideal. When $n = 100$, the boxplot of the 500 RMSEs for link function is shown in Figure 4(b), it shows the RMSEs of the estimates for link function are small.

3.2. A real data example

Now, we use a Boston data-set to illustrate the proposed methodology. The number of observations in this data-set is 506. The interesting variables include the median value of owner-occupied home (*MEDV*), lower status of the population (*LSTAT*) and proportion of owner-occupied units built prior to 1940 (*AGE*). Here we use model (1) to fit the data-set. Among them, *MEDV* is the response variable Y, $\sqrt{LSTAT}$ is the covariate U, *AGE* is the covariate X, where the time is random selection. We will use kernel regression estimate to determine the function $f(\cdot, \cdot)$. From Figure 5(a), it tells us that the function $f(\cdot, \cdot)$ may be a quadratic function, so we assume that the form of function $f(\cdot, \cdot)$ is $f(X, \beta_1, \beta_2, \beta_3) = \beta_1 X^2 + \beta_2 X + \beta_3$. Then we use the proposed method to estimate $\beta_1, \beta_2, \beta_3$ and σ_b^2, where the initial value $\beta^0 = (\beta_1^0, \beta_2^0, \beta_3^0)$ can be obtained by fitting the linear model. By using the proposed methodology, we get the estimators of the parameter β and variance component σ_b^2 are $(-0.00052, 0.02220, 1.94410)^T$ and 0.39412, respectively. We also consider the estimated curve of $g(\cdot)$, which is shown in Figure 5(b).



(a) (b)

Figure 4. When $n = 100$, the result of simulation for Example 3.2: (a) solid curve shows the real curve of $g(\cdot)$ and dashed curve shows the estimated curve of $g(\cdot)$; (b) shows the boxplots of the 500 RMSEs of the estimates $\hat{g}(\cdot)$.

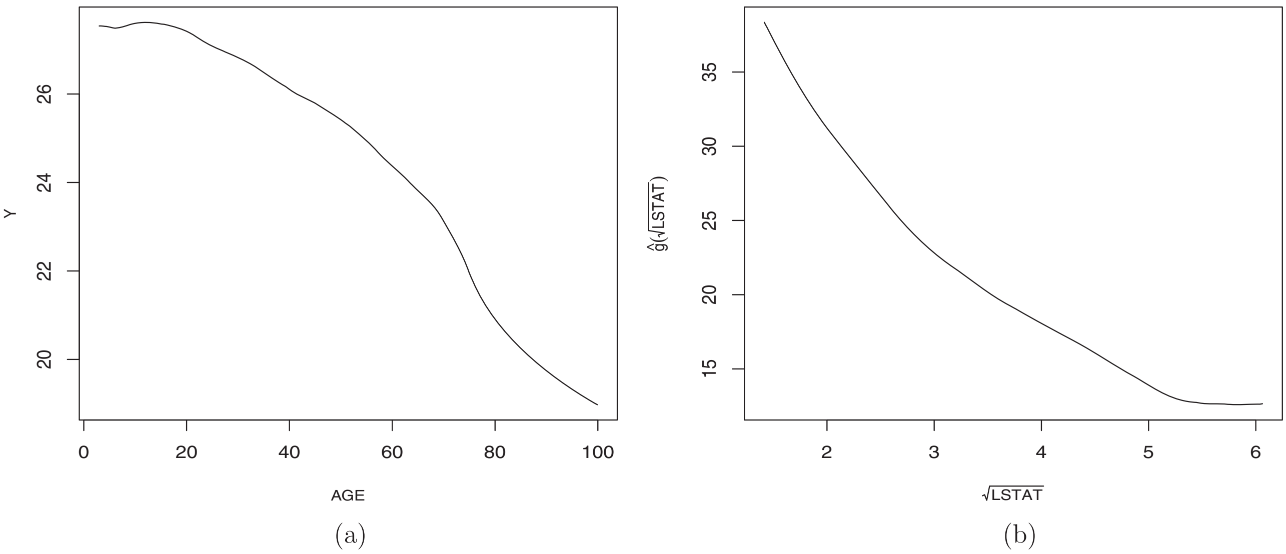


Figure 5. Applied to the Boston data, (a) shows the relationship between AGE and MEDV and (b) shows the estimated curve of $g(\cdot)$.

Acknowledgements

We thank the editor, the associate editor and two referees for their helpful and constructive comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Natural Science Foundation of China [grant number 11471160], [grant number 11101114], [grant number 11571112]; the National Statistical Science Research Key Program of China [grant number 2013LZ45]; the Fundamental Research Funds for the Central Universities [grant number 30920130111015]; the Jiangsu Provincial Basic Research Program (Natural Science Foundation) [grant number BK20131345] and sponsored by Qing Lan Project.

Notes on contributors

Ye Que is a master of statistics of School of Science in Nanjing University of Science and Technology.

Zhensheng Huang is a professor of statistics of School of Science in Nanjing University of Science and Technology.

Riquan Zhang is a professor and chair of School of Statistics in East China Normal University.

References

Ai, C., & Mcfadden, D. (1997). Estimation of some partially specified nonlinear models. *Journal of Econometrics, 76*, 1–37.

Bates, D. M., & Watts, D. G. (1988). Nonlinear regression analysis and its applications. *Journal of Marketing Research, 60*, 194–196.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics, 21*, 196–216.

Gu, C., & Ma, P. (2005). Optimal smoothing in nonparametric mixed-effect models. *Annals of Statistics, 33*, 1357–1379.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems (with discussion). *Journal of the American Statistical Association, 72*, 320–338.

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management, 5*, 81–102.

Jiang, J. M. (2007). *Linear and generalized linear mixed models and their applications*. New York: Springer.

Ke, C. L., & Wang, Y. D. (2001). Semiparametric nonlinear mixed-effects models and their applications(with discussion). *Journal of the American Statistical Association, 96*, 1294–1298.

Lindstrom, M. J., & Bates, D. M. (2010). Nonlinear mixed-effects models for repeated measures data. *Biometrics, 46*, 673–687.

Li, T. Z., & Mei, C. G. (2013). Estimation and inference for varying coefficient partially nonlinear models. *Journal of Statistical Planning and Inference, 143*, 2023–2037.

Li, R., & Nie, L. (2008). Efficient statistical inference procedures for partially nonlinear models and their applications. *Biometrics, 64*, 904–911.

Müller, S., & Vial, C. (2009). Partially linear model selection by the bootstrap. *Australian and New Zealand Journal of Statistics, 51*, 183–200.

Masry, E., & Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory, 11*, 258–289.

Pang, Z., & Xue, L. G. (2012). Estimation for the single-index models with random effects. *Computational Statistics and Data Analysis, 56*, 1837–1853.

Song, L. X., Zhao, Y., & Wang, X. G. (2010). Sieve least squares estimation for partially nonlinear models. *Statistics and Probability Letters, 80*, 1271–1283.

Xiao, Y. T., Tian, Z., & Li, F. X. (2013). Empirical likelihood-based inference for parameter and nonparametric function in partially nonlinear models. *Journal of the Korean Statistical Society, 43*, 367–379.

Zhong, X. P., Fung, W. K., & Wei, B. C. (2002). Estimation in linear models with random effects and errors-in-variables. *Annals of the Institute of Statistical Mathematics*, *54*, 595–606.

Appendix

The following regular conditions will be used in the proofs of theorems.

- (C_1) $(\tau_1, \dots, \tau_J$ are the internal nodes of spline function, namely, we assume $z_0 = 0$, $z_{k+1} = 1$, $h_i = z_i - z_{i-1}$, $h = \max_{1 \le i \le k+1} h_i$ and exist a constant M_0, making $\frac{h}{\min_i} \le M_0$ and $\max_i |h_{i+1} - h_i| = o(\frac{1}{k})$, so $h = o(\frac{1}{k})$.
- (C_2) The design point sequence $\{U_i, i = 1, \dots, n\}$ has bounded support set $\Re$, and the density function $\varphi(u)$ satisfies Lipschitz continuous and bounded from 0.
- (C_3) There is a constant c_0, making $E(\epsilon^2) \le c_0 < \infty$, where $\epsilon = Z^T b + \varepsilon$.
- (C_4) For $\forall X$, $f(X, \beta)$ is a continuous function about β and the second derivatives of $f(X, \beta)$ about β are continuous. Let Ω is compactly supported set and $\beta \in \Omega$.
- (C_5) Let $f(X_i, \beta) = [\partial f(X_i, \beta)/\partial \beta]_{P \times 1}$, then $E\{\|f(X, \beta)\|^4\} < \infty$.

Proof of Theorem 2.1. Let $\eta(\beta) = \{f(X_1, \beta), \dots, f(X_n, \beta)\}^T$ and $\eta'(\beta) = \{f(X_1, \beta), \dots, f(X_n, \beta)\}^T$. From (3), we get $f(X, \beta) = f(X, \beta^0) + f(X, \beta^0)^T(\beta - \beta^0)$, so $Y_i \approx g(U_i) + f(X_i, \beta^0) + f'(X_i, \beta^0)^T(\beta - \beta^0) + Z_i^T b_i + \varepsilon_i$. Let

$$z_i^* = Y_i - f(X_i, \beta^0) + f'(X_i, \beta^0)^T \beta^0, \quad \text{(A.1)}$$

we have $z_i^* = g(U_i) + f'(X_i, \beta^0)^T \beta + Z_i^T b_i + \varepsilon_i$. Now, we note that

$$\hat{\beta} - \beta_0 = \{\eta'(\beta^0)^T \eta'(\beta^0)\}^{-1} \eta'(\beta^0)^T \{z^* - \eta'(\beta^0)\beta_0\},$$

where $z^* = (z_1^*, \dots, z_n^*)^T$. Under the condition of (C_1) $\sim$ (C_5) and using (A.5) in Li and Nie (2008), we can get $\frac{1}{n}\eta'(\beta^0)^T \eta'(\beta^0) = M\{1 + o_p(1)\}$. Now, we have

$$\frac{1}{\sqrt{n}} \eta'(\beta^0)^T [z^* - \eta'(\beta^0)\beta_0] = \sqrt{n}\xi_n + o_P(1), \quad \text{(A.2)}$$

where $\xi_n = n^{-1} \sum_{i=1}^n [f'(X_i, \beta_0) - E\{f'(X, \beta_0)|U = U_i\}]\varepsilon_i$. From (A.1), we get $z^* - \eta'(\beta^0)\beta_0 = Y - \eta(\beta^0) + \eta'(\beta^0)(\beta^0 - \beta_0)$. Similar to (A.2) in Li and Nie (2008), we get $\frac{1}{\sqrt{n}}\eta'(\beta_0)^T\{Y - \eta(\beta_0)\} = \sqrt{n}\xi_n + o_P(1)$. To get (A.1), we show that

$$\frac{1}{\sqrt{n}} \eta'(\beta_0)^T \{\eta(\beta_0) - \eta(\beta^0) + \eta'(\beta^0)(\beta^0 - \beta_0)\} = o_p(1), \quad \text{(A.3)}$$

and

$$\frac{1}{\sqrt{n}} \{\eta'(\beta^0) - \eta'(\beta_0)\}^T \{z^* - \eta'(\beta^0)\beta_0\} = o_p(1). \quad \text{(A.4)}$$

Through calculation and $\|\beta^0 - \beta_0\| = O_P(n^{-\frac{1}{2}})$, we can get the left of (A.3) is of the order $O_P(\sqrt{n}\|\beta^0 - \beta_0\|^2) = O_P(\frac{1}{\sqrt{n}})$. We also can get the left of (A.4) is of the order $O_P(c_n\|\beta^0 - \beta_0\| = O_P(c_n/\sqrt{n})$, where $c_n = \frac{1}{\sqrt{n}}$. So, we prove (A.2). Now, we have $\sqrt{n}(\hat{\beta} - \beta_0) = M\{1 + o_P(1)\}^{-1}\{\sqrt{n}\xi_n + o_P(1)\}$. Using the central limit Theorem and the Slutsky Theorem, we get $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, VM^{-1})$, then the proof of Theorem 2.1 is completed. □

Proof of Theorem 2.2: Using (A.6) in Pang and Xue (2012) and conditions (C_1) $\sim$ (C_5), we can get $\hat{g}(u_0, \hat{\beta}) - \hat{g}(u_0) = [\varphi(u_0)]^{-1}\xi_{n,0}(u_0, \hat{\beta}) + o_P(n^{-\frac{1}{2}})$. Through $\|\hat{\beta} - \beta_0\| = O_P(n^{-\frac{1}{2}})$, we can prove that $\xi_{n,0}^*(u_0, \hat{\beta}) = \frac{1}{n}\sum_{i=1}^n [f(X_i, \beta_0) + Z_i^T b_i + \varepsilon_i](U_i - u_0) + O_P(n^{-\frac{1}{2}})$. Through the Theorem 4.4 in Masry and Tjøstheim (1995), we can get $\sqrt{n}\xi_{n,0}^*(u_0, \hat{\beta}) \xrightarrow{D} N(0, \gamma^2(u_0))$. Thus, we complete the proof of the Theorem 2.2. □

Proof of Theorem 2.3: The proof of Theorem 2.3 is similar to Theorem 4 in Pang and Xue (2012), here we omit the proof process. □