Published in final edited form as:

Stat Probab Lett. 2017 March ; 122: 11–19. doi:10.1016/j.spl.2016.10.020.

Focused Information Criterion and Model Averaging with Generalized Rank Regression

Qingzhao Zhang^{1,2}, Xiaogang Duan³, and Shuangge Ma^{1,2,*}

¹School of Economics and the Wang Yanan Institute for Studies in Economics, Xiamen University

²Department of Biostatistics, Yale University

³Department of Statistics, Beijing Normal University

Abstract

Generalized rank regression, which is a class of weighted rank regression with weights based on factor space, provides a powerful tool for conducting robust estimation. In this article, we first establish the asymptotic properties of generalized rank regression under local model misspecification. We then apply the generalized rank regression to the focus information criterion and frequentist model averaging and establish their properties.

Keywords

Focus Information Criterion; Frequentist Model Averaging; Generalized Rank Regression; Local Model Misspecification

1 Introduction

Variable and model selection is an important objective in statistical practice and can facilitate both model interpretation and prediction. “Traditional” variable and model selection procedures aim at identifying a single model with overall good properties and take this model as the “true” one for all downstream analysis. It has been realized that this single selected model is not necessarily well tailored to a specific parameter under focus. See Section 4.4 of Claeskens and Hjort (2003) for an inspiring example. In addition, this procedure neglects the uncertainty in the selection process and can lead to underestimation in inference.

Hjort and Claeskens (2003) and Claeskens and Hjort (2003) partly address the above concerns by developing a focus information criterion (FIC) and a frequentist model averaging (FMA) procedure under a locally misspecified likelihood framework. The FIC encourages different models to be selected for different parameters of interest, while the

* shuangge.ma@yale.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

FMA estimates the target parameter by a weighted average of the estimators associated with a series of plausible models. In addition to linear regression, the FIC and FMA have been extended to the Cox's proportional hazard model (Hjort and Claeskens 2006), semiparametric models (Claeskens and Carroll 2007, Zhang and Liang 2011), quantile regression (Xu et al. 2013), among others. See Claeskens and Hjort (2008) for a comprehensive survey.

In the existing FIC and FMA studies, estimation and inference have been based on the likelihood, least squares, and quantile regression. Beyond these techniques, a competitive estimation and inference procedure is the generalized rank (GR) regression (Hettmansperger and McKean 1998; Terpstra and McKean 2005), which includes the rank regression (McKean 2004) as a special case. The GR regression is robust to outliers in both response variables and predictors and usually more efficient than the least squares and quantile regression for long-tailed error distributions. With such appealing features, the GR regression has been well studied in the "traditional" variable selection context. For example, Wang and Li (2009) studies the weighted Wilcoxon-type smoothly clipped absolute deviation method for robust estimation and selection of important variables. Leng (2010) considers variable selection and coefficient estimation using the regularized rank regression. However, these studies still target at selecting the universally optimal model and may not be the best choice for a specific parameter under focus.

Motivated by these considerations, in this study, our goal is to investigate the FIC and FMA under the framework of locally misspecified linear regression, with the GR regression technique for estimation. Statistical properties are rigorously established, rendering the proposed approach a solid theoretical ground.

2 GR estimation under the local misspecification framework

The local model misspecification framework has been established in the literature (Claeskens and Hjort 2003). Briefly, consider the regression model

$$Y = \alpha^* + X^{\text{T}} \beta^* + Z^{\text{T}} \delta / \sqrt{n} + \epsilon, \quad (1)$$

where $Y \in$ is the univariate response variable, α^* is the true value of intercept parameter, X is the length- p vector of random covariates that are certainly included in the model, β^* is the vector of regression coefficients corresponding to X , Z is the length- q vector of random covariates that are not guaranteed to be included in the model, δ is a fixed length- q vector that does not vary with n , and ϵ is the error term with density function f and distribution function F . Without loss of generality, assume that each component of X and Z has mean zero. For more details on this modeling framework, refer to Hjort and Claeskens (2003, 2006). Following Hjort and Claeskens (2003), we call model (1) – which does not include Z – as the "narrow model" and the one including all covariates as the "wide model".

Denote $(y_i, x_i, z_i)_{i=1}^n$ as the n iid observations generated from model (1). Let π_S be an $|S| \times q$ projection matrix, which maps a column vector $a = (a_1, \dots, a_q)^{\text{T}}$ to its subvector a_S whose

components have indexes in S . Here $|S|$ denotes the cardinality of S . When referring to a submodel S , we mean that all estimation procedures are based on $(y_i, x_i, \pi_S z_i)_{i=1}^n$. Let $y_{ij} = y_i - y_j$, $x_{ij} = x_i - x_j$ and $z_{ij} = z_i - z_j$, $z_{ij,S} = \pi_S z_{ij}$. Denote $(\alpha, \beta^\top, \gamma_S^\top)^\top$ as the parameter vector in the submodel S , and S^c as the complement of set S in $\{1, \dots, q\}$.

The GR estimate $(\hat{\beta}_S^\top, \hat{\gamma}_S^\top)^\top$ is defined as the minimizer of the objective function

$$G_{n,S}(\beta, \gamma_S) = (2n)^{-1} \sum_{i,j} b_{ij} |y_{ij} - x_{ij}^\top \beta - z_{ij,S}^\top \gamma_S|, \quad (2)$$

where b_{ij} 's are positive and symmetric (that is, $b_{ij} = b_{ji}$) functions designed to accommodate disperse predictors. With a slight abuse of notation, we use $\hat{\beta}_S$ to denote the estimate of β^* based on the submodel indexed by S . When b_{ij} 's are constant, minimizing $G_{n,S}(\beta, \gamma_S)$ is equivalent to minimizing the Wilcoxon-type dispersion function. In this sense, Wang and Li (2009) also calls the approach based on minimizing (2) as the weighted Wilcoxon-type method. A popular choice for weights (Terpstra and McKean 2005; Wang and Li 2009) is:

$b_{ij} = h(t_i)h(t_j)$, where $t_i = (x_i^\top, z_i^\top)^\top$ and

$$h(t_i) = \min \left\{ 1, \frac{b}{(t_i - \hat{\mu})^\top \hat{\Sigma}^{-1} (t_i - \hat{\mu})} \right\},$$

with $(\hat{\mu}, \hat{\Sigma})$ being the robust minimum volume ellipsoid estimator of the location and scatter and b being the 95th percentile of $\chi^2(p+q)$.

To establish asymptotic properties, we assume the following regularity conditions.

- (i) ϵ is independent of X and Z , and its distribution satisfies $F(0) = 0.5$.
- (ii) $f(\cdot)$ is absolutely continuous with finite Fisher information, i.e., $\int f^{-1}(e)f'(e)^2 de < \infty$.
- (iii) There exist $(p+q) \times (p+q)$ positive-definite matrices C and V such that

$$n^{-1} \left\{ \sum_i^n \sum_j^n \sum_k^n w_{ij} w_{kj} (x_{ij}^\top, z_{ij}^\top)^\top (x_{kj}^\top, z_{kj}^\top)^\top \right\} \xrightarrow{p} V, \\ (2n)^{-1} \sum_{i,j} w_{ij} (x_{ij}^\top, z_{ij}^\top)^\top (x_{ij}^\top, z_{ij}^\top)^\top \xrightarrow{p} C,$$

where for $i, j = 1, \dots, n$,

$$w_{ij} = \begin{cases} -n^{-1}b_{ij}, & i \neq j, \\ n^{-1}\sum_{k \neq i} b_{ik}, & i = j. \end{cases}$$

Let $\tau = \left(\sqrt{12} \int f^2(t) dt \right)^{-1}$. For the positive-definite matrices $V = (V_{lk})_{l,k=0,1}$ and $C = (C_{lk})_{l,k=0,1}$, denote their inverse matrices as $V^{-1} = (V^{lk})_{l,k=0,1}$ and $C^{-1} = (C^{lk})_{l,k=0,1}$, respectively. Denote $Q = V^{11} = (V_{11} - V_{10}V_{00}^{-1}V_{01})^{-1}$ and $K = C^{11} = (C_{11} - C_{10}C_{00}^{-1}C_{01})^{-1}$. Further denote

$$C_s = \begin{pmatrix} C_{00} & C_{01}\pi_s^\top \\ \pi_s C_{10} & \pi_s C_{11}\pi_s^\top \end{pmatrix}, \text{ with its inverse } C_s^{-1} = \begin{pmatrix} C^{00,S} & C^{01,S} \\ C^{10,S} & C^{11,S} \end{pmatrix}.$$

Denote $K_s = C^{11,S} = (\pi_s K^{-1} \pi_s^\top)^{-1}$. It can be shown that

$$C^{00,S} = C_{00}^{-1} + C_{00}^{-1} C_{01} \pi_s^\top K_s \pi_s C_{10} C_{00}^{-1} \text{ and } C^{01,S} = -C_{00}^{-1} C_{01} \pi_s^\top K_s.$$

Let M be a random vector of size p and U be a random vector of size q such that $(M^\top, U^\top)^\top \sim N(\mathbf{0}_{(p+q) \times 1}, \tau^2 V)$. Denote $U_s = \pi_s U$. The asymptotic properties of $(\hat{\beta}_s^\top, \hat{\gamma}_s^\top)^\top$ can be established as follows.

Lemma 1

For the true data generating model (1), under conditions (i)–(iii),

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_s - \beta^*) \\ \sqrt{n}\hat{\gamma}_s \end{pmatrix} \xrightarrow{D} C_s^{-1} \left\{ \begin{pmatrix} C_{01} \\ \pi_s C_{11} \end{pmatrix} \delta + \begin{pmatrix} M \\ U_s \end{pmatrix} \right\}. \quad (3)$$

Note that $G_{n,s}(\beta, \gamma_s)$ does not depend on α . Estimating α demands one more step. After

obtaining $(\hat{\beta}_s^\top, \hat{\gamma}_s^\top)^\top$, we define the estimator of α^* as

$$\hat{\alpha}_s = \text{median of } \left\{ y_i - x_i^\top \hat{\beta}_s - z_{i,s}^\top \hat{\gamma}_s \right\}_{i=1}^n \text{ based on condition (i), where } z_{i,s} = \pi_s z_i.$$

Let $\tau_1 = (2\kappa(0))^{-1}$. The following lemma establishes the asymptotic properties of $\hat{\alpha}_s$.

Lemma 2

For the true data generating model (1), under conditions (i)–(iii),

$$\sqrt{n}(\hat{\alpha}_s - \alpha^*) \xrightarrow{D} N(0, \tau_1^2). \quad (4)$$

The proofs of Lemma 1 and 2 are presented in Section 5. Combining these two lemmas, we can establish Theorem 1 following similar arguments as in the proof of Theorem 3.5.11 in Hettmansperger and McKean (2003).

Theorem 1

For the true data generating model (1), under conditions (i)–(iii),

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_S - \alpha^* \\ \hat{\beta}_S - \beta^* \\ \hat{\gamma}_S \end{pmatrix} \xrightarrow{D} \begin{pmatrix} 1 & \mathbf{0}_{1 \times p} & \mathbf{0}_{1 \times |S|} \\ \mathbf{0}_{p \times 1} & C^{00,S} & C^{01,S} \\ \mathbf{0}_{|S| \times 1} & C^{10,S} & C^{11,S} \end{pmatrix} \left\{ \begin{pmatrix} \mathbf{0}_{1 \times q} \\ C_{01} \\ \pi_S C_{11} \end{pmatrix} \delta + \begin{pmatrix} J \\ M \\ U_S \end{pmatrix} \right\}, \quad (5)$$

where J is independent of M and U , and $J \sim N(0, \tau_1^2)$.

In the same line as Hettmansperger and McKean (1998), Terpstra and McKean (2005), and Wang and Li (2009), Theorem 1 also shows that the GR estimate under the local misspecification framework is robust to outliers in both responses and covariates. When $b_{ij} \equiv 1$, following the same discussions in Hettmansperger and McKean (1998) and Leng (2010), the asymptotic relative efficiency of the rank method with respect to the least squares method is never below 0.864 and is much higher than 1 for long-tailed error distributions. Moreover, the rank estimate is asymptotically much more efficient than the least absolute deviation estimate – which is a special case of quantile regression – for many common distributions.

Remark 1

Let $\theta^* = (\alpha^*, \beta^{*\top})^\top$ be the protected coefficients (their corresponding covariates are included in all of the submodels), and I_p be the $p \times p$ identity matrix. It can be seen from Theorem 1 that, for the wide model with $S = \{1, 2, \dots, q\}$,

$$\sqrt{n} (\hat{\theta}_{wide} - \theta^*) \xrightarrow{D} N \left(\mathbf{0}_{(p+1) \times 1}, \begin{bmatrix} \tau_1^2 & \mathbf{0}_{1 \times p} \\ \mathbf{0}_{p \times 1} & \tau^2 (I_p, \mathbf{0}_{p \times q}) C^{-1} V C^{-1} (I_p, \mathbf{0}_{p \times q})^\top \end{bmatrix} \right).$$

For the narrow model with $S = \emptyset$,

$$\sqrt{n} (\hat{\theta}_{narrow} - \theta^*) \xrightarrow{D} N \left(\begin{pmatrix} 0 \\ C_{00}^{-1} C_{01} \delta \end{pmatrix}, \begin{bmatrix} \tau_1^2 & \mathbf{0}_{1 \times p} \\ \mathbf{0}_{p \times 1} & \tau^2 C_{00}^{-1} V_{00} C_{00}^{-1} \end{bmatrix} \right).$$

Calculation shows that the asymptotic variance of $\sqrt{n} (\hat{\theta}_{wide} - \theta^*)$ is bigger than that of $\sqrt{n} (\hat{\theta}_{narrow} - \theta^*)$, while the latter is asymptotically biased. This suggests that there is a tradeoff in terms of squared bias and variance between the wide and narrow models. In the next section, we introduce criteria to effectively deal with this tradeoff.

3 FIC and FMA with the GR estimation

In this section, we apply the GR estimation to FIC and FMA. Compared to that based on for example likelihood-based estimation, this allow us to conduct more robust estimation for focus parameters. Building on the asymptotic results obtained in Section 2, we develop an FIC model selection and an FMA estimator.

3.1 FIC

Here we develop the FIC based on the GR estimation for a general focus parameter, say $\mu \equiv \mu(\theta^*, \delta / \sqrt{n})$, of the true model (1) with $\theta^* = (\alpha^*, \beta^{*\top})^\top$. For example, μ can be the median regression surface or the standard deviation function evaluated at a specific point, a quotient between two values of the mean regression function, and others. We refer to Claeskens and Hjort (2003) and Hjort and Claeskens (2003) for discussions on the focus parameter. Define

$$R = C_{10}C_{00}^{-1} - V_{10}V_{00}^{-1}, \quad H_S = \pi_S^\top K_S \pi_S,$$

and $\omega = C_{10}C_{00}^{-1} \frac{\partial \mu}{\partial \beta} - \frac{\partial \mu}{\partial \gamma}$ with the partial derivatives evaluated at $(\theta^{*\top}, \mathbf{0}^\top)^\top$. We first establish the following theorem which is crucial for examining the FIC.

Theorem 2—Assume the conditions in Section 2 and that μ is continuously differentiable at $(\theta, \mathbf{0})$. Let $\hat{\mu}_S = \mu(\hat{\theta}_S, (\hat{\gamma}_S, \mathbf{0}_{Sc}))$ be the estimator of μ based on the submodel indexed by S , and I_q be the $q \times q$ identity matrix. Then

$$\sqrt{n}(\hat{\mu}_S - \mu) \xrightarrow{D} \Lambda_S^{GR},$$

where

$$\Lambda_S^{GR} = \frac{\partial \mu}{\partial \alpha} J + \left\{ \left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1} + \omega^\top H_S R \right\} M + \omega^\top H_S Q^{-1} T + \omega^\top [I_q - H_S K^{-1}] \delta,$$

with $J \sim N(0, \tau_1^2)$, $M \sim N(\mathbf{0}_{p \times 1}, \tau^2 V_{00})$, and $T \sim N(\mathbf{0}_{q \times 1}, \tau^2 Q)$ mutually independent.

From Theorem 2, it is easy to see that Λ_S^{GR} is normally distributed with mean squared error

$$\begin{aligned} r(S) = & \tau_1^2 \left(\frac{\partial \mu}{\partial \alpha} \right)^2 + \tau^2 \omega^\top H_S Q^{-1} H_S \omega + \omega^\top [I_q - H_S K^{-1}] \delta \delta^\top [I_q - K^{-1} H_S] \omega \\ & + \tau^2 \left\{ \left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1} + \omega^\top H_S R \right\} V_{00} \left\{ \left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1} + \omega^\top H_S R \right\}^\top. \end{aligned} \quad (6)$$

We can assess and compare estimates based on different submodels using $\kappa(S)$'s. From Lemma 1, for the wide model we have

$$\Gamma_n \equiv \sqrt{n}\hat{\gamma}_{wide} \xrightarrow{D} \Gamma \equiv N(\delta, \tau^2[\mathbf{0}_{q \times p}, I_q]C^{-1}VC^{-1}[\mathbf{0}_{q \times p}, I_q]^\top).$$

It is reasonable to replace $\delta\delta^\top$ by $n\hat{\gamma}_{wide}\hat{\gamma}_{wide}^\top - \tau^2[\mathbf{0}_{q \times p}, I_q]C^{-1}VC^{-1}[\mathbf{0}_{q \times p}, I_q]^\top$. Following Hjort and Claeskens (2003) and (6), the FIC of submodel S is defined as

$$FIC(S) \equiv \Pi_1 H_S \omega + \omega^\top H_S \Pi_2 H_S \omega, \quad (7)$$

where

$$\begin{aligned} \Pi_1 &= 2\tau^2 \left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1} V_{00} R - 2\omega^\top \left[n\hat{\gamma}_{wide}\hat{\gamma}_{wide}^\top - \tau^2[\mathbf{0}_{q \times p}, I_q]C^{-1}VC^{-1}[\mathbf{0}_{q \times p}, I_q]^\top \right] K^{-1}, \\ \Pi_2 &= \tau^2 Q^{-1} + \tau^2 R V_{00} R + K^{-1} \left[n\hat{\gamma}_{wide}\hat{\gamma}_{wide}^\top - \tau^2[\mathbf{0}_{q \times p}, I_q]C^{-1}VC^{-1}[\mathbf{0}_{q \times p}, I_q]^\top \right] K^{-1}. \end{aligned}$$

It is noted that the FIC measure is a modification of $\kappa(S)$ by ignoring the terms shared by all submodels and rearranging the terms. It can be naturally used to balance the squared bias and variance.

The FIC measure contains unknown quantities. In practice, we can define an estimated FIC, $\widehat{FIC}(S)$, by replacing the unknown quantities by their sample counterparts. For more details, refer to Section 3.2 of Claeskens and Hjort (2003). Let $\hat{S}$ be defined through $\widehat{FIC}(\hat{S}) = \min_S \widehat{FIC}(S)$. Then the best estimator of μ is $\mu(\hat{\theta}_{\hat{S}}, (\hat{\gamma}_{\hat{S}}, \mathbf{0}_{\hat{S}^c}))$.

3.2 FMA

The definition, merit, and applications of FMA have been discussed elsewhere (Hjort and Claeskens 2003; Claeskens and Hjort 2008) under the likelihood and least squares estimation frameworks and will not be reiterated here. In this section, we examine FMA with the robust GR estimation.

Consider the focus parameter $\mu = \mu(\theta^*, \delta/\sqrt{n})$. Its model averaging estimate is defined as $\hat{\mu} = \sum_S w(S|\Gamma_n) \hat{\mu}_S$, where $\hat{\mu}_S = \mu(\hat{\theta}_S, (\hat{\gamma}_S, \mathbf{0}_{S^c}))$ is the estimate of μ based on the submodel indexed by S , and $w(S|\Gamma_n)$ is a nonnegative weight function that sums to one (over all submodels). For example, based on the FIC measures, we can build the smoothed FIC weights,

$$w(S|\Gamma_n) = \exp\left(-\frac{FIC(S)}{\tau^2 \omega^\top K Q^{-1} K \omega}\right) / \sum_S \exp\left(-\frac{FIC(S)}{\tau^2 \omega^\top K Q^{-1} K \omega}\right).$$

See Hjort and Claeskens (2003) for more discussions. Then we can establish the following results for the FMA estimator $\hat{\mu}$.

Theorem 3—Assume the conditions in Section 2 and that the discontinuities of the weight function $w(S|\Gamma)$ have a zero measure with respect to the distribution of Γ . Then it follows that

$$\begin{aligned}\sqrt{n}(\hat{\mu} - \mu) &\xrightarrow{D} \Lambda^{GR} = \sum_S w(S|\Gamma) \Lambda_S^{GR} \\ &= \frac{\partial \mu}{\partial \alpha} J + \left\{ \left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1} + \omega^\top \tilde{H}(E) R \right\} M \\ &\quad - \omega^\top \tilde{H}(E) Q^{-1} T + \omega^\top \left[I_q - \tilde{H}(E) K^{-1} \right] \delta,\end{aligned}$$

where J , M , and T are defined as in Theorem 2, and $\tilde{H}(E) = \sum_S w(S|\Gamma) H_S$.

This theorem establishes the asymptotic properties of the model averaging estimate under mild conditions. Inference can be conducted based on the asymptotic distribution result.

4 A special case

Consider b_{ij} in (2). The most important case is with $b_{ij} \equiv 1$ because of the simplicity as well as tight connection with the familiar Wilcoxon approach. Consider the estimation criterion

$$L_{n,S}(\beta, \gamma_S) = (2n)^{-1} \sum_{i,j} \left| y_{ij} - x_{ij}^\top \beta - z_{ij,S}^\top \gamma_S \right|. \quad (8)$$

With a slight abuse of notation, we still use the same notations as in the previous sections.

For this special case, $C = V = \Sigma$, where $n^{-1} \sum_{i=1}^n \begin{pmatrix} x_i^\top & z_i^\top \end{pmatrix}^\top \begin{pmatrix} x_i^\top & z_i^\top \end{pmatrix} \xrightarrow{p} \Sigma$. Let

$$\Sigma = \begin{pmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{pmatrix}, \Pi = \left(\Sigma_{11} - \Sigma_{10} \Sigma_{00}^{-1} \Sigma_{01} \right)^{-1}, \quad \Pi_S = \left(\pi_S \Pi^{-1} \pi_S^\top \right)^{-1},$$

$L_S = \pi_S^\top \Pi_S \pi_S$, and $\omega_1 = \Sigma_{10} \Sigma_{00}^{-1} \frac{\partial \mu}{\partial \beta} - \frac{\partial \mu}{\partial \gamma}$ with the partial derivatives evaluated at $(\theta^{*\top}, \mathbf{0}^\top)^\top$. Then we have the following corollary.

Corollary 1

Assume the conditions in Theorem 2 and $b_{ij} \equiv 1$. The limit distribution of $\sqrt{n}(\hat{\mu}_S - \mu)$ reduces to

$$\Lambda_S^R = \frac{\partial \mu}{\partial \alpha} J + \left(\frac{\partial \mu}{\partial \beta} \right)^\top \Sigma_{00}^{-1} M_1 + \omega_1^\top L_S \Pi^{-1} T_1 + \omega_1^\top \left(I_q - L_S \Pi^{-1} \right) \delta,$$

where $J \sim N(0, \tau_1^2)$, $M_1 \sim N(\mathbf{0}_{p \times 1}, \tau^2 \Sigma_{00})$, and $T_1 \sim N(\mathbf{0}_{q \times 1}, \tau^2 \boldsymbol{\pi})$ are mutually independent.

With the assistance of Corollary 1, we can re-examine the FIC statistic. By Lemma 1 with $b_{ij} \equiv 1$, $G_n \equiv \sqrt{n} \hat{\gamma}_{wide} \xrightarrow{D} G \sim N(\delta, \tau^2 \Pi)$. Following a similar way as in Section 3, we have

$$FIC(S) \equiv \omega_1^\top \left\{ (I_q - L_S \Pi^{-1}) \left[n \hat{\gamma}_{wide} \hat{\gamma}_{wide}^\top - \tau^2 \Pi \right] (I_q - L_S \Pi^{-1}) + \tau^2 L_S \Pi^{-1} L_S \right\} \omega_1.$$

We can also obtain the properties of FMA under this special case. Denote

$\hat{\mu} = \sum_s w(S|G_n) \hat{\mu}_s$, where notations have similar implications as in Section 3. Similarly, we can use the smoothed FIC weights,

$$w(S|G_n) = \exp\left(-\frac{FIC(S)}{\tau^2 \omega_1^\top \Pi \omega_1}\right) / \sum_S \exp\left(-\frac{FIC(S)}{\tau^2 \omega_1^\top \Pi \omega_1}\right).$$

Corollary 2

Assume the conditions in Theorem 3 and $b_{ij} \equiv 1$. Then it follows that

$$\begin{aligned} \sqrt{n}(\hat{\mu} - \mu) &\xrightarrow{D} \Lambda^R = \sum_S w(S|G) \Lambda_S^R \\ &= \frac{\partial \mu}{\partial \alpha} J + \left(\frac{\partial \mu}{\partial \beta}\right)^\top \Sigma_{00}^{-1} F + \omega_1^\top \sum_S w(S|G) L_S \Pi^{-1} (T_1 - \delta) + \omega_1^\top \delta, \end{aligned}$$

where J , M_1 , and T_1 are defined as in Corollary 1.

5 Proof

Proof of Lemma 1

Denote $u = \sqrt{n}(\beta_S - \beta^*)$ and $v_S = \sqrt{n}\gamma_S$. Consider

$$\begin{aligned} H_n(u, v_S) &= G_{n,S}(\beta, \gamma_S) - G_{n,S}(\beta^*, \delta / \sqrt{n}) \\ &= \frac{1}{2n} \sum_{i,j} b_{ij} \left\{ |y_{ij} - x_{ij}^\top \beta - z_{ij,S}^\top \gamma_S| - |y_{ij} - x_{ij}^\top \beta^* - z_{ij}^\top \delta / \sqrt{n}| \right\} \\ &= \frac{1}{2n} \sum_{i,j} b_{ij} \left\{ \left| \epsilon_{ij} - \frac{x_{ij}^\top u + z_{ij,S}^\top v_S - z_{ij}^\top \delta}{\sqrt{n}} \right| - |\epsilon_{ij}| \right\}. \end{aligned}$$

Obviously, $H_n(u, v_S)$ is minimized at $\hat{u}_S \equiv \sqrt{n}(\hat{\beta}_S - \beta^*)$ and $\hat{v}_S \equiv \sqrt{n}\hat{\gamma}_S$.

By applying the identity

$$(|r - s| - |r|) / 2 = s \{I(r < 0) - 1/2\} + \int_0^s \{I(r \leq t) - I(r \leq 0)\} dt \quad (9)$$

and let $\phi(S, \delta) = (u^\top, (v_S - \delta_S)^\top, -\delta_{S^c}^\top)^\top$, we have

$$\begin{aligned} H_n(u, v_S) &= n^{-3/2} \sum_{i,j}^n \{I(\epsilon_{ij} < 0) - 1/2\} b_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right) \phi(S, \delta) \\ &\quad + n^{-1} \sum_{i,j}^n b_{ij} \int_0^{\frac{\left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right) \phi(S, \delta)}{\sqrt{n}}} \{I(\epsilon_{ij} \leq t) - I(\epsilon_{ij} \leq 0)\} dt \\ &\equiv \Delta_n^\top \phi(S, \delta) + B_n(\delta), \end{aligned} \quad (10)$$

where $a^{\bar{r}}$ denotes a rearrangement of vector/matrix a to ensure that the row and/or column index of a coincides with (S, S^c) . Following Lemma 5.2.3 of Hettmansperger and McKean (2003),

$$\begin{aligned} \Delta_n &= n^{-3/2} \sum_{i,j}^n \{I(\epsilon_{ij} < 0) - 1/2\} b_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top \\ &= n^{-3/2} \sum_{i,j}^n [1 - 2F(\epsilon_j)] b_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top + o_p(1). \end{aligned} \quad (11)$$

Let $T(n, S) = n^{-3/2} \sum_{i,j} [1 - 2F(\epsilon_j)] b_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top$. It is easy to see that

$$T(n, S) = n^{-1/2} \sum_j \left[\frac{1}{n} \sum_i b_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top \right] [1 - 2F(\epsilon_j)]. \quad (12)$$

Obviously, by condition (i) and the fact that $F(\epsilon_j) \sim U[0, 1]$, $E[T(n, S)] = 0$. In addition, recall the definition of w_{ij} , we have

$$\begin{aligned} \text{Var}[T(n, S)] &= n^{-3} E \left\{ \sum_j \sum_i b_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top [1 - 2F(\epsilon_j)] \right\}^2 \\ &= n^{-3} \sum_j E \left\{ \sum_i b_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top [1 - 2F(\epsilon_j)] \right\}^2 \\ &= n^{-3} \sum_j E \left\{ [1 - 2F(\epsilon_j)]^2 \sum_i \sum_k b_{ij} b_{kj} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top \left(x_{kj}^\top, z_{kj}^{\bar{r}\top} \right) \right\} \\ &= (3n)^{-1} E \left\{ \sum_i \sum_j \sum_k w_{ij} w_{kj} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top \left(x_{kj}^\top, z_{kj}^{\bar{r}\top} \right) \right\} \\ &\rightarrow V^{\bar{r}}/3, \end{aligned} \quad (13)$$

where

$$V^{\bar{r}} = \begin{pmatrix} V_{00} & V_{01}\pi_S^\top & V_{01}\pi_{S^c}^\top \\ \pi_S V_{10} & \pi_S V_{11}\pi_S^\top & \pi_S V_{11}\pi_{S^c}^\top \\ \pi_{S^c} V_{10} & \pi_{S^c} V_{11}\pi_S^\top & \pi_{S^c} V_{11}\pi_{S^c}^\top \end{pmatrix}.$$

Moreover, from (12), conditional on (X, Z) , $T(n, S)$ is a sum of independent but not identically distributed random variables, whose conditional normality can be established via the Lindeberg-Feller central limit theorem. Then by Slutsky's Lemma, together with (11) and (13), we have

$$\Delta_n \xrightarrow{D} \Delta, \quad \text{with} \quad \Delta \sim N_{p+q} \left(\mathbf{0}, V^{\bar{r}}/3 \right). \quad (14)$$

Next, we want to verify that $B_n(\delta) = D_n(\delta) + o_p(1)$, where

$$D_n(\delta) = n^{-1} \sum_{i,j} b_{ij} \int_0^n \frac{\left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right) \phi(S, \delta)}{\sqrt{n}} \{E_\epsilon [F(\epsilon+t) - F(\epsilon)]\} dt. \quad (15)$$

In fact, $Var(B_n(\delta)|X, Z) \rightarrow 0$, then $D_n(\delta) = E[B_n(\delta)|X, Z]$. Moreover, by conditions (ii), (iii) and the definition of C , we have

$$\begin{aligned} D_n(\delta) &= n^{-1} \sum_{i,j} b_{ij} \int_0^n \frac{\left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right) \phi(S, \delta)}{\sqrt{n}} \int f(\epsilon)^2 d\epsilon t (1+o(1)) dt \\ &\xrightarrow{p} \left(4\sqrt{3}\tau \right)^{-1} \phi(S, \delta)^\top \left\{ n^{-1} \sum_{i,j} w_{ij} \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right)^\top \left(x_{ij}^\top, z_{ij}^{\bar{r}\top} \right) \right\} \phi(S, \delta) \\ &\xrightarrow{p} \left(2\sqrt{3}\tau \right)^{-1} \phi(S, \delta)^\top C^{\bar{r}} \phi(S, \delta). \end{aligned} \quad (16)$$

Therefore, by (10), (14), and (16),

$$H_n(u, v_S) \xrightarrow{D} H(u, v_S) = \Delta^\top \phi(S, \delta) + \left(2\sqrt{3}\tau \right)^{-1} \phi(S, \delta)^\top C^{\bar{r}} \phi(S, \delta). \quad (17)$$

Recall that

$$C = \begin{pmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{pmatrix} \quad \text{and} \quad C_S = \begin{pmatrix} C_{00} & C_{01}\pi_S^\top \\ \pi_S C_{10} & \pi_S C_{11}\pi_S^\top \end{pmatrix}.$$

Let $\Delta = (\Delta_0^\top, \Delta_1^\top)^\top$, $\Delta_S = (\Delta_0^\top, \Delta_1^\top \pi_S^\top)^\top$, and $S^c = \pi_S^c$. Consequently,

$$\begin{aligned}
H(u, v_s) &= (2\sqrt{3}\tau)^{-1} \begin{pmatrix} u \\ v_s - \delta_s \end{pmatrix}^\top C_s \begin{pmatrix} u \\ v_s - \delta_s \end{pmatrix} + (2\sqrt{3}\tau)^{-1} \delta_{sc}^\top \pi_{sc} C_{11} \pi_{sc}^\top \delta_{sc} \\
&\quad - (\sqrt{3}\tau)^{-1} \left\{ \begin{pmatrix} C_{01} \pi_{sc}^\top \\ \pi_s C_{11} \pi_{sc}^\top \end{pmatrix} \delta_{sc} - \sqrt{3}\tau \Delta_s \right\}^\top \begin{pmatrix} u \\ v_s - \delta_s \end{pmatrix} - \Delta_{sc}^\top \delta_{sc}.
\end{aligned}$$

Thus, $H_S(u, v_S)$ is minimized at

$$\begin{aligned}
&\begin{pmatrix} \mathbf{0} \\ \delta_s \end{pmatrix} + C_s^{-1} \left\{ \begin{pmatrix} C_{01} \pi_{sc}^\top \\ \pi_s C_{11} \pi_{sc}^\top \end{pmatrix} \delta_{sc} - \sqrt{3}\tau \Delta_s \right\} \\
&= C_s^{-1} \left\{ C_s \begin{pmatrix} \mathbf{0} \\ \delta_s \end{pmatrix} + \begin{pmatrix} C_{01} \pi_{sc}^\top \\ \pi_s C_{11} \pi_{sc}^\top \end{pmatrix} \delta_{sc} - \sqrt{3}\tau \Delta_s \right\} \\
&= C_s^{-1} \left\{ \begin{pmatrix} C_{01} \pi_s^\top \delta_s \\ \pi_s C_{11} \pi_s^\top \delta_s \end{pmatrix} + \begin{pmatrix} C_{01} \pi_{sc}^\top \delta_{sc} \\ \pi_s C_{11} \pi_{sc}^\top \delta_{sc} \end{pmatrix} - \sqrt{3}\tau \Delta_s \right\} \\
&= C_s^{-1} \left\{ \begin{pmatrix} C_{01} \\ \pi_s C_{11} \end{pmatrix} \delta - \sqrt{3}\tau \Delta_s \right\} \equiv C_s^{-1} \left\{ \begin{pmatrix} C_{01} \\ \pi_s C_{11} \end{pmatrix} \delta + \begin{pmatrix} M \\ U_s \end{pmatrix} \right\}, \quad (18)
\end{aligned}$$

where

$$\begin{pmatrix} M \\ U_s \end{pmatrix} = -\sqrt{3}\tau \Delta_s \sim N(0, \tau^2 V_s).$$

Since $L_{n,S}$ is convex, following Knight (1998), the desired result can be proved.

Proof of Lemma 2

Denote the expectation of $(A^\top, B^\top)^\top$ by $u_{ab} = (u_a^\top, u_b^\top)^\top$ for any (A, B) . Note that

$\hat{\alpha}_S = \arg \min_{\alpha} \sum_i |y_i - \alpha - x_i^\top \hat{\beta}_S - z_{i,S}^\top \hat{\gamma}_S|$. Define $\eta = \sqrt{n}(\alpha_S - \alpha^*)$. Then

$\hat{\eta} \equiv \sqrt{n}(\hat{\alpha}_S - \alpha^*)$ minimizes the objective function

$$\begin{aligned}
\Phi_n(\eta) &= \sum_i^n \left\{ \left| y_i - \alpha - x_i^\top \hat{\beta}_S - z_{i,S}^\top \hat{\gamma}_S \right| - \left| y_i - \alpha^* - x_i^\top \beta^* - z_{i,S}^\top \delta / \sqrt{n} \right| \right\} \\
&= \sum_i^n \left\{ \left| \epsilon_i - \frac{(x_i^\top, z_{i,S}^\top) \varphi_S + \eta}{\sqrt{\eta}} \right| - |\epsilon_i| \right\},
\end{aligned}$$

where $\varphi_S = \left\{ \sqrt{n}(\hat{\beta}_S - \beta^*)^\top, (\sqrt{n}\hat{\gamma}_S - \delta_S)^\top, -\delta_{sc}^\top \right\}^\top$. By using (9), we have

$$\begin{aligned}
\Phi_n(\eta) &= \frac{1}{\sqrt{n}} \sum_i^n \{I(\epsilon_i < 0) - 1/2\} \left[\left(x_i^\top, z_i^{\bar{r}\top} \right) \varphi_S + \eta \right] \\
&\quad + \sum_i^n \int_0^{\frac{\left(x_i^\top, z_i^{\bar{r}\top} \right) \varphi_S + \eta}{\sqrt{n}}} \{I(\epsilon_i \leq t) - I(\epsilon_i \leq 0)\} dt \\
&\equiv K(n, S)^\top \varphi_S + R_{n,S}(\delta).
\end{aligned}$$

Similar to the proof of Lemma 1,

$$\begin{aligned}
R_{n,S}(\delta) &= \sum_i^n \int_0^{\frac{\left(x_i^\top, z_i^{\bar{r}\top} \right) \varphi_S + \eta}{\sqrt{n}}} \{f(0)t\} dt + o_p(1) \\
&= f(0) \varphi_S^\top \frac{1}{n} \sum_i^n \begin{pmatrix} x_i^\top \\ z_i^{\bar{r}\top} \end{pmatrix} \left(x_i^\top, z_i^{\bar{r}\top} \right) \varphi_S / 2 + f(0) \eta^2 / 2 \\
&\quad + f(0) \varphi_S^\top \frac{\eta}{n} \sum_i^n \begin{pmatrix} x_i^\top \\ z_i^{\bar{r}\top} \end{pmatrix} + o_p(1) \\
&\xrightarrow{p} \frac{1}{2} f(0) \varphi_S^\top C^{\bar{r}} \varphi_S^\top + \frac{1}{2} f(0) \eta^2 + f(0) \varphi_S^\top \begin{pmatrix} u_x^\top, u_z^{\bar{r}\top} \end{pmatrix}^\top \eta.
\end{aligned}$$

Thus, $\Phi_n(\eta) - \Phi(\eta) \xrightarrow{p} 0$, where

$$\begin{aligned}
\Phi(\eta) &= \frac{1}{\sqrt{n}} \sum_i^n \{I(\epsilon_i < 0) - 1/2\} \left(x_i^\top, z_i^{\bar{r}\top} \right) \varphi_S + \frac{\eta}{\sqrt{n}} \sum_i^n \{I(\epsilon_i < 0) - 1/2\} \\
&\quad + \frac{1}{2} f(0) \varphi_S^\top C^{\bar{r}} \varphi_S^\top + \frac{1}{2} f(0) \eta^2 + f(0) \varphi_S^\top \begin{pmatrix} u_x^\top, u_z^{\bar{r}\top} \end{pmatrix}^\top \eta.
\end{aligned}$$

Note that the first and third terms of $\Phi(\eta)$ are not related to η . Thus we have that

$$\tilde{\eta} = \frac{1}{f(0)} \frac{1}{\sqrt{n}} \sum_i^n \{1/2 - I(\epsilon_i < 0)\} - \varphi_S^\top \begin{pmatrix} u_x^\top, u_z^{\bar{r}\top} \end{pmatrix}^\top$$

minimizes $\Phi(\eta)$. Note that $\frac{1}{f(0)} \frac{1}{\sqrt{n}} \sum_i^n \{1/2 - I(\epsilon_i < 0)\} \xrightarrow{D} N\left(0, \left(4f^2(0)\right)^{-1}\right)$. Recall the definition of C_S^{-1} in Section 2. By (17) and (18), it follows that

$$\begin{aligned}
\varphi_S^\top \begin{pmatrix} u_x^\top, u_z^{\bar{r}\top} \end{pmatrix}^\top &= \begin{pmatrix} u_x^\top, u_{z_S}^\top \end{pmatrix} C_S^{-1} \left\{ \begin{pmatrix} C_{01} \\ \pi_S C_{11} \end{pmatrix} \delta - \sqrt{3}\tau \Delta_S \right\} - u_{z_{S^c}}^\top \delta_{S^c} \\
&= u_{xz}^\top \begin{pmatrix} C^{00,S} & C^{01,S} & \mathbf{0}_{p \times |S^c|} \\ C^{10,S} & C^{11,S} & \mathbf{0}_{|S| \times |S^c|} \\ \mathbf{0}_{|S^c| \times p} & \mathbf{0}_{|S^c| \times |S|} & -I_{|S^c|} \end{pmatrix} \left\{ \begin{pmatrix} C_{01} \\ \pi_S C_{11} \end{pmatrix} \delta - \sqrt{3}\tau \begin{pmatrix} \Delta_S \\ \mathbf{0}_{|S^c| \times 1} \end{pmatrix} \right\} \\
&\xrightarrow{D} N \left(u_{xz}^\top \begin{pmatrix} C^{00,S} & C^{01,S} & \mathbf{0}_{p \times |S^c|} \\ C^{10,S} & C^{11,S} & \mathbf{0}_{|S| \times |S^c|} \\ \mathbf{0}_{|S^c| \times p} & \mathbf{0}_{|S^c| \times |S|} & -I_{|S^c|} \end{pmatrix} \begin{pmatrix} C_{01} \\ \pi_S C_{11} \end{pmatrix} \delta, \tau^2 u_{xz}^\top C_S^{-1} V_S C_S^{-1} u_{xz_S} \right).
\end{aligned}$$

Furthermore, the two terms of $\tilde{\eta}$ are asymptotically independent. By Knight (1998), $u_{xz} = \mathbf{0}_{(p+q) \times 1}$ and $\tau_1 = (2f(0))^{-1}$, we have $\sqrt{n}(\hat{\alpha}_s - \alpha^*) \xrightarrow{D} N(0, \tau_1^2)$.

Proof of Theorem 2

From Theorem 1, J is independent of M and U while $(M^\top, U^\top)^\top \sim N(\mathbf{0}_{(p+q) \times 1}, \tau^2 V)$. Write $T = Q(U - V_{10}V_{00}^{-1}M)$. Then $T \sim N(0, \tau^2 Q)$, and T and M are independent. Obviously, $U = Q^{-1}T + V_{10}V_{00}^{-1}M$. Recall that $U_S = \pi_S U$, the structure of C_S^{-1} , and

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_s - \alpha^* \\ \hat{\beta}_s - \beta^* \\ \hat{\gamma}_s \end{pmatrix} \xrightarrow{D} \begin{pmatrix} 1 & \mathbf{0}_{1 \times p} & \mathbf{0}_{1 \times |S|} \\ \mathbf{0}_{p \times 1} & C^{00,S} & C^{01,S} \\ \mathbf{0}_{|S| \times 1} & C^{10,S} & C^{11,S} \end{pmatrix} \left\{ \begin{pmatrix} \mathbf{0}_{1 \times q} \\ C_{01} \\ \pi_S C_{11} \end{pmatrix} \delta + \begin{pmatrix} J \\ M \\ U_S \end{pmatrix} \right\}.$$

Moreover, recall that $R = C_{10}C_{00}^{-1} - V_{10}V_{00}^{-1}$ and $H_s = \pi_s^\top K_s \pi_s$. It follows that

$$\begin{aligned} \sqrt{n}(\hat{\alpha}_s - \alpha_s^*) &\xrightarrow{D} J, \\ \sqrt{n}(\hat{\beta}_s - \beta_s^*) &\xrightarrow{D} (C_{00}^{-1} + C_{00}^{-1}C_{01}H_sR)M - C_{00}^{-1}C_{01}H_sQ^{-1}T + C_{00}^{-1}C_{01}[I_q - H_sK^{-1}]\delta, \\ \sqrt{n}\hat{\gamma}_s &\xrightarrow{D} -K_s\pi_sRM + K_s\pi_sQ^{-1}T + K_s\pi_sK^{-1}\delta. \end{aligned}$$

With these results, this theorem is a direct consequence of Lemma 1 combined with Taylor's expansion. More specifically,

$$\begin{aligned} \sqrt{n}(\hat{\mu}_s - \mu) &= \sqrt{n} \left\{ \mu(\hat{\theta}_s, (\hat{\gamma}_s, \mathbf{0}_{S^c})) - \mu(\theta^*, \delta/\sqrt{n}) \right\} \\ &= \sqrt{n} \left\{ \mu(\hat{\theta}_s, (\hat{\gamma}_s, \mathbf{0}_{S^c})) - \mu(\theta^*, \mathbf{0}) \right\} - \sqrt{n} \left\{ \mu(\theta^*, \delta/\sqrt{n}) - \mu(\theta^*, \mathbf{0}) \right\} \\ &\doteq \frac{\partial \mu}{\partial \alpha} \sqrt{n}(\hat{\alpha}_s - \alpha_s^*) + \left(\frac{\partial \mu}{\partial \beta} \right)^\top \sqrt{n}(\hat{\beta}_s - \beta_s^*) + \left(\frac{\partial \mu}{\partial \gamma} \right)_s^\top \sqrt{n}\hat{\gamma}_s - \left(\frac{\partial \mu}{\partial \gamma} \right)^\top \delta \\ &\xrightarrow{D} \Lambda_s \equiv \frac{\partial \mu}{\partial \alpha} J + \left\{ \left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1} + \left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1}C_{01}H_sR - \left(\frac{\partial \mu}{\partial \gamma} \right)^\top H_sR \right\} M \\ &\quad + \left\{ -\left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1}C_{01}H_s + \left(\frac{\partial \mu}{\partial \gamma} \right)^\top H_s \right\} Q^{-1}T \\ &\quad + \left\{ -\left(\frac{\partial \mu}{\partial \beta} \right)^\top C_{00}^{-1}C_{01} - \left(\frac{\partial \mu}{\partial \gamma} \right)^\top \right\} [I_q - H_sK^{-1}]\delta, \end{aligned}$$

where $\left(\frac{\partial \mu}{\partial \gamma} \right)_s = \pi_s \left(\frac{\partial \mu}{\partial \gamma} \right)$. Let $\omega = C_{10}C_{00}^{-1} \frac{\partial \mu}{\partial \beta} - \frac{\partial \mu}{\partial \gamma}$ with the partial derivatives evaluated at $(\theta^{*\top}, \mathbf{0}^\top)^\top$. Together with the definitions of J , M and T , we prove the result.

Acknowledgements

We thank the Editor and reviewer for their careful review and insightful comments. This study has been supported by National Natural Science Foundation of China (11401561, 210100113) and National Institutes of Health (R21CA191383, R01CA204120).

References

- [1]. Claeskens G, Hjort NL. The Focused Information Criterion (with discussions). *Journal of the American Statistical Association*. 2003; 98:900–916.
- [2]. Claeskens, G., Hjort, NL. *Model selection and model averaging*. Cambridge University Press; New York: 2008.
- [3]. Claeskens G, Carroll RJ. An asymptotic theory for model selection inference in general semiparametric problems. *Biometrika*. 2007; 94:249–265.
- [4]. Hettmansperger, TP., McKean, JW. *Robust nonparametric statistical methods*. Arnold/Wiley; London and New York: 1998.
- [5]. Hjort NL, Claeskens G. Frequentist Model Average Estimators (with discussions). *Journal of the American Statistical Association*. 2003; 98:879–899.
- [6]. Hjort NL, Claeskens G. Focussed information criteria and model averaging for Cox’s hazard regression model. *Journal of the American Statistical Association*. 2006; 101:1449–1464.
- [7]. Knight K. Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics*. 1998; 26:755–770.
- [8]. Leng C. Variable selection and coefficient estimation via regularized rank regression. *Statistica Sinica*. 2010; 20:167–181.
- [9]. McKean JW. Robust analysis of linear models. *Statistical Science*. 2004; 19:562–570.
- [10]. Terpstra JT, McKean JW. Rank-based analysis of linear models using R. *Journal of Statistical Software*. 2005; 14:1–26.
- [11]. Wang L, Li RZ. Weighted wilcoxon-type smoothly clipped absolute deviation method. *Journal of the American Statistical Association*. 2009; 104:1631–1645. [PubMed: 20657760]
- [12]. Xu GG, Wang SJ, Huang JH. Focused information criterion and model averaging based on weighted composite quantile regression. *Scandinavian Journal of Statistics*. 2014; 41:365–381.
- [13]. Zhang XY, Liang H. Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics*. 2011; 39:174–200.