Back-fitting Estimation of Semi-parametric Partially Linear Varying-coefficient Models with PCA

Ming Xing Zhang[1, a *], Zi Xin Liu[1, b] and Jian Nan Qiao[1, c]

[1]Shcool of .Guizhou university of finance and economics, China

zmxgraduate@163.com, [b]xinxin905@163.com, [c]qjn560@163.com

Keywords: Partially linear varying-coefficient model；Principal component analysis；Back-fitting procedure； Multi-collinearity； Semi-parametric estimation.

Abstract. This paper investigates the estimation problem of semi-parametric partially linear varying-coefficient models by the technique of back-fitting. In order to avoid the disturbance of multi-collinearity and improve estimation efficiency, we apply principal component analysis to semi-parametric partially linear varying-coefficient models due to principal components are those uncorrelated linear combinations. And then we obtain the estimators of original parametric component and nonparametric component respectively. Model estimation and some statistic inferences about the property of the estimators are also derived theoretically.

Introduction

In recent three decades, with the rapid development of computing techniques, the semi-parametric models have gained more and more attention in various areas. As we know, semi-parametric models have many forms such as partially linear models, varying-coefficient models, additive models and so on. In this paper, we consider the partially linear varying-coefficient model which is a useful extension as follows:

$$Y = X^T\beta + Z^T\alpha(U) + \varepsilon, \tag{1}$$

where Y is response, and (X, Z, U) are the associated covariates. For simplicity, we assume U is univariate. ε is an independent random error with $E(\varepsilon \mid X, Z, U) = 0$ and $Var(\varepsilon \mid X, Z, U) = \delta^2$. $\beta = (\beta_1, \ldots \beta_q)^T$ is a q-dimensional vector of unknown parametric component, $\alpha(U) = (\alpha_1(U), \ldots \alpha_p(U))$ is a p-dimensional vector of unknown coefficient functions.

Obviously, when $X = 0$ Eq.1 reduces to varying-coefficient model, which has been widely studied in the literature, see the work of Hu and Xia[1], Cai et al[2], Fan and Zhang[3], and among others. When $p = 1$ and $Z = 1$ Eq.1 becomes partially linear regression model, which was proposed by Engle et al[4] when they researched the influence of weather on electricity demand. A series of literature (Chen[5], Hu et al[6]) regarding partially linear regression model have provided corresponding statistical inference.

Recently, Eq.1 has been widely studied by Fan and Huang[7], Wei and Wu[8], You and Chen[9] and so on. In [9], You and Chen studied the estimation of partially linear varying-coefficient model under the circumstance that some covariates were measured with additive errors.

However, in practice, there may exists multi-collinearity relation among explanatory variable X. In that case, we introduce principal components into the semi-parametric model to eliminate the influence of multi-collinearity. With the purpose of not losing information, we choose principal components as the same dimension as explanatory variables. And then we get the estimators for the parametric component and nonparametric component based on the approach of back-fitting.

Principal Component Analysis

Suppose that $\{Y_i, X_i, Z_i, U_i\}_{i=1}^n$ is a random sample from Eq.1, that is to say, they satisfy

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_q X_{iq} + \alpha_1(U_i) Z_{i1} + \cdots + \alpha_P(U_i) Z_{ip} + \varepsilon_i \ . \tag{2}$$

Let X be a $n \times q$ matrix, where n and q are the number of observations and the number of the variables from the linear part. The covariance matrix of X is Σ, which is a $q \times q$ matrix. $\{\lambda_1, \ldots, \lambda_q\}$ are the corresponding eigenvalues of the covariance matrix Σ, where $\lambda_1 \geq \lambda_2 \ldots \lambda_q \geq 0$. Additionaly, $l_j^T = (l_{j1} \ldots, l_{jq}), j = 1, \ldots, q.$ stands for the eigenvectors. Thus, we have

$$\begin{cases} V_{i1} = l_1^T X_i = l_{11} X_{i1} + l_{12} X_{i2} + \ldots + l_{1q} X_{iq} \\ \quad \ldots \\ V_{iq} = l_q^T X_i = l_{q1} X_{i1} + l_{q2} X_{i2} + \ldots + l_{qq} X_{iq} \end{cases}, \tag{3}$$

where

$$L^T = (l_1, l_2 \ldots l_q)^T = \begin{pmatrix} l_{11} \ldots l_{1q} \\ \vdots \\ l_{q1} \ldots l_{qq} \end{pmatrix}, \qquad V = \begin{pmatrix} V_{11} \ldots V_{1q} \\ \vdots \\ V_{n1} \ldots V_{nq} \end{pmatrix}.$$

By Eq.3, we structure the $V_{ij}, i = 1, \ldots, n, j = 1, \ldots, q$ as principal components which are the linear combinations of original variables. Where $D(V_i) = L_i^T \Sigma L_i$ with $\Lambda = diag(\lambda_1, \ldots, \lambda_q)$.Without loss of information, we note that the semi-parametric partially linear varying-coefficient model with principal components can be written as

$$Y_i = \gamma_1 V_{i1} + \gamma_2 V_{i2} + \ldots \gamma_q V_{iq} + \alpha_1(U_i) Z_{i1} + \ldots \alpha_p(U_i) Z_{ip} + \varepsilon_i, \tag{4}$$

where $\gamma = (\gamma_1, \ldots \gamma_q)$ is the parametric component of model (4). In that case, based on Eq.2 and Eq.4, we have $\beta = L\gamma$.

Back-fitting Estimation

As for Eq.4, we use back-fitting procedure to obtain the estimators for parametric component and nonparametric component respectively. Firstly, suppose that nonparametric component $\alpha(U_i)$ is known, Eq.4 can be written as

$$Y_i - \sum_{k=1}^{p} \alpha_k(U_i) Z_{ik} = \sum_{k=1}^{q} \gamma_k V_{ik} + \varepsilon_i. \tag{5}$$

By Eq.5 we get the least squares estimator of γ , say $\hat{\gamma}$, satisfies $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_q)^T = (V^T V)^{-1} V^T (Y - M)$,where $M = \left(X_1{}^\tau \alpha(U_1), \ldots, X_n{}^\tau \alpha(U_\mathrm{n}) \right)^\tau$, $Y = (Y_1 \ldots Y_n)$.

Further then if γ as the parametric component is known, the Eq.4 becomes

$$Y_i - \sum_{k=1}^{q} \gamma_k V_{ik} = \sum_{k=1}^{p} \alpha_k(U_i) Z_{ik} + \varepsilon_i \ . \tag{6}$$

According to Fan and Huang[7], we apply the local linear regression to the coefficient functions. For U in a small neighbourhood of u_0, one can approximate $a_j(U)$ locally by a linear function

$$\alpha_j(U) \approx \alpha_j(u_o) + \alpha'_j(U)(U - u_0) \equiv a_j + b_j(U - u_0).$$

This leads to the following weighted least-squares problem: find $\{(a_j, b_j), j = 1,...p\}$ to minimize

$$\sum_{i=1}^{n}[Y_i^* - \sum_{j=1}^{p}\{a_j + b_j(U_i - u_0)X_{ij}]^2 K_h(U_i - u_0), \tag{7}$$

where $Y_i^* = Y_i - V_i^T\gamma, K_h(.) = K(./h)/h$. We denote $K(.)$ is a kernel function and h is a bandwidth.

Therefore the estimator of $\alpha(U)$ is given by

$$\hat{\alpha}(U) = (I_p 0_p)\{D_{u_0}^T W_{u_0} D_{u_0}\}^{-1} D_{u_0}^T W_{u_0}(Y - V\gamma),$$

where $W_{u_0} = diag(K_h(U_1 - u_0),...K_h(U_n - u_0))$. According to the above results, we can obtain the estimator of M, it is shown that

$$\hat{M} = S(Y - V\gamma), \tag{8}$$

where

$$D_{u_0} = \begin{pmatrix} Z_1^T & \frac{U_1 - u_0}{h} Z_1^T \\ & \vdots \\ Z_n^T & \frac{U_n - u_0}{h} Z_n^T \end{pmatrix}, \qquad S = \begin{pmatrix} (V_1^T, 0)\{D_{u_1}^T W_{u_1} D_{u_1}\}^{-1} D_{u_1}^T W_{u_1} \\ \vdots \\ (V_n^T, 0)\{D_{u_n}^T W_{u_n} D_{u_n}\}^{-1} D_{u_n}^T W_{u_n} \end{pmatrix}.$$

Based on backing-fitting theory, we have $M = S(Y - V\gamma)$, $V\gamma = V(V^TV)^{-1}V^T(Y - M)$.

Thus, under the assumption that the matrix of $V^T(I - S)V$ is invertible, the estimators of γ and M can be defined as

$$\begin{cases} \hat{\gamma} = (V^T(I - S)V)^{-1}V^T(I - S)Y \\ \hat{M} = S(Y - V\hat{\gamma}) \end{cases}. \tag{9}$$

Meanwhile, the fitting values of response are further obtained, namely, $\hat{Y} = (Y_1,...,Y_n)^T = V\hat{\gamma} + \hat{M} = EY$, where $E = S + (I - S)V[V^T(I - S_2)V]^{-1}V^T(I - S)$.

Note that $\beta = L\gamma$, the estimator of original parametric component β satisfies

$$\hat{\beta}_{BF} = L[V^\tau(I - S)V]^{-1}V^T(I - M)Y.$$

The Property of Estimators

Let $u_i = \int u^i K(u)du$, $v_i = \int u^i K^2(u)du$, and we denote $\Gamma(U) = E(ZZ^T|U)$, $\Phi(U) = E(ZV^T|U)$, in that case some statistical inferences are given as follows.

Theorem 1 By the estimator of $\hat{\beta}_{BF}$, we have

$$E(\hat{\beta}_{BF} - \beta | Z,V,U) = L[V^T(I - S)V]^{-1}V^T(I - S)M,$$

$$Var(\hat{\beta}_{BF|Z,V,U}) = \sigma^2 L[V^T(I - S)V]^{-1}V^T(I - S)(I - S)^T V[V^T(I - S)V]^{-1}L^T.$$

Suppose that the assumption conditions hold, the above results can be simplified as

$$E(\hat{\beta}_{BF} - \beta | Z,V,U) = -\frac{h^2}{2}u_2(K)LE(V_i Z_i^T \alpha''(U_i))\Xi^{-1} + O_P(h^2),$$

$$Var(\hat{\beta}_{BF} \mid Z,V,U) = L\frac{\sigma^2}{n}\Xi^{-1}L^T + O_p(\frac{1}{n}),\ \ \Xi = E(VV^T) - E[E(VZ^T \mid U)E(ZZ^T \mid U)E(ZV^T \mid U)].$$

Corollary 1[10] If $h\, n^{r/5}$,where $-1 < r < -1/4,$ we have $\sqrt{n}(\hat{\beta}_{BF} - \beta) \longrightarrow N(0, \sigma^2\Sigma^{-1}).$

Proofs of the Conclusion

We begin with the following assumptions needed to prove the theorems for the proposed methods.

Assumption 1. The random variable U has a bounded support Ω. Its density function $f(.)$ is Lipschitz continuous and bounded away from 0 on its support.

Assumption 2. For each $U \in \Omega$, $E(ZZ^T \mid U)$ is non-singular. $E(ZZ^T \mid U)$, $E(ZZ^T \mid U)^{-1}$ and $E(ZV^T \mid U)$ are all Lipschitz continuous.

Assumption 3. There is an $s > 2$ such that $E \| X \|^{2s} < \infty$ and $E \| Z \|^{2s} < \infty$ and for some $\varepsilon < 2 - s^{-1}$ such that $n^{2\varepsilon - 1}h \to \infty$.

Assumption 4. $\{\alpha_j(.), j = 1,...p\}$ have continuous second derivatives in $U \in \Omega$.

Assumption 5. The function K(.) is a symmetric density function with compact support and the bandwidth satisfies $nh^8 \to 0$ and $nh^2 / (logn)^2 \to \infty.$

Lemma A.1 Let $(X_1, Y_1)...(X_n, Y_n)$, be independent and identically distributed random vectors, where the Y_i are scalar random variables. Further assume that $E \mid y \mid^s < \infty$ and $\sup_x \int \mid y \mid^s f(x,y)dy < \infty,$ where f means the joint density of (X,Y). Let K be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Given that $n^{2\varepsilon-1}h \to \infty$ for some $\varepsilon < 1 - s^{-1}$, then

$$\sup_x \mid \frac{1}{n}\sum_{i=1}^{n}[K_h(X_i - x)Y_i - E\{K_h(X_i - x)Y_i\}] \mid = O_p(\{\frac{log(1/h)}{nh}\}^{1/2}).$$

The proof of Lemma A.1 can be found in Mack and Silverman [11].

Proof of Theorem 1 By the definition of $\hat{\beta}_{BF}$

$$\hat{\beta}_{BF} = L[V^T(I-S)V]^{-1}V^T(I-S) = \beta + L[V^T(I-S)V]^{-1}V^T(I-S)(M+\varepsilon),$$

we have $E(\hat{\beta}_{BF} - \beta \mid Z,V,U) = L[V^T(I-S)V]^{-1}V^T(I-S)M.$

In order to get the asymptotic property of the deviation, we discuss $V^T(I-S)V$, and $V^T(I-S)M$ respectively. Similar to the proof of LemmaA.2 in Fan and Huang[7], observes that

$$D_{u_0}^T W_{u_0} D_{u_0} = nf(U)\Gamma(U) \otimes \begin{pmatrix} 1 & 0 \\ 0 & u_2 \end{pmatrix}\{1 + O_P(c_n)\},\ D_{u_0}^T W_{u_0} V = nf(U)\Phi(U) \otimes (1,0)^T\{1 + O_P(c_n)\}.$$

And then $SV = [Z^T, 0]\{D_{u_0}^T W_{u_0} D_{u_0}\}^{-1} D_{u_0}^T W_{u_0} V = Z^T\Gamma^{-1}(U)\Phi(U).$

Now, with the above results it is easy to show that

$$\frac{1}{n}V^T(I-S)V = \frac{1}{n}\sum_{i=1}^{n}V_iV_i^T - \frac{1}{n}\sum_{i=1}^{n}V_iZ_i^T\Gamma^{-1}(U_i)\Phi(U_i)\{1 + O_P(c_n)\} \to \Xi.$$

By Taylor expansion, we have $\alpha(U_i) = \alpha(u_O) + ha'(u_0)(\frac{U_i - u_0}{h}) + \frac{h^2}{2}\alpha''(u_0)(\frac{U_i - u_0}{h})^2 + 0_p(h^2).$

Then it follows that

$$M=\begin{pmatrix} Z_1^T\{\alpha(u_0)+h\alpha'(u_0)(\frac{U_1-u_0}{h})+\frac{h^2}{2}\alpha''(u_0)(\frac{U_1-u_0}{h})^2\} \\ \vdots \\ Z_1^T\{\alpha(u_0)+h\alpha'(u_0)(\frac{U_1-u_0}{h})+\frac{h^2}{2}\alpha''(u_0)(\frac{U_1-u_0}{h})^2\} \end{pmatrix}+O_p(h^2)=D_{U_0}\begin{pmatrix}\alpha(u_0)\\ h\alpha'(u_O)\end{pmatrix}+\frac{h^2}{2}\begin{pmatrix} Z_1^T(\frac{U_1-u_0}{h})^2 \\ \vdots \\ Z_n^T(\frac{U_n-u_0}{h})^2\end{pmatrix}\alpha''(u_0)+0_P(h^2).$$

Note that

$$D_{u_0}^T W_{u_0}\begin{pmatrix} Z_1^T(\frac{U_1-u_0}{h})^2 \\ \vdots \\ Z_n^T(\frac{U_n-u_0}{h})^2\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^n Z_iZ_i^T(\frac{U_i-u_0}{h})^2K_h(U_i-u_0) \\ \sum_{i=1}^n Z_iZ_i^T(\frac{U_i-u_0}{h})^3K_h(U_i-u_0)\end{pmatrix}=nu_2(K)f(U)\Gamma(U)\otimes(1,0)^T\{1+O_p(c_n)\}.$$

Furthermore, $SM=Z^T\alpha(U)+\frac{h^2}{2}u_2(K)Z_n^T\alpha''(U)\{1+O_p(c_n)\}$.

Hence, $\frac{1}{n}V^T(I-S)M=-\frac{h^2}{2}u_2(K)E(V_iZ_i^T\alpha''(U_i))+O_p(h^2)$.

By the definition of $\hat{\beta}_{BF}$, $Var(\beta_{BF})=\sigma^2L[V^T(I-S)V]^{-1}V^T(I-S)(I-S)^TV[V^T(I-S)V]^{-1}L^T$.

Similar to the above proof of $V^T(I-S)V$,and by Lemma A.1 we have $\frac{1}{n}V^T(I-S)(I-S)^TV\to\Xi$.

And then $Var(\hat{\beta}_{BF}|V,Z,U)=L\frac{\sigma^2}{n}\Xi^{-1}L^T+O_p(\frac{1}{n})$.

Together with the result, we prove the Theorem 1.

Acknowledgements

The authors thank the financial support of the National Natural Science Foundation of China, grant 61472093.

References

[1] T. Hu and Y. C. Xia .Adaptive Semi-varying Coefficient Model Selection, Statistica Sinica. Vol.22 (2010), p.575-599.

[2] Z. W. Cai, J. Q. Fan and Q.W.Yao. Efficient Estimation and Inference for Varying Coefficient Model, Journal of the American Statistical Association, Vol.95 (2000), No.451, p.888-902.

[3] J. Q. Fan and W. Y. Zhang. Statistical Estimation in Varying-coefficient Models, The Annals of Statistics, Vol.27 (1999), No.5, p.1491-1581.

[4] R. F. Engle, W. J. Granger, J. Rice and A .Weiss. Semi-parametric Estimates of the Relation Between Weather and Electricity Techniques. Journal of American Statistical Association, Vol.80 (1986), p.310-319.

[5] H. Chen. Convergence Rates for Parametric Components in a Partially Linear Model. The American Statistical Association, Vol.88 (1988), p.136-146.

[6] Z. H. Hu, N Wang and R. J .Carroll. Profile-kernel Versus Back-fitting in the Partially Linear Models for Longitudinal/clusterde data. Biometrika. Vol.91 (2004), p.251-262.

[7] J. Q. Fan and T. Huang. Profile Likelihood Inferences on Semi-parametric Varying-coefficient Partially Linear Models. Scandinavian Journal of Statistics. Vol.11.(2005), No.6, p.1031-1057.

[8] C H Wei. Statistical Inference in Partially Linear Varying-coefficient Errors-in-variables Models. Journal of Statistical Planning and Inference. Vol.142(2012), p.2464-2472.

[9] J .H.You and G.M. Chen. Estimation of a Semi-parametric Varying-coefficient Partially Linear Errors-in-variables model. Journal of Multivariate Analysis, Vol.97(2006), p.324-341.

[10] C H Wei: The Theory and Application of Semi-parametric Model (Science and Technology Publications, Beijing 2013).

[11] Y. P Mack, B W. Silverman. Weak and Strong Uniform Consistency of Kernel Regression Estimates.Z.W ahrsch. Verw. Gebiete. Vol.61(1982), p.405-415.