Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models

Arman Jahangiri

December 5^{th}, 2023

UNIVERSITY OF CALGARY

Outline

Outline

Literature Review

- During the past three decades, advancements in computing have fueled a surge of interest in nonparametric models, driven by the realization that parametric models often fall short in capturing complex relationships.
  However, the flexibility of nonparametric approaches can hinder drawing concise conclusions.

- Semiparametric approaches are of high interest since they can handle a trade-off between interpretability and flexibility.

Literature Review

- To address limitations in tools for making inferences on semiparametric and nonparametric models, Fan et al. (2001) proposed the generalized likelihood ratio (GLR) statistic.
  The GLR test, under the null hypothesis, follows an asymptotically χ^2 distribution, a property known as the Wilk's phenomenon.

- Focus of this project:
  Fan, J. Q., & Huang, T. (2005). ***Profile likelihood inferences on semiparametric varying-coefficient partially linear models*. Bernoulli, 11(6), 1031–1057.**

Outline

Introduction

- Varying-coefficient partially linear models (VC-PLM) are used particularly when dealing with complex relationships and heterogeneous data (data coming from different sources in various formats).

- VC-PLM can capture nonlinearities in the relationships between some covariates and the response variable by allowing the coefficients to vary across observations.

- VC-PLM combines parametric and nonparametric components. (i,e. linear and nonlinear terms)

Model Structure and Assumptions

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

- $\alpha(\cdot) = (\alpha_1(\cdot), \cdots \alpha_p(\cdot))^T$: A p-dimensional vector of unkown coefficients

- $\beta^T = (\beta_1, \cdots, \beta^q)$ relates the variables $Z_1 \cdots Z_q$ to the response through a linear relationship.

Further Model Assumptions

- $E[\epsilon] = 0$;
- $Var(\epsilon) = \sigma^2$;
- $\epsilon \perp (U, X^T, Z^T)$
- Presence of interaction between the covariates U and X so that different levels of the covariate U corresponds to different linear models.

Profile least squares is a powerful approach and can be shown to have optimal trade-off between efficiency and flexibility for the proposed model.

Incorporating sample into the Model

Given a random sample of size n, namely $\{U_i, \mathbf{X}_i, \mathbf{Z}_i, Y_i : i = 1 \cdots n\}$, where $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})$ and $\mathbf{Z}_i = (Z_{i1}, \cdots, Z_{iq})$.

$$Y_i = \sum_{j=1}^{p} \alpha_j(U_i) X_{ij} + \sum_{j=1}^{n} \beta_j Z_{ij} + \epsilon_i \tag{2}$$

Varying-Coefficient Partially Linear Model (VC-PLM) ($\alpha_j(U_i)$ is not a

fixed constant such as β_j)) Can be converted to

$$Y_i^* = \sum_{j=1}^{p} \alpha_j(U_i) X_{ij} + \epsilon_i \tag{3}$$

with

$$Y_i^* = Y_i - \sum_{j=1}^{n} \beta_j Z_{ij} \tag{4}$$

Varying-coefficient model

Steps for estimating the coefficients

Needed Tools

- **Local Linear Regression (LLR)**: To estimate $\{\alpha_i(\cdot) : i = 1, \cdots, p\}$

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha'_j(u_0)(u - u_0) \equiv a_j + b_j(u - u_0) \tag{5}$$

- **Score functions**:

$$(\hat{a}, \hat{b}) = \underset{\text{a,b}}{argmin} \sum_{i=1}^{n} [Y_i^* - \sum_{j=1}^{p} \{a_j + b_j(U_i - u_0)X_{ij}\}]^2 K_h(U_i - u_0) \tag{6}$$

with K a kernel function and h a bandwidth and $k_h(\cdot) = K(\cdot/h)/h$.

Score Functions

- $\frac{\partial U(a,b)}{\partial a_j} = \sum_{i=1}^{n}[Y_i^* - \sum_{j=1}^{p}\{a_j + b_j(U_i - u_0)X_{ij}\}]K_h(U_i - u_0)X_{ij} = 0$
- $\frac{\partial U(a,b)}{\partial b_j} = \sum_{i=1}^{n}[Y_i^* - \sum_{j=1}^{p}\{a_j + b_j(U_i - u_0)X_{ij}\}]K_h(U_i - u_0)(U_i - u_0)X_{ij} = 0$

Conversion to Matrix Notation

Define

$$D_u = \begin{bmatrix} X_1^T & \frac{U_1 - u}{h} X_1^T \\ \vdots & \vdots \\ X_n^T & \frac{U_n - u}{h} X_n^T \end{bmatrix} \quad ; \quad W_u = \mathrm{diag}(K_h(U_i - u_0); i = 1, \cdots, n) \tag{10}$$

- D_u: Similar to X matrix in a general linear model setup, whose second set of values have been adjusted by the weights $\frac{U_i - u_0}{h}$.

- W_u: Similar to W (weight) matrix in the general linear model setup.

Answer to the weighted least square is then given by

$$[\hat{\mathbf{a}}, h\hat{\mathbf{b}}]^T = (D_u^T W_u D_u)^{-1} D_u^T W_u (Y - Z\beta) \tag{9}$$

Conversion to Matrix Notation

Defining

$$M = \begin{bmatrix} \alpha^T(U_1)X_1 \\ \vdots \\ \alpha^T(U_n)X_n \end{bmatrix} = \begin{bmatrix} \alpha^T(U_1) \\ \vdots \\ \alpha^T(U_n) \end{bmatrix} \times \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix} \tag{11}$$

results in (3) being converted to $Y - Z\beta = M + \epsilon$.
We thus have

$$\hat{M} = \begin{bmatrix} \begin{bmatrix} X_1^T & 0 \end{bmatrix} (D_{u1}^T W_{u_1} D_{u_1})^{-1} D_{u_1}^T W_{u_1} \\ \vdots \\ \begin{bmatrix} X_n^T & 0 \end{bmatrix} (D_{u_n}^T W_{u_n} D_{u_n})^{-1} D_{u_n}^T W_{u_n} \end{bmatrix} (Y - Z\beta) = S(Y - Z\beta) \tag{12}$$

Substituting $\hat{M}$ in $Y - Z\beta = M + \epsilon$, we get,

$$Y - Z\beta = S(Y - Z\beta) + \epsilon \Longrightarrow (I - S)Y = (I - S)Z\beta + \epsilon \tag{13}$$

Using the IWLS method in gerenal linear models once again for finding the coefficients vector β, we get

$$\hat{\beta} = \{Z^T(I - S)^T(I - S)Z\}^{-1} Z^T(I - S)^T(I - S)Y \tag{14}$$

Outline

Coefficients Assymptotic Distribution

Theorem 1. Under the conditions given in the appendix, the profile likelihood estimator of β is asymptotically normal,

$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, \Sigma)$$

where $\Sigma = \sigma^2\{E(ZZ^T) - E[E(ZX^T|U)E(XX^T|U)^{-1}E(XZ^T|U)]\}^{-1}$.

Sketch of proof

Proof of Theorem 4.1. By (2.6), we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{n}(\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})(\mathbf{M} + \boldsymbol{\varepsilon}).$$

By Lemmas A.2 and A.4, the bias term

$$\sqrt{n}(\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})\mathbf{M} = O_p(\sqrt{n}c_n^2).$$

Consider the stochastic term. By Lemma A.2, we have

$$n^{1/2}(\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})\boldsymbol{\varepsilon} = n^{-1/2}\sigma^{-2}\boldsymbol{\Sigma}\widetilde{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})\boldsymbol{\varepsilon}\{1 + o_P(1)\}. \tag{A.5}$$

Note that,

$$\widetilde{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})\boldsymbol{\varepsilon} = \sum_{i=1}^{n} \widetilde{\mathbf{Z}}_i\{\varepsilon_i - [\mathbf{X}_i^{\mathrm{T}}, 0]\{\mathbf{D}_{u_i}^{\mathrm{T}}\mathbf{W}_{u_i}\mathbf{D}_{u_i}\}^{-1}\mathbf{D}_{u_i}^{\mathrm{T}}\mathbf{W}_{u_i}\boldsymbol{\varepsilon}\}.$$

By using the same argument as before, we have

$$[\mathbf{X}^{\mathrm{T}}, 0]\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\boldsymbol{\varepsilon} = \mathbf{X}^{\mathrm{T}}\Gamma(U)^{-1}\mathrm{E}(\mathbf{X}|U)O_p(c_n).$$

Then we can show that

$$\widetilde{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})\boldsymbol{\varepsilon} = \sum_{i=1}^{n}\{\mathbf{Z}_i - \Phi(U_i)^{\mathrm{T}}\Gamma(U_i)^{-1}\mathbf{X}_i\}\varepsilon_i\{1 + o_P(1)\}.$$

By the Slutsky theorem and the central limit theorem, we have

$$n^{-1/2}\sigma^{-2}\widetilde{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})\boldsymbol{\varepsilon} \to N(0, \boldsymbol{\Sigma}^{-1}).$$

This, together with (A.5), proves the result. □

Wald Statistic Asymptotic Distribution

Theorem 2. Under the null hypothesis $A\beta = 0$ and the conditions in the appendix, the Wald statistic $W_n(h)$ follows the assymptotic χ^2 distribution with l degrees of freedom.

$$W_n(h) = \hat{\beta}^T A^T (A \hat{\sum}_h A^T)^{-1} A\hat{\beta} \xrightarrow{P} \chi_l^2(\lambda)$$

- Theorem 2 follows from 1 and Slutsky theorem.
  (if $X \sim N_p(\mu, \sum)$ then $(X-\mu)^T \sum^{-1}(X-\mu) \sim \chi^2(p)$.)

Outline

Profile Likelihood Ratio (PLR) Test

- **Null Hypothesis:**

$$H_0 : A\beta = 0$$

- **Alternative Hypothesis:**

$$H_1 : A\beta \neq 0$$

where:

$A_{l \times q}$ is a known full-rank matrix specifying the linear combination,

$\beta_{q \times 1}$ is the vector of regression coefficients,

Here, the problem is that we cannot use the MLR test since the nonparametric MLEs do not exist for $\alpha(.)$

- **Question.** How to proceed?

1. Find a reasonable non-parametric estimate for $\alpha(.)$
(done by via Profile Likelihood Estimation)

2. Plug it in when they appear in the Likelihood Ratio statistic

- Generalized Likelihood Ratio (GLR) statistic;
- Profile Likelihood Ratio (PLR) statistic as the nonparametric estimate of $\alpha(.)$ is found via profile likelihood technique.

- Assume (WLOG: Fan et. al 2001): $\epsilon_i \sim N(0, \sigma^2)$,
- $Y_i = \alpha(U_i)^T X_i + \beta^T Z_i + \epsilon_i \mid U_i,\, X_i,\, Z_i, \sim N(\alpha(U_i)^T X_i + \beta^T Z_i, \sigma^2)$.
- log-likelihood function

$$l(\alpha, \beta, \sigma) = -n \log(\sqrt{2\pi}\sigma) - \frac{RSS1}{2\sigma^2} \tag{15}$$

with $RSS1 = \sum_{i=1}^{n} (Y_i - \alpha(U_i)^T X_i - \beta^T Z_i)^2$

- Estimate $\alpha(.)$ using LLR for fixed β, with $\hat{\alpha}(.;\beta)$,
- Substitute $\hat{\alpha}(.;\beta)$ into (15):
- $$l(\hat{\alpha}.;\beta),\beta,\sigma)) = -nlog(\sqrt{2\pi}\sigma) - \frac{\hat{R}SS1}{2\sigma^2} \tag{16}$$
- Differentiating (16) with respect to σ, we have

  $$\frac{\partial l(\hat{\alpha}.;\beta),\beta,\sigma))}{\partial\sigma^2} = -n/\sigma - \frac{RSS1}{\sigma^3} = 0 \Longrightarrow \hat{\sigma}^2 = n^{-1}RSS1$$

  Using a similar argument and using the derivations in section **3**, we find $\hat{\beta}$ as well to have the formula found in that section.

Substituting the found estimators into the log-likelihood function yields the Profile Likelihood

$$l(H_1) = -\frac{n}{2}log(\frac{2\pi}{n}) - \frac{n}{2}log(RSS1) - \frac{n}{2}$$

A similar approach can be adopted to find $l(H_0)$, the maxized log-likelihood function under the null hypothesis $H_0: \ A\beta = 0$. Then, we have,

$$T_n = l(H_1) - l(H_0) = \frac{n}{2}log(\frac{RSS_0}{RSS_1}) \approx \frac{n}{2}\frac{RSS_0 - RSS1}{RSS_1}$$

using the Taylor series expansion $-log(1-x) = x + \frac{x^2}{2} + \cdots$ with the first term being incorporated and $x = \frac{RSS_1 - RSS_0}{RSS_1}$.

T_n is referred to as the PLR statistic, which is not the same as the MLR test statistic.

The PLR statistic is derived analogously to the maximum likelihood ratio statistic. However, they are different since $\alpha_i(.)$ is found through LLR.

Outline

1. Literature Review
2. Introducing the Model
3. Derivation of the Asymptotic results
4. Hypothesis Testing
5. Comparison
6. Conclusion
7. References

Comparison with traditional LRT Distribution

Question Is PLR statistic asymptotically χ^2 (similiar to the traditional MLR statistic)?

Theorem 3 Under the null hypothesis $A\beta = 0$ and the conditions in the Appendix, the PLR statistic $2T_n(h)$ follows the asymptotic $\chi^2(df = l)$.

Furthermore,

Wald Test The testing problem $A\beta = 0$ can also be performed using the Wald statistic,

$$W_u(h) = \hat{\beta}^T A^T (A \hat{\sum_h} A^T)^{-1} A\hat{\beta}$$

where $\hat{\sum}_h$ is an estimation of $Cov(\beta)$. The asymptotic distribution of the Wald statistic also follows a $\chi^2(df = l)$ distribution, which was illustrated in **Theorem 2**.

Outline

Conclusion

- In the ever-evolving landscape of statistical modeling, exploration into semiparametric varying-coefficient partially linear models represents a significant importance toward addressing the complexities of parametric estimation.
- The adoption of the Generalized Likelihood Ratio (GLR) statistic, with a focus on the Profile Likelihood Ratio (PLR) test, gives us a powerful test comparable to the traditional LRT.
- In an era marked by complex datasets, the need for methodologies that balance flexibility and precision is necessary. The asymptotic properties established in the derivations lay a foundation for confidence in the estimators and tests proposed, further emphasizing their utility in practical scenarios.
- Beyond the confines of this project, the implications of our findings resonate with the broader landscape of statistical modeling.

Outline

References

[1] Fan, J. Q., & Huang, T. (2001). *Generalized Likelihood Ratio Statistics and Wilks Phenomenon*. Bernoulli, 11(6), 1031–1057.

[2] Fan, J. Q., & Huang, T. (2005). *Profile likelihood inferences on semiparametric varying-coefficient partially linear models*. Bernoulli, 11(6), 1031–1057.

[3] He, B., Yu, M., & Zhou, J. (2021). *Statistical inference for partially linear errors-in-variables panel data models with fixed effects. Systems Science & Control Engineering*, 9(1), 1-10. DOI: 10.1080/21642583.2020.1856212

Thank You!

Questions...