

## روش‌های تشخیصی در مدل‌های خطی ریح نیمه پارامتری با خطا در اندازه‌گیری

هادی امامی<sup>۱</sup>

گروه آمار، دانشگاه زنجان

تاریخ دریافت: ۱۳۹۵/۶/۱۶ تاریخ پذیرش: ۱۳۹۷/۵/۲۲

**چکیده:** در این مقاله، مباحث تشخیصی در مدل‌های خطی نیمه پارامتری با خطا در اندازه‌گیری با وجود همخطی چندگانه مطالعه می‌شود. در ابتدا برای برطرف نمودن اثرات همخطی چندگانه، با استفاده از روش ماکزیمم درست‌نمایی تصحیح‌شده توانیده برآوردگرهای ریح پیشنهاد می‌شود و سپس براساس رویکرد حذف موردی، آماره‌های تشخیصی برای شناسایی و ارزیابی مشاهدات مؤثر در برآوردگرهای پیشنهادی معرفی می‌شوند. در ادامه نشان داده می‌شود که این آماره‌ها تابعی از مقادیر باقی‌مانده و داده‌های نافذ هستند. علاوه بر رویکرد حذف موردی، رویکرد تأثیر موضعی بر پایه لگاریتم درست‌نمایی تصحیح‌شده توانیده جهت شناسایی و ارزیابی مشاهدات دورافتاده و مؤثر در مدل مورد بحث تعمیم داده شده است. سپس با استفاده از مطالعه شبیه‌سازی کارایی برآوردگرهای پیشنهادی با توجه به معیار میانگین خطای مجانبی ارزیابی می‌شود و در پایان روش‌های تشخیصی مطرح‌شده به یک مجموعه داده واقعی مورد تحلیل و بررسی قرار گرفته است.

**واژه‌های کلیدی:** اسپلاین هموارساز، برآوردگر ریح، خطا در اندازه‌گیری، معیار اعتبارسنجی، مدل‌های خطی نیمه پارامتری، هم خطی چندگانه.

رده‌بندی ریاضی (۲۰۱۰): ۶۲J۰۵، ۶۲J۰۷

### ۱- مقدمه

در سه دهه اخیر مباحث تشخیصی از جمله شناسایی مشاهدات مؤثر و دورافتاده و ارزیابی مناسبیت مدل، توجه زیادی را به خود معطوف نموده است طوری که این مقوله بیش از همه در مدل‌های خطی پارامتری مورد توجه قرار گرفته است. در این زمینه کوک [۱]، کوک و ویزبرگ [۲] و بلسلی

و همکاران [۳] مطالعات وسیعی انجام داده‌اند. در مدل‌های رگرسیون نا پارامتری و نیمه پارامتری مباحث تشخیصی به ندرت مورد مطالعه قرار گرفته‌اند، از میان مطالعات پایه‌ای انجام شده کیم [۴] و کیم و همکاران [۵] باقیمانده‌ها، داده‌های نافذ و آماره کوک را در مدل‌های خطی نا پارامتری و نیمه پارامتری مورد کنکاش قرار داده‌اند. مشابه پژوهش آن‌ها فونگ و همکاران [۶] چنین مباحثی را به مدل‌های آمیخته خطی نیمه پارامتری تعمیم داده‌اند. از طرف دیگر وجود همخطی چندگانه در مدل‌های خطی نیمه پارامتری دور از تصور نیست. شبیه مدل‌های خطی پارامتری وجود همخطی در مدل‌های خطی نیمه پارامتری مسئله‌ساز است و اثرات منفی آن روی اجزای مختلف مدل شناخته شده است. به عنوان مثال حضور همخطی در این مدل‌ها تورم واریانس و یا حذف متغیرهای مهم از مدل را به همراه دارد. برای جزئیات بیشتر درباره اثرات همخطی می‌توان به بلسلی [۷] مراجعه نمود. در صورت وجود همخطی بین متغیرهای توضیحی برآوردگرهای جایگزین دیگری پیشنهاد شده‌اند که عموماً اریب می‌باشند. از میان این برآوردگرها، برآوردگر ریچ هورل و کنارد [۸] نگاه‌های گسترده‌ای را به خود معطوف داشته است. دیدگاه رگرسیون ریچ اخیراً در مدل‌های خطی نیمه پارامتری معمولی (بدون خطا در اندازه‌گیری) توسط بسیاری از پژوهشگران از جمله دوران و آکدنس [۹]، روزبه و آرش [۱۰] و روزبه [۱۱] توسعه یافته است. با گسترش مدل‌های خطی نیمه پارامتری ریچ، امامی [۱۲] و [۱۳] تحلیل مباحث تشخیصی را به ترتیب از دو رویکرد تحلیل حذف موردی و تحلیل تأثیر موضعی مورد تحقیق و کنکاش قرار داده است.

یکی دیگر از مهم‌ترین فرضیه‌های پایه تحلیل‌های آماری این است که مشاهدات مدل درست اندازه‌گیری شده‌اند. نقض چنین فرضی حاکی از نفوذ خطای اندازه‌گیری در داده‌هاست که در صورت وجود موجب کم اعتباری آزمون‌های آماری است. در رابطه با این موضوع به فولر [۱۴] رجوع شود. در مواجهه با نقض چنین فرضیه‌ای دو روش در پژوهش هانفلت و لیانگ [۱۵] اشاره شده است. روش نخست استفاده از درست‌نمایی تصحیح شده ناکامورا [۱۶] است که به خوبی خطاهای اندازه‌گیری را در مدل‌های نرمال، پواسن و گاما تصحیح می‌کند. به نظر می‌رسد استفاده از این روش برای محاسبات کاربردی راحت‌تر است. روش دوم توسط استفانسکی و کارول [۱۷] مطرح شده است که بیشتر براساس یک تابع از برآوردگر ناریب است تا تقریب مناسبی از تابع درست‌نمایی. برای رویارویی با مشکل وجود توأم همخطی چندگانه و خطا در اندازه‌گیری در مدل‌های خطی پارامتری، راسخ [۱۸]، صالح و شالب [۱۹] (با استفاده از روش دوم) و قپانی و همکاران [۲۰] (با استفاده از روش نخست) استفاده از برآوردگرهای ریچ را پیشنهاد داده‌اند. از آنجایی که برآوردگرهای مطرح شده در مقابل مشاهدات مؤثر و دورافتاده حساس هستند راسخ [۲۱] و قپانی و همکاران [۲۲] به ترتیب با روش تحلیل تأثیر موضعی و روش انتقال میانگین

مباحث تشخیصی در مدل‌های خطی پارامتری ریح با خطا در اندازه‌گیری را مورد مطالعه و بررسی قرار داده‌اند.

با توجه به مطالب بالا حضور توأم همخطی و مشاهدات مؤثر در مدل‌های خطی نیمه‌پارامتری با خطا در اندازه‌گیری دور از امکان نیست. از نظر نویسندگان تاکنون پژوهشی در زمینه رویارویی با همخطی و طرح مباحث تشخیصی در مدل‌های خطی نیمه‌پارامتری با خطا در اندازه‌گیری با مشکل همخطی مورد توجه قرار نگرفته است. بنابراین در این مقاله ما برای رویارویی با مشکل همخطی ابتدا برآوردگر ریح را در مدل‌های خطی نیمه‌پارامتری با خطا در اندازه‌گیری و با استفاده از ماکزیمم درست‌نمایی تصحیح‌شده پیشنهاد می‌دهیم و سپس با استفاده از روش حذف موردی و تحلیل تأثیر موضعی مباحث تحلیل تشخیصی جهت شناسایی مشاهدات دورافتاده را توسعه خواهیم داد. لذا در بخش بعدی مدل خطی نیمه‌پارامتری با خطا در اندازه‌گیری تعریف می‌شود. در بخش ۳ برآوردگرهای ریح با استفاده از مدل اسپلاین جزئی به دست می‌آیند و خواص مجانبی بخش پارامتری مدل بررسی می‌شود سپس برآورد ریح بخش پارامتری مدل با استفاده از معیار میانگین مربعات خطا با برآوردگر غیر ریح مقایسه می‌شود. در بخش ۴ شاخص‌های تشخیصی با استفاده از دو رویکرد حذف موردی و تأثیر موضعی بحث شده‌اند. در بخش ۵ داده‌های شبیه‌سازی و داده‌های سفال مصری مورد تحلیل و بررسی قرار می‌گیرد و در بخش ۶ نتایج کلی مطرح می‌شود.

## ۲- پیش‌زمینه و تعاریف

متغیرهای توضیحی در مدل‌های خطی نیمه‌پارامتری ممکن است با یک خطای غیرقابل اغمازی اندازه‌گیری شده باشند. در چنین مواقعی در نظر گرفتن مدل‌های خطی نیمه‌پارامتری با اندازه‌گیری در خطا به شکل زیر می‌تواند مطلوب باشد.

$$\begin{aligned} y &= \mathbf{Z}\beta + g(t) + \epsilon \\ \mathbf{X} &= \mathbf{Z} + \delta \end{aligned} \quad (1)$$

در مدل بالا  $y = (y_1, \dots, y_n)'$  بردار  $n$  تایی مشاهدات و  $\mathbf{Z} = (z_1, \dots, z_n)'$  ماتریس  $n \times p$  از متغیرهای توضیحی پنهان است. بردار  $g(t) = (g(t_1), \dots, g(t_n))'$  مقادیر تابع نامعلوم نا پارامتری است. معمولاً  $t_i$  ها روی یک بازه واحد کراندار هستند و به شکل  $t_1 \leq t_2 \leq \dots \leq t_n$  مرتب می‌شوند.  $\epsilon$  یک بردار با  $n$  مؤلفه از خطاهای مستقل و دارای توزیع نرمال با میانگین صفر و واریانس  $\sigma^2 \mathbf{I}_n$  است که در آن  $\mathbf{I}_n$  ماتریس همانی از مرتبه  $n$  است. ماتریس  $\mathbf{X} = (x_1, \dots, x_n)'$ ، ماتریس مشاهده شده  $\mathbf{Z}$  با خطای اندازه‌گیری  $\delta$  است.  $\delta$  یک ماتریس تصادفی  $n \times p$  مستقل از مجموعه  $(y, \mathbf{X}, t, \epsilon)$  است که سطرهای آن از هم مستقل‌اند طوری

که هر سطر دارای توزیع نرمال با میانگین صفر و کوواریانس ماتریس  $\Lambda$  (بعد سطری و ستونی  $p$ ) است. در واقع  $\delta \sim N(0, \mathbf{I}_n \otimes \Lambda)$  که در آن  $\otimes$  علامت ضرب کرونگر است.

## ۲-۱- روش درست‌نمایی تصحیح‌شده

در این بخش به اختصار روش درست‌نمایی تصحیح‌شده در مدل‌های خطی با خطا در اندازه‌گیری توضیح داده می‌شود.

فرض کنید تابع لگاریتم درست‌نمایی  $l(\beta, \mathbf{Z}, y)$ ، تابع امتیاز  $U(\beta, \mathbf{Z}, y)$ ، تابع فیشر  $I(\beta, \mathbf{Z})$  و تابع اطلاع مشاهده‌شده  $J(\beta, \mathbf{Z}, y) = -\partial U(\beta, \mathbf{Z}, y) / \partial \beta$ ، توابعی نسبت به  $\beta$  با شرط  $\mathbf{Z}$  و  $y$  باشند. اگر  $\beta_0$  مقدار پارامتر درست و  $E^+$  امید نسبت به متغیر تصادفی  $y$  باشد بنابراین

$$E^+ \{U(\beta_0, \mathbf{Z}, y)\} = 0, E^+ \{J(\beta, \mathbf{Z}, y)\} = \mathbf{I}(\beta, \mathbf{Z}). \quad (2)$$

وقتی که  $\mathbf{Z}$  منوط به خطا و  $\mathbf{X}$  مقدار مشاهده‌شده  $\mathbf{Z}$  است بنابراین رابطه  $E^+ \{U(\beta_0, \mathbf{X}, y)\} = 0$  الزاماً برقرار نیست و برآوردی که از رابطه  $U(\beta, \mathbf{X}, y) = 0$  به دست می‌آید به‌طور حتم سازگار نخواهد بود. برای تصحیح این رابطه ناکامورا (۱۹۹۰) تابع درست‌نمایی تصحیح‌شده  $l^*(\beta, \mathbf{X}, y)$  را پیشنهاد داد که در رابطه زیر صدق می‌کند.

$$E^* \{l^*(\beta, \mathbf{X}, y)\} = l(\beta, \mathbf{Z}, y)$$

که در آن  $E^*$  میانگین شرطی نسبت به  $\mathbf{X}$  به شرط  $\mathbf{Z}$  و  $y$  داده‌شده است.

فرض کنید  $U^*(\beta, \mathbf{X}, y) = \partial l^*(\beta, \mathbf{X}, y) / \partial \beta$  و  $J^*(\beta, \mathbf{X}, y) = -\partial U^*(\beta, \mathbf{X}, y) / \partial \beta$  به ترتیب تابع امتیاز و تابع اطلاع تصحیح‌شده باشند. اگر  $E^*$  و  $\beta$  قابل تعویض باشند خواهیم داشت:

$$E^* \{U^*(\beta, \mathbf{X}, y)\} = U(\beta, \mathbf{Z}, y), E^* \{J^*(\beta, \mathbf{X}, y)\} = J(\beta, \mathbf{Z}, y) \quad (3)$$

مقدار  $\hat{\beta}$  که از رابطه  $U^*(\hat{\beta}, \mathbf{X}, y) = 0$  حاصل می‌شود یک برآوردگر تصحیح‌شده درست‌نمایی از  $\beta$  نامیده می‌شود. با توجه به توضیحات ناکامورا [۱۶] تابع درست‌نمایی تصحیح‌شده  $l^*$  و برآورد تصحیح‌شده درست‌نمایی  $\hat{\beta}$  دارای ویژگی‌های خوبی است (هانفلیت و لیانگ [۱۵]). فرض کنید  $E = E^+ E^*$  امید کامل را نشان دهد آنگاه از رابطه (۲) و (۳) خواهیم داشت:

$$E \{U^*(\beta_0, \mathbf{X}, y)\} = E^+ E^* \{U^*(\beta_0, \mathbf{X}, y)\} = E^+ \{U^*(\beta_0, \mathbf{Z}, y)\} = 0 \quad (4)$$

رابطه (۴) نشان می‌دهد که تابع امتیاز تصحیح‌شده ناریب است. می‌توان نشان داد که تحت برخی شرایط نظم  $\hat{\beta}$  دارای خواص سازگاری و توزیع نرمال مجانبی است (ناکامورا [۱۶]).

### ۳- مدل‌های اسپلاین جزئی و برآورد ریح

مشکل همخطی در داده‌ها برآورد پارامترهای رگرسیون را به شدت تحت تأثیر قرار می‌دهد. در مواجهه با همخطی در مدل‌های رگرسیونی به جای استفاده از برآوردگرهای ناریب معمولی استفاده از برآوردگرهای اریب مناسب پیشنهاد می‌شود. از میان این برآوردگرها، برآورد رگرسیون ریح هورل و کنارد [۲۳] حائز اهمیت است. همان‌طور که قبلاً گفته شد امکان بروز همخطی در مدل‌های خطی نیمه پارامتری دور از انتظار نیست. بنابراین در این بخش، با توجه به پژوهش‌های روزبه و آرش [۱۰] و دوران و همکاران [۲۴] روش‌های رگرسیون ریح با استفاده از تابع درست‌نمایی تصحیح‌شده توانیده به مدل‌های خطی نیمه پارامتری با خطا در اندازه‌گیری تعمیم می‌یابد تا برآوردگرهای ریح مؤلفه‌های پارامتری و نا پارامتری حاصل شوند. بدین منظور ابتدا باید تابع ماکزیمم درست‌نمایی توانیده تصحیح‌شده به دست آید. لذا ابتدا: تابع لگاریتم درست‌نمایی توان-های دوم توانیده ریح به شکل زیر تعریف می‌شود:

$$l(\theta, \mathbf{Z}, y) = -\frac{n}{\lambda} \log(\lambda \pi \sigma^2) - \frac{1}{\lambda \sigma^2} \sum_{i=1}^n \{y_i - z_i' \beta - g(t_i)\}^2 - \frac{\lambda}{\lambda \sigma^2} \int g''(t) dt - \frac{k}{\lambda \sigma^2} (\beta' \beta - d) \quad (5)$$

که  $k > 0$  ضریب لاگرانژ و  $\theta = (\beta, g)'$  است. در این مقاله انتخاب پارامتر هموارساز  $\lambda > 0$  با مینیمم کردن معیار اعتبار سنجی تقاطعی  $GCV(\lambda)$  به دست می‌آید. برآورد پارامترها را می‌توان شبیه به مدل اسپلاین جزئی به دست آورد. توضیحات بیشتر در مورد مدل‌های اسپلاین جزئی و کاربرد آماری آن‌ها در گرین و سیلورمن (۱۹۹۴) بیان شده است. فرض کنید مقادیر مرتب‌شده و مجزای  $t_1, \dots, t_n$  با  $s_1, \dots, s_q$  نشان داده شود. رابطه بین  $t_1, \dots, t_n$  و  $s_1, \dots, s_q$  توسط میانگین ماتریس  $n \times q$  با نام  $\mathbf{N}$  با درایه‌های  $\mathbf{N}_{ij} = 1$  اگر  $t_i = s_j$  و در غیر این صورت  $\mathbf{N}_{ij} = 0$ ، توضیح داده می‌شود. از آنجایی که فرض شده است  $t_i$  ها همه یکسان نیستند بنابراین  $q \geq 2$  خواهد بود. اکنون می‌توان  $g$  را به شکل برداری از مقادیر  $g(s_j) = a_j$  نوشت. با توجه به گرین و سیلورمن [۲۵] ماتریسی مانند  $\mathbf{M}$  وجود دارد که تنها به گره‌های  $\{s_j\}$  وابسته است طوری که مینیمم مقادیر  $\int g''(t) dt$  مساوی  $g' \mathbf{M} g$  است. بنابراین تابع لگاریتم درست‌نمایی توان‌های دوم توانیده ریح به شکل

$$l(\theta, \mathbf{Z}, y) = \frac{-n}{2} \log(\nu\pi\sigma^2) - \frac{1}{2\sigma^2} \|y - \mathbf{Z}\beta - \mathbf{N}g\|^2 - \frac{\lambda}{2\sigma^2} g' \mathbf{M}g - \frac{k}{2\sigma^2} (\beta' \beta - d). \quad (6)$$

بازنویسی می‌شود. چون متغیر  $\mathbf{Z}$  قابل مشاهده نیست برای برآورد پارامترها از روش درست‌نمایی تصحیح‌شده استفاده می‌کنیم. برای این کار لگاریتم درست‌نمایی تصحیح‌شده جمع توان‌های دوم تاوانیده ریج به شکل

$$l^*(\theta, \mathbf{X}, y) = \frac{-n}{2} \log \log(\nu\pi\sigma^2) - \frac{1}{2\sigma^2} \{ \|y - \mathbf{X}\beta - \mathbf{N}g\|^2 - n\beta' \Lambda \beta \} - \frac{\lambda}{2\sigma^2} g' \mathbf{M}g - \frac{k}{2\sigma^2} (\beta' \beta - d) \quad (7)$$

خواهد بود، که با مشتق‌گیری (۷) نسبت به  $\beta$  و  $g$ ، برآوردگرهای ریج  $\hat{\beta}_k$  و  $\hat{g}_k$  به ترتیب به صورت

$$\hat{\beta}_k = \{ \mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} + k\mathbf{I}_p - n\Lambda \}^{-1} \mathbf{X}'(\mathbf{I}_n - \mathbf{S})y \quad (8)$$

$$\hat{g}_k = (\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1} \mathbf{N}'(y - \mathbf{X}\hat{\beta}_k) \quad (9)$$

به دست می‌آیند (ضمیمه رابطه (۱) را ببینید)، که در آن  $\mathbf{S}$  یک ماتریس هموارساز و برابر  $\mathbf{S} = \mathbf{N}(\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1} \mathbf{N}'$  است. با قرار دادن  $k = 0$  در رابطه (۸) و (۹) برآوردگرهای معمولی  $\beta$  و  $g$  برابر است با:

$$\hat{\beta} = \{ \mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} - n\Lambda \}^{-1} \mathbf{X}'(\mathbf{I}_n - \mathbf{S})y$$

$$\hat{g} = (\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1} \mathbf{N}'(y - \mathbf{X}\hat{\beta}).$$

در این حالت برآورد ماکزیمم درست‌نمایی تصحیح‌شده برای  $\sigma^2$  برابر است با:

$$\hat{\sigma}^2 = \frac{1}{n} \{ \|y - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{g}\|^2 - n\beta' \Lambda \beta + \lambda \hat{g}' \mathbf{M} \hat{g} \}$$

از رابطه (۸) داریم:

$$\mathbf{X}\hat{\beta}_k = \mathbf{X} \{ \mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} + k\mathbf{I}_p - n\Lambda \}^{-1} \mathbf{X}'(\mathbf{I}_n - \mathbf{S})y = \tilde{\mathbf{H}}_k y$$

و از رابطه (۹) خواهیم داشت:

$$\mathbf{N}\hat{\mathbf{g}}_k = \mathbf{N}(\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1}\mathbf{N}'(y - \mathbf{X}\hat{\beta}_k) = \mathbf{S}(\mathbf{I}_n - \tilde{\mathbf{H}})y = \mathbf{H}_k^* y.$$

از روابط (۸) و (۹) مقادیر تقریبی  $\hat{y}$  را می‌توان به شکل

$$\hat{y} = \mathbf{X}\hat{\beta}_k + \mathbf{N}\mathbf{g}_k = \mathbf{H}y$$

نوشت. که در آن

$$\mathbf{H} = \mathbf{S} + (\mathbf{I}_n - \mathbf{S})\mathbf{X}\{\mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} + k\mathbf{I}_p - n\Lambda\}^{-1}\mathbf{X}'(\mathbf{I}_n - \mathbf{S}) = \tilde{\mathbf{H}}_k + \mathbf{H}_k^*,$$

ماتریس مقادیر نافذ می‌باشد. بردار باقیمانده‌ها از رابطه

$$e = (\mathbf{I}_n - \mathbf{H})y,$$

به دست خواهد. یکی از نکات کلیدی در رگرسیون رنج انتخاب پارامتر  $k$  است. روش‌های متفاوتی برای انتخاب این پارامتر ارائه شده است از جمله روش تکراری هورل و کنارد [۲۳]، معیار آماره  $C_p$  و معیار  $GCV$  روش‌های عمومی هستند.

### ۳-۱- بررسی خواص جانبی برآورد رنج پارامتر مدل

در این بخش ابتدا توزیع جانبی برآورد  $\hat{\beta}_k$  در قالب یک قضیه به اختصار بررسی می‌شود. سپس با توجه به نتایج این قضیه دو برآوردگر  $\hat{\beta}_k$  و  $\hat{\beta}$  طبق معیار میانگین مربعات خطای جانبی باهم مقایسه می‌شوند. فرض می‌شود که مقادیر حدی،  $\lim_{n \rightarrow \infty} n^{-1}\mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z}$  و  $\lim_{n \rightarrow \infty} n^{-1}\mathbf{D}$  وجود دارند، که  $\mathbf{D} = (\mathbf{Z}\beta + \mathbf{N}g)'(\mathbf{I}_n - \mathbf{S})'(\mathbf{Z}\beta + \mathbf{N}g)$  است.

قضیه ۱: تحت فرضیات قبل،  $\hat{\beta}_k$  دارای توزیع نرمال جانبی با امید ریاضی و واریانس جانبی

$$\text{avar}(\hat{\beta}_k) = \mathbf{C}_k^{-1}\Phi\mathbf{C}_k^{-1}E(\hat{\beta}_k) = \mathbf{C}_k\mathbf{C}_k^{-1}\beta$$

است که در آن  $\mathbf{C}_k = \mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z} + k\mathbf{I}_p$ ،  $\mathbf{C}_0 = \mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z}$  و  $\Phi = \mathbf{D} + \mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z}\sigma^2$  هستند.

اثبات: ضمیمه را ببینید.

از قضیه ۱ می‌توان میانگین مربعات خطای جانبی برآوردگر  $\hat{\beta}_k$  را به شکل

$$\begin{aligned} amse(\hat{\beta}_k) &= avar(\hat{\beta}_k) + bias(\hat{\beta}_k)bias(\hat{\beta}_k)' \\ &= C_k^{-1} \Phi C_k^{-1} + k' C_k^{-1} \beta \beta' C_k^{-1} \end{aligned}$$

به دست آورد. در ابتدا برای مقایسه  $amse$  دو برآوردگر  $\hat{\beta}_k$  و  $\hat{\beta}$  برای  $k > 0$  رابطه

$$\begin{aligned} \Pi &= avar(\hat{\beta}) - avar(\hat{\beta}_k) \\ &= C_0^{-1} \Phi C_0^{-1} - C_k^{-1} \Phi C_k^{-1} > 0 \end{aligned}$$

برقرار است. ماتریس  $\Pi$  معین مثبت است (طبق تعریف ۱ ضمیمه) چون  $C_k > C_0$  بنابراین

$C_0^{-1} > C_k^{-1}$  و در نتیجه  $C_0^{-1} \Phi C_0^{-1} > C_k^{-1} \Phi C_k^{-1}$ . اکنون با توجه به  $\Pi$  اختلاف بین

ماتریس‌های میانگین مربعات خطای مجانبی  $\hat{\beta}_k$  و  $\hat{\beta}$  برابر با

$$\Delta_k = amse(\hat{\beta}) - amse(\hat{\beta}_k) = \Pi - k' C_k^{-1} \beta \beta' C_k^{-1}$$

است. با توجه به لم ۲ ضمیمه  $\Delta_k > 0$  اگر و تنها اگر

$$d = k - \frac{1}{\sqrt{\beta' C_k^{-1} \Pi^{-1} C_k^{-1} \beta}} \leq 0. \quad (10)$$

### ۲-۳- برآورد $k$ و $\lambda$

از رابطه (۱۰) مشخص هست که مقدار  $k$  مناسب را به راحتی نمی توان یافت. یک روش برای

پیدا کردن  $k$  رسم  $\Delta_k$  در برابر یک بازه پیوسته از  $k$  است. مقدار  $k$  متناظر به ماکزیمم  $\Delta_k$

می تواند به عنوان  $k$  بهینه در نظر گرفته شود [۱۰]. برای پیدا کردن پارامتر ریح و پارامتر

هموارساز مناسب از معیار اعتبار سنجی متقابل تعمیم یافته ( $GCV$ ) استفاده می شود. اخیراً

روزبه [۲۶] و آرشی و روزبه [۲۷] این روش را برای برآورد هم‌زمان پارامترهای  $k$  و  $\lambda$

به کار برده‌اند. فرض کنید با حذف  $\lambda$  آمین سطر از داده‌ها مقدارهای  $\hat{\beta}_{k(i)}$  و  $\hat{g}_{k(i)}$  برآوردهای ریح

$\beta$  و  $g$  باشند. آنگاه برآورد مناسب  $k$  و  $\lambda$  مقدارهایی هستند که به ازای آن‌ها مؤلفه

$\mathbf{W}\hat{\beta}_{k(i)} + \mathbf{N}\hat{g}_{k(i)}$  یعنی  $[\mathbf{W}\hat{\beta}_{k(i)} + \mathbf{N}\hat{g}_{k(i)}]_i$  پیش‌بینی خوبی برای  $y_i$  است، به عبارتی

مقدارهای مناسب برای  $k$  و  $\lambda$  مقدارهایی هستند که تابع زیر را مینیمم کنند:

$$GCV(k, \lambda) = n^{-1} \sum_{i=1}^n \left( y_i - [\mathbf{W}\hat{\beta}_{k(i)} + \mathbf{N}\hat{g}_{k(i)}]_i \right)^2$$



همچنین  $\lambda$  را می‌توان به‌جز روش هم‌زمان برآورد از طریق دومرحله‌ای نیز به دست آورد برای این کار با قرار دادن  $k$  معلوم مناسب در تابع بالا معیار اعتبارسنجی تعمیم‌یافته را می‌توان تنها برحسب  $\lambda$  نوشت و  $\lambda$  را با مینیمم کردن تابع حاصل به دست آورد [۱۰].

#### ۴- شاخص‌های مؤثر در برآورد لگاریتم درست‌نمایی تصحیح‌شده جمع توان‌های دوم تاوانیده رنج

در مباحث ویژه برای شناسایی مشاهدات مؤثر که به شکل جدی نتایج تحلیل آماری را تحت تأثیر قرار می‌دهند دو رویکرد اصلی وجود دارد. رویکرد نخست براساس روش حذف موردی است که در آن اثر حذف یک مشاهده در برآوردها به‌طور مستقیم توسط برخی از معیارهای معتبر نظیر آماره کوک [۱] ارزیابی می‌شود. رویکرد دوم تأثیر موضعی است که توسط کوک [۲۸] مطرح شده است. در این رویکرد با وارد کردن اغتشاش‌های جزئی به اجزای مختلف مدل به بررسی ثبات برآوردهای مدل پس از اغتشاش می‌پردازد. در ادامه ابتدا رویکرد حذف موردی و سپس تأثیر موضعی بحث می‌شود.

#### ۴-۱- حذف موردی

فرض کنید وقتی مشاهده  $i$  از مجموعه داده حذف شده است مقادیر  $\hat{\beta}_{k(i)}$  و  $\hat{g}_{k(i)}$  برآوردهای لگاریتم درست‌نمایی تصحیح‌شده تاوانیده رنج  $\beta$  و  $g$  هستند.

قضیه ۲: برآورد پارامترهای  $\theta$  بعد از حذف مشاهده  $i$  ام به شکل زیر به دست می‌آیند:

$$\hat{\beta}_{k(i)} = \hat{\beta}_k - \frac{\{X'(I_n - S)X + kI_p - n\Lambda\}^{-1} X'(I_n - S)e_i \xi_i}{1 - h_{ii}}$$

$$\hat{g}_{k(i)} = \hat{g}_k - \frac{(N'N + \lambda M)^{-1} N'[I_n - X\{X'(I_n - S)X + kI_p - n\Lambda\}^{-1} X'(I_n - S)e_i \xi_i]}{1 - h_{ii}}$$

که  $\xi_i$  برداری با بعد  $n$  است که  $i$ -امین درایه آن ۱ و سایر درایه‌های آن مقدار صفر می‌گیرند.  $e_i$  نیز  $i$ -امین مؤلفه از بردار باقیمانده تاوانیده رنج  $e$  است.

اثبات: فرض کنید  $y^* = (y_1^*, \dots, y_n^*)$  و  $y_j^* = \begin{cases} y_i^*, & j = i \\ y_j, & j \neq i \end{cases}$  طوری که

را برای هر  $\beta$  و تابع هموار  $g$  مینیمم کنند:  $y_i^* = x_i' \hat{\beta}_{k(i)} + \hat{g}_{k(i)}(t_i)$  است. با توجه به تعریف  $\hat{\beta}_{k(i)}$  و  $\hat{g}_{k(i)}$  مقادیری هستند که تابع زیر

$$\begin{aligned} & \|y^* - \mathbf{X}\beta - \mathbf{N}g\|^2 + \lambda \int g''(t) dt + k(\beta'\beta - d) - n\beta'\Lambda\beta \\ & \geq \sum_{j \neq i} (y_j - x_j'\beta - g(t_j))' + \lambda \int g''(t) dt + k(\beta'\beta - d) - n\beta'\Lambda\beta \end{aligned} \quad (11)$$

$$\begin{aligned} & \geq \|y^* - \mathbf{X}\hat{\beta}_{k(i)} - \mathbf{N}\hat{g}_{k(i)}\|^2 + \lambda \int \hat{g}_{k(i)}''(t) dt + k(\hat{\beta}_{k(i)}'\hat{\beta}_{k(i)} - d) - n\hat{\beta}_{k(i)}'\Lambda\hat{\beta}_{k(i)} \\ & \beta_{k(i)} = \hat{\beta}_{k(i)} \text{ با جایگزینی تابع با جایگزینی } \hat{g}_{k(i)} = g_{k(i)} \text{ در رابطه} \end{aligned}$$

$$\|y^* - \mathbf{X}\hat{\beta}_{k(i)} - \mathbf{N}g_{k(i)}\|^2 + \lambda g_{k(i)}' \mathbf{M} g_{k(i)} + k(\beta_{k(i)}'\beta_{k(i)} - d) - n\beta_{k(i)}'\Lambda\beta_{k(i)}$$

حاصل می‌شود. بنابراین با مشتق‌گیری از رابطه بالا نسبت به  $\beta_{k(i)}$  و  $g_{k(i)}$  برآوردها برابر

$$\hat{\beta}_{k(i)} = \{\mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} + k\mathbf{I}_p - n\Lambda\}^{-1} \mathbf{X}'(\mathbf{I}_n - \mathbf{S})y^*$$

$$\hat{g}_{k(i)} = (\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1} \mathbf{N}' \left( y^* - \mathbf{X}\hat{\beta}_{k(i)} \right)$$

به دست می‌آیند (کافی است در رابطه (۱) ضمیمه به جای  $y$  مقدار  $y^*$  و به جای  $\beta$  و  $g$  مقادیر  $\beta_{k(i)}$  و  $g_{k(i)}$  قرار گیرند). با قرار دادن  $y^* = y - (y - y^*)$  در روابط بالا و ذکر این نکته که  $y - y^* = \xi_i(y_i - y_i^*)$  رابطه

$$\hat{\beta}_{k(i)} = \hat{\beta}_k - \{\mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} + k\mathbf{I}_p - n\Lambda\}^{-1} \mathbf{X}'(\mathbf{I}_n - \mathbf{S})\xi_i(y_i - y_i^*)$$

حاصل می‌شود. در ادامه برای  $\hat{g}_{k(i)}$  رابطه زیر به دست می‌آید:

$$\hat{g}_{k(i)} = \hat{g}_k - (\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1} \mathbf{N}' [\mathbf{I}_n - \mathbf{X}\{\mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} + k\mathbf{I}_p - n\Lambda\}^{-1} \mathbf{X}'(\mathbf{I}_n - \mathbf{S})] \xi_i(y_i - y_i^*)$$

از طرفی داریم:

$$\begin{aligned} y_i - y_i^* &= \xi_i' (y - \mathbf{X}\hat{\beta}_{k(i)} - \mathbf{N}\hat{g}_{k(i)}) = \xi_i' (y - \mathbf{S}y^* - (\mathbf{I}_n - \mathbf{S})\mathbf{X}\hat{\beta}_{k(i)}) \\ &= \xi_i' (y - \mathbf{H}y^*) = \xi_i' [y - \mathbf{H}y + \mathbf{H}(y - y^*)] \\ &= \xi_i' (y - \mathbf{H}y) + \xi_i' \mathbf{H}\xi_i (y_i - y_i^*) = e_i + h_{ii} (y_i - y_i^*). \end{aligned}$$

بنابراین  $y_i - y_i^* = \frac{e_i}{1 - h_{ii}}$  به دست می‌آید و اثبات کامل می‌شود.

#### ۴-۱-۱ تأثیر روی $\hat{\beta}_k$

حداقل دو نوع آماره کوک به شکل زیر می‌تواند برای برآورد لگاریتم درست‌نمایی تصحیح‌شده کمترین توان‌های دوم تاوانیده ریح می‌توان مطرح شود.

$$D_{\beta i} = \frac{(\hat{\beta}_k - \hat{\beta}_{k(i)})' \mathbf{X}' \mathbf{X} (\hat{\beta}_k - \hat{\beta}_{k(i)})}{\sigma^2 \text{tr}(\tilde{\mathbf{H}}_k)}$$

$$D_{\beta i}^* = \frac{(\hat{\beta}_k - \hat{\beta}_{k(i)})' \{ \mathbf{X}' (\mathbf{I}_n - \mathbf{S}) \mathbf{X} + k \mathbf{I}_p - n \Lambda \} (\mathbf{X}' (\mathbf{I}_n - \mathbf{S}) \mathbf{X})^{-1} \{ \mathbf{X}' (\mathbf{I}_n - \mathbf{S}) \mathbf{X} + k \mathbf{I}_p - n \Lambda \} (\hat{\beta}_k - \hat{\beta}_{k(i)})}{\sigma^2 \text{tr}(\tilde{\mathbf{H}}_k)}$$

که در آن  $\tilde{\mathbf{H}}_k = \mathbf{X} \{ \mathbf{X}' (\mathbf{I}_n - \mathbf{S}) \mathbf{X} + k \mathbf{I}_p - n \Lambda \}^{-1} \mathbf{X}' (\mathbf{I}_n - \mathbf{S})$  با درایه  $\tilde{h}_{ij}$  است.  $D_{\beta i}$  را می‌توان با استفاده از روابط قضیه ۱ به صورت تابعی از باقیمانده‌ها و داده‌های نافذ نیز نوشت (ضمیمه را ببینید)

$$D_{\beta i} = \frac{\left[ \sum_{j=1}^n \tilde{h}_{ij}^{\gamma} \right] e_i^{\gamma}}{\sigma^2 \text{tr}(\tilde{\mathbf{H}}_k) (1 - h_{ii})^{\gamma}}$$

#### ۴-۱-۲ تأثیر روی $\hat{g}_k$

برای سنجش تأثیر  $i$ -آمین مشاهده روی  $\hat{g}_k$  فاصله کوک زیر تعریف می‌شود:

$$D_{gi} = \frac{(\hat{g}_{k(i)} - \hat{g}_k)' \mathbf{N}' \mathbf{N} (\hat{g}_{k(i)} - \hat{g}_k)}{\sigma^2 \text{tr}(\mathbf{H}_k^*) (1 - h_{ii})^{\gamma}}$$

شبهه  $D_{\beta i}$ ،  $D_{gi}$  را می‌توان به صورت باقیمانده‌ها و مقادیر نافذ به شکل زیر نوشت:

$$D_{gi} = \frac{\left[ \sum_{j=1}^n h_{ij}^{*\gamma} \right] e_i^{\gamma}}{\sigma^2 \text{tr}(\mathbf{H}_k^*) (1 - h_{ii})^{\gamma}}$$

که  $h_{ij}^*$  درایه ماتریس  $\mathbf{H}_k^* = \mathbf{S} (\mathbf{I}_n - \tilde{\mathbf{H}}_k)$  است.

#### ۴-۱-۳ تأثیر روی $\hat{y}_k$

معیار سنجش برای میزان تأثیر  $i$ -آمین مشاهده روی بردار مقادیر برازش به شکل

$$D_{yi} = \frac{(\hat{y}_k - \hat{y}_{k(i)})' (\hat{y}_k - \hat{y}_{k(i)})}{\sigma^2 tr(\mathbf{H})}$$

تعریف می‌شود.  $D_{yi}$  را می‌توان به صورت تقریبی تابعی از باقیمانده‌ها و داده‌های نافذ نیز نوشت:

$$D_{yi} \approx \frac{h_{ii} e_i^2}{\sigma^2 tr(\mathbf{H})(1 - h_{ii})^2}$$

#### ۴-۱-۴ روابط بین فاصله کوک‌ها

برای بررسی رابطه نظری بین  $D_{yi}$ ،  $D_{gi}$ ،  $D_{\beta i}$  با توجه به  $\hat{y} = \mathbf{X}\hat{\beta}_k + \mathbf{N}\hat{g}_k$  و محاسبات ۴-۱-۱ و ۴-۱-۲ خواهیم داشت.

$$\begin{aligned} \sigma^2 tr(\mathbf{H}) D_{yi} &= (\hat{y}_k - \hat{y}_{k(i)})' (\hat{y}_k - \hat{y}_{k(i)}) \\ &= \{ \mathbf{X}(\hat{\beta}_k - \hat{\beta}_{k(i)}) + \mathbf{N}(\hat{g}_{k(i)} - \hat{g}_k) \}' \{ \mathbf{X}(\hat{\beta}_k - \hat{\beta}_{k(i)}) + \mathbf{N}(\hat{g}_{k(i)} - \hat{g}_k) \} \\ &= (\hat{\beta}_k - \hat{\beta}_{k(i)})' \mathbf{X}' \mathbf{X} (\hat{\beta}_k - \hat{\beta}_{k(i)}) + (\hat{g}_{k(i)} - \hat{g}_k)' \mathbf{N}' \mathbf{N} (\hat{g}_{k(i)} - \hat{g}_k) \\ &\quad + 2(\hat{\beta}_k - \hat{\beta}_{k(i)})' \mathbf{X}' \mathbf{N} (\hat{g}_{k(i)} - \hat{g}_k) \\ &= \sigma^2 tr(\tilde{\mathbf{H}}_k) D_{\beta i} + \sigma^2 tr(\mathbf{H}_k^*) D_{gi} + 2(\hat{\beta}_k - \hat{\beta}_{k(i)})' \mathbf{X}' \mathbf{N} (\hat{g}_{k(i)} - \hat{g}_k) \end{aligned}$$

#### ۴-۲-۲ رویکرد تأثیر موضعی

رویکرد تأثیر موضعی ابتدا توسط کوک [۲۸] معرفی شد و سپس سایر افراد از جمله توماس و کوک [۲۹] و کوان و فونگ [۳۰] آن را مورد بررسی و کنکاش بیشتری قرار دادند. در این بخش ابتدا روابط پایه‌ای تأثیر موضعی مرور می‌شود و سپس این رویکرد به مدل‌های خطی ریح نیمه-پارامتری با خطا در اندازه‌گیری تعمیم داده می‌شود. شایان ذکر است که این تعمیم بر پایه لگاریتم درست‌نمایی تصحیح‌شده تاوانیده است. مطابق روش کوک [۲۸]، ابتدا انحراف  $\omega$  را که ماهیت آن در ادامه بیشتر مشخص می‌شود بررسی می‌کنیم. ابتدا فرض می‌شود که مقدار  $\omega$  برای داده‌ها و مدل اصلی برابر صفر است.  $\hat{\theta}_\omega$  برآورد  $\theta$  تحت انحراف  $\omega$  است. لگاریتم درست‌نمایی تصحیح‌شده تاوانیده تحت انحراف داده‌شده با  $l^*(\theta, \omega)$  نشان داده می‌شود طوری که  $l^*(\theta, 0)$  برابر با  $l^*(\theta)$  در رابطه (۷) است و برای هر  $\omega$ ،  $l^*(\theta, \omega)$  در  $\theta = \hat{\theta}_\omega$  مینیمم می‌شود. با

هر بردار واحد سویه  $d$  در فضای  $\omega$ ، انحنا نرمال نگاشت از  $\omega$  به  $l^*(\hat{\theta}_\omega)$  در  $\omega = \circ$  و در جهت  $d$  برابر  $d^T \ddot{\mathbf{F}} d$  تعریف می‌شود که در آن  $\ddot{\mathbf{F}}$  برابر

$$\ddot{\mathbf{F}} = \frac{\partial^2 l^*(\hat{\theta}_\omega)}{\partial \omega \partial \omega'} \quad (۱۲)$$

است. در نقطه  $\omega = \circ$ ،  $\ddot{\mathbf{F}}$  ماتریس تأثیر برای  $\theta$  نامیده می‌شود. از آنجایی که  $\frac{\partial l^*(\hat{\theta}_\omega, \omega)}{\partial \theta} = \circ$  است محاسبات مستقیم با استفاده از قاعده زنجیره‌ای رابطه

$$\ddot{\mathbf{F}} = \Delta' \ddot{\mathbf{L}}^{-1} \Delta \quad (۱۳)$$

را نتیجه می‌دهد. که در آن  $\Delta = \frac{\partial^2 l^*(\theta, \omega)}{\partial \theta \partial \omega'}$  در نقطه  $\theta = \hat{\theta}$  و  $\omega = \circ$  است و عبارت میانی

برای  $\ddot{\mathbf{F}}$  در  $\theta = \hat{\theta}$   $\ddot{\mathbf{L}} = \frac{\partial^2 l^*(\theta)}{\partial \theta \partial \theta'}$  در نظر گرفته می‌شود. کوک [۲۸] نشان داد که ساختار ویژه  $\ddot{\mathbf{F}}$  برای اهداف تشخیصی می‌تواند مفید باشد. عناصر قطری  $\ddot{\mathbf{F}}$ ، حساسیت موضعی به لگاریتم درست‌نمایی را توسط انحرافات مؤلفه-محور اندازه‌گیری می‌کند. برای مثال، اگر پاسخ انحراف داده شده  $\gamma + \omega$  است لذا مقدار نسبتاً بزرگی از اولین عنصر قطری ماتریس تأثیر، اثر نسبتاً بزرگی از اولین مشاهده روی  $\theta$  را نشان می‌دهد. هرچند تحلیل‌های تأثیر معمولاً برای اندازه‌های نمونه کوچک تعریف می‌شوند لکن انتظار می‌رود که اندازه‌های تأثیر روی  $g$  هنگامی که  $n$  بزرگ است، ماتریس تأثیر  $\ddot{\mathbf{F}}$  را دربر داشته باشد. لذا توصیه می‌شود که تأثیر جزئی را با تمرکز روی جزئی از  $\theta$  یعنی  $\beta$  در نظر بگیریم. بدین منظور،  $\hat{g}$  را به‌عنوان تابعی از  $\hat{\beta}$  در نظر می‌گیریم. و ماتریس تأثیر جزئی را از انحنا نرمال نگاشت شده از  $\omega$  به  $l^*(\hat{\beta}_\omega, \hat{g}(\hat{\beta}_\omega))$  تعریف می‌کنیم. ماتریس تأثیر جزئی برای  $\beta$  از فرمول

$$\ddot{\mathbf{F}}(\beta) = \Delta' \left( \ddot{\mathbf{L}}^{-1} - \begin{pmatrix} \cdot & \cdot \\ \cdot & \mathbf{L}_g^{-1} \end{pmatrix} \right) \Delta, \quad (۱۴)$$

به دست می‌آید که  $\mathbf{L}_g$  زیر ماتریسی از  $\ddot{\mathbf{L}}$  می‌باشد که متناظر با  $g$  است. تأثیر جزئی برای بخش دیگر (یا معادل آن) از همان طریق ارائه می‌شود. برای جزئیات بیشتر به کوک ([۲۸]، ص ۱۴) مراجعه نمایید.

#### ۴-۲-۱ انحراف وزن‌های نمونه

نخست داده‌ها را با تغییر وزنی هر نمونه در معیارهای حداقل مربعات انحراف می‌دهیم. این کار معادل انحراف واریانس  $\epsilon_i$  در مدل می‌باشد. با وزن‌های  $1 + \omega_i$  داده شده به  $t$ -آمین مشاهده، رابطه زیر را داریم:

$$l^*(\theta, \omega) = \frac{-1}{2\sigma^2} \sum_{i=1}^n (1 + \omega_i) (y_i - x_i' \beta - g(t_i))^2 + \frac{n}{2\sigma^2} \beta' \Lambda \beta - \frac{\lambda}{2\sigma^2} \int g''(t) dt - \frac{k}{2\sigma^2} (\beta' \beta - d) - \frac{n}{2} \log 2\pi\sigma^2$$

با  $\omega = (\omega_1, \dots, \omega_n)' \in \mathbf{R}^n$  محاسبه مستقیم رابطه‌های زیر

$$\Delta = \hat{\sigma}^2 (\mathbf{X}, \mathbf{N})' \hat{\mathbf{E}} \quad \text{و} \quad \ddot{\mathbf{L}} = \begin{pmatrix} \mathbf{X}'\mathbf{X} + k\mathbf{I}_p - n\Lambda & \mathbf{X}'\mathbf{N} \\ \mathbf{N}'\mathbf{X} & \mathbf{N}'\mathbf{N} + \lambda\mathbf{M} \end{pmatrix}$$

به دست می‌آیند. که  $\hat{\mathbf{E}} = \text{diag}(y - \mathbf{X}\hat{\beta}_k - \mathbf{N}\hat{g}_k)$ ، ماتریس قطری است. بنابراین ماتریس تأثیر

$$\ddot{\mathbf{F}}_{\omega} = \hat{\mathbf{E}}\mathbf{H}\hat{\mathbf{E}} \quad (15)$$

به دست می‌آید. عناصر قطری  $\ddot{\mathbf{F}}_{\omega}$  کاملاً به منحنی‌های تأثیر نمونه بر پایه حذف نمونه مرتبط هستند (برای مثال کوک و ویزبرگ [۲] را ببینید). همچنین ماتریس‌های تأثیر جزئی برای مؤلفه‌های پارامتری به شکل زیر به دست می‌آیند:

$$\ddot{\mathbf{F}}_{\omega}(\beta) = \hat{\mathbf{E}}(\mathbf{H} - \mathbf{S})\hat{\mathbf{E}} \quad \text{و} \quad \ddot{\mathbf{F}}_{\omega}(g) = \hat{\mathbf{E}}(\mathbf{H} - \mathbf{X}(\mathbf{X}'\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1}\mathbf{X}')\hat{\mathbf{E}} \quad (16)$$

محاسبه ماتریس تأثیر (۱۵) آسان است، ولی همانند فاصله‌ی کوک طبق حذف موردی، وابسته به باقیمانده‌ها و مقادیر نافذ است. یک تفاوت آشکار رابطه (۱۵) بافاصله کوک این است که از باقیمانده‌های غیراستاندارد استفاده می‌کند. این کار می‌تواند به انتخاب انحرافات نسبت داده شود. اگر انحراف  $\omega$  خود به‌طور صحیح مقیاس بندی شود، طوری که وزن  $\sigma_{\omega}$  باشد، باقیمانده‌ها نیز در رابطه (۱۴) توسط  $\sigma$  استانداردسازی می‌شوند.

#### ۴-۲-۲ انحراف متغیر پاسخ

اکنون با جایگزینی  $y$  با  $y + \omega$  وقتی که  $\omega \in \mathbf{R}^n$  انحراف متغیر پاسخ را بررسی می‌کنیم، در این مورد خواهیم داشت:

$$l^*(\theta, \omega) = \frac{-1}{2\sigma^2} \|y + \omega - \mathbf{X}\beta - \mathbf{N}g\|^2 + \frac{n}{2\sigma^2} \beta' \Lambda \beta - \frac{\lambda}{2\sigma^2} g' \mathbf{M}g - \frac{k}{2\sigma^2} (\beta' \beta - d) - \frac{n}{2} \log 2\pi\sigma^2.$$

آنگاه ماتریس تأثیر به شکل ساده

$$\ddot{\mathbf{F}}_y = \mathbf{H} \quad (17)$$

به دست خواهد آمد و ماتریس‌های تأثیر جزئی برابر  $\ddot{\mathbf{F}}_y(\beta) = \mathbf{H} - \mathbf{S}$  است که در آن  $\ddot{\mathbf{F}}_y(g) = \mathbf{H} - \mathbf{X}(\mathbf{X}'\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1} \mathbf{X}'$  تأثیر موضعی تحت انحراف پاسخ متناظر به داده‌های نافذ است تعجب‌آور نیست. اینجا تأثیر جزئی ابزار درک مقادیر نافذ شرطی و حاشیه‌ای در مدل‌های خطی نیمه پارامتری را فراهم می‌کند. می‌توان  $\mathbf{H} - \mathbf{S}$  را به‌عنوان ماتریس مقادیر نافذ جهت برآورد  $\beta$  بعد از تطبیق برآورد  $g$  و  $\mathbf{H} - \mathbf{X}(\mathbf{X}'\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1} \mathbf{X}'$  را به‌عنوان ماتریس مقادیر نافذ برای برآورد  $g$  بعد از تطبیق برآورد  $\beta$  در نظر بگیریم. به‌طور مشابه می‌توان  $\mathbf{X}(\mathbf{X}'\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1} \mathbf{X}'$  و  $\mathbf{S}$  را به‌عنوان ماتریس‌های مقادیر نافذ حاشیه‌ای برای مؤلفه‌های پارامتری و نا پارامتری در نظر گرفت. به‌طور خلاصه داریم:

$$\text{مقادیر نافذ شرطی} + \text{مقادیر نافذ حاشیه‌ای} = \text{مقادیر نافذ توأم}$$

#### ۴-۲-۳ انحراف متغیرهای کمکی

انحراف متغیرهای کمکی تأثیر بسیار پیچیده‌ای روی برآورد دارد. واضح است که خطاهای اندازه‌گیری روی متغیرهای کمکی می‌تواند منجر به اریبی نسبتاً جدی در برآورد ضرایب رگرسیون خطی شود [۱۴]. در مدل‌های خطی نیمه پارامتری پژوهشگران اندکی روی مبحث اریبی کار کردند. تحلیل تأثیر موضعی تحت انحراف متغیرهای کمکی ممکن است یک دید متناوب به مدل‌های خطا در اندازه‌گیری فراهم نماید.

انحراف  $\mathbf{X}$  به  $\mathbf{X}_\omega = \mathbf{X} + \omega l_i'$  را که  $\omega \in R^n$  و  $l_i \in R^p$  یک بردار یکه با  $i$ -امین عنصر برابر با یک است، را در نظر بگیرید، طوری که تنها  $i$ -امین متغیر کمکی انحراف داده می‌شود. در این مورد رابطه

$$l^*(\theta, \omega) = \frac{-1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\beta - \mathbf{N}g - \omega l_i'\|^2 + \frac{n}{2\sigma^2} \beta' \Lambda \beta - \frac{\lambda}{2\sigma^2} g' \mathbf{M}g - \frac{k}{2\sigma^2} (\beta' \beta - d) - \frac{n}{2} \log 2\pi\sigma^2.$$

برقرار است. با توجه به محاسبات در ضمیمه، ماتریس تأثیر رابطه

$$\ddot{\mathbf{F}}_{x,i} = (l'_i \hat{\beta}_k)' \mathbf{S} + \mathbf{A}' (\mathbf{X}' (\mathbf{I} - \mathbf{S}) \mathbf{X} - n\mathbf{\Lambda} + k\mathbf{I}_p)^{-1} \mathbf{A} \quad (18)$$

را نتیجه می‌دهد، که  $\mathbf{A} = l'_i e' - (l'_i \hat{\beta}_k)' \mathbf{X}' (\mathbf{I} - \mathbf{S})$  است. برای ماتریس‌های تأثیر جزئی  $\beta$  و  $g$  به ترتیب رابطه‌های:

$$\begin{aligned} \ddot{\mathbf{F}}_{x,i}(\beta) &= \ddot{\mathbf{F}}_{x,i} - (l'_i \hat{\beta}_k)' \mathbf{S} \\ \ddot{\mathbf{F}}_{x,i}(g) &= \ddot{\mathbf{F}}_{x,i} - \left( (l'_i \hat{\beta}_k)' \mathbf{X} - e'_i \right) (\mathbf{X}' \mathbf{X})^{-1} \left( (l'_i \hat{\beta}_k)' \mathbf{X} - e'_i \right)' \end{aligned}$$

حاصل می‌شود. توجه شود که رابطه (۱۸) به دو بخش تقسیم شده است. اولین بخش رابطه (۱۷)،  $(l'_i \hat{\beta}_k)' \mathbf{S}$  بخشی از برآورد نا پارامتری است. باقیمانده فرمول (۱۸) همان شکلی را داراست که ماتریس تأثیر برای مدل‌های خطی دارد. برای تفسیر و بحث مفصل راجع این بخش به کوک [۲۸] بخش ۵ رجوع کنید.

### ۵-۱- شبیه‌سازی

در این بخش برای ارزیابی تئوری‌های بیان شده در بخش ۴ یک مجموعه داده‌های شبیه‌سازی شده از مدل (۱) در نظر گرفته شده است. در واقع مدل خطی نیمه پارامتری با خطا در اندازه‌گیری موردنظر برابر

$$\begin{aligned} y &= \mathbf{Z}\beta + \sin 2\pi t + \varepsilon \\ \mathbf{X} &= \mathbf{Z} + \delta \end{aligned}$$

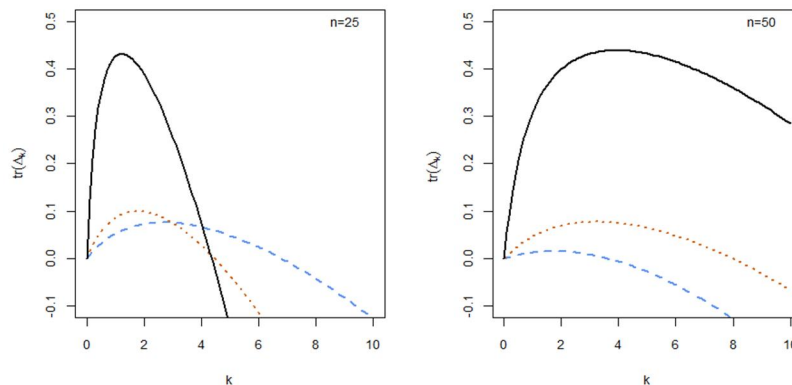
است که در آن درایه‌های مربوط به متغیرهای  $\mathbf{Z}$  برای داشتن همخطی چندگانه در ستون‌ها به شکل زیر تولید می‌شوند:

$$z_{ij} = (1 - \rho^2)^{\frac{1}{2}} u_{ij} + \rho u_{ij} \quad j = 1, 2, 3 \quad i = 1, \dots, n$$

که در آن  $u_{ij}$  از توزیع نرمال استاندارد تولید می‌شوند و ضریب  $\rho$  اختصاص داده شده میزان همبستگی بین ستون‌های ماتریس  $\mathbf{Z}$  را نشان می‌دهد. در هر تکرار بردار  $\beta$  برابر بردار ویژه متناظر به بزرگ‌ترین مقدار ویژه ماتریس  $\mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z}$  انتخاب شده است و بردار  $t$  از توزیع یکنواخت صفر و یک تصادفی به دست می‌آید. عناصر  $\varepsilon$  از هم مستقل هستند و به‌طور تصادفی از توزیع نرمال استاندارد تولید می‌شوند. کوواریانس ماتریس  $\mathbf{\Lambda}$  به شکل  $diag \mathbf{\Lambda} = (0.15, 0.175, 1)$  انتخاب شده است. برای تولید داده‌هایی با میزان همبستگی متفاوت، مقادیر  $\rho$  برابر ۰/۷۵، ۰/۸۵ و ۰/۹۵ در نمونه‌هایی با اندازه‌های  $n = 25$  و  $n = 50$  در نظر گرفته شده است. داده‌ها در نرم‌افزار



R و با ۱۰۰۰ تکرار شبیه‌سازی شده‌اند. نتیجه حاصل از یک نمونه تصادفی از این ۱۰۰۰ تکرار برای نمونه‌های با اندازه متفاوت در شکل ۱ و جدول ۱ نشان داده شده است.



شکل (۱): رسم اثر  $\Delta_k$  در مقابل  $k$  به ازای  $\rho$  های متفاوت (خط ممتد  $\rho = 0/95$ ، خط نقطه‌ای  $\rho = 0/85$  و خط بریده  $\rho = 0/75$ )

شکل ۱ رسم مقادیر  $\Delta_k$  را به ازای  $k$  متفاوت نشان می‌دهد از هر کدام از دو تصویر شکل ۱ مشخص است که به ازای هر  $\rho$  متفاوت مقادیری از  $k$  وجود دارند که به ازای آن‌ها مقدار اثر  $amse(\hat{\beta}_k)$  از اثر  $amse(\hat{\beta})$  کوچک‌تر است به عبارتی  $\Delta_k$  مثبت است. جدول ۱ مقادیر برآورد  $k$ ،  $\lambda$ ، اثر  $amse(\hat{\beta}_k)$ ،  $amse(\hat{\beta})$  و  $d$  را نشان می‌دهد. لازم به ذکر برآورد  $\lambda$  در دو مرحله به دست آمده است. ابتدا با محاسبه  $\Delta_k$  به ازای مقادیر مختلف  $k$ ،  $k$ ی متناظر به ماکزیمم  $\Delta_k$  مقدار  $k$  مناسب را می‌دهد. سپس با قرار دادن  $k$  مناسب در  $GCV(k, \lambda)$  این تابع نسبت به  $\lambda$  مینیمم می‌شود و  $\lambda$  مشخص می‌شود. با توجه به جدول مقادیر منفی  $d$  رابطه (۱۰) شرط مثبت بودن  $\Delta_k$  برای  $k$ های بهینه را تأیید می‌کنند. در ادامه برای بررسی تحلیل تشخیصی به روش حذف موردی مدل شبیه‌سازی شده بالا را دوباره با اندازه نمونه‌های متفاوت در نظر می‌گیریم. این بار سه مجموعه داده با اندازه‌های ۲۵، ۵۰ و ۱۰۰ تولید می‌شوند. مقدار ضریب همبستگی  $\rho$  و بردار پارامتر  $\beta$  به ترتیب مقدار اختیاری  $\rho = 0/95$  و  $\beta' = (1, -3, 5)$  در نظر گرفته می‌شود. در هر مجموعه داده برای داشتن مقادیر دورافتاده و مؤثر مشاهدات ۵، ۱۰ و ۲۵ با یک مجموعه داده‌های دورافتاده مؤثر جایگزین می‌شود. به این شکل که  $y_{25}$  و  $y_{10}$  را برابر میانگین  $y$  بعلاوه سه برابر انحراف معیار  $y$  و  $\sin 2\pi t_1$  را برابر ۱ قرار می‌دهیم. جدول

۲ مقادیر آماره‌های کوک را با ۵۰۰۰ تکرار نشان می‌دهد. با توجه به جدول مشخص است که هر سه آماره به درستی مشاهدات مؤثر را شناسایی می‌کنند.

جدول (۱): نتایج مربوط به داده‌های شبیه‌سازی

$d$	$amse(\hat{\beta}_k)$	$amse(\hat{\beta})$	$\lambda$	$k$	$\rho$	$n$
-۰/۶۳	۰/۲۴۳	۰/۳۲۳	۰/۱۹	۲/۷	۰/۷۵	
-۰/۵۶	۰/۱۲۰	۰/۲۲۱	۰/۱۱	۱/۸	۰/۸۵	۲۵
-۰/۸۱	۰/۲۰۱	۰/۶۳۸	۰/۰۸	۱/۲	۰/۹۵	
-۰/۴۰	۰/۰۵۱	۰/۱۰۰	۰/۱۳	۱/۷	۰/۷۵	
-۰/۲۲	۰/۱۰۳	۰/۱۸۱	۰/۰۹	۳/۲	۰/۸۵	۵۰
-۰/۱۳	۰/۱۵۲	۰/۵۶۲	۰/۰۵	۳/۹	۰/۹۵	

جدول (۲): مقادیر آماره‌های  $D_{yi}$ ،  $D_{gi}$  و  $D_{\beta i}$  برای داده‌های شبیه‌سازی شده. مقادیر داخل پرانتز مقادیر انحراف استاندارد می‌باشند.

$D_{yi}$	$D_{gi}$	$D_{\beta i}$	مشاهدات	اندازه نمونه
۰/۱۰۵ (۰/۰۰۸۲)	۰/۰۵۵ (۰/۰۰۲۳)	۰/۰۹۱ (۰/۰۰۹۶)	۵	۲۵
۰/۱۴۷ (۰/۰۰۷۷)	۰/۱۵۱ (۰/۰۰۲۶)	۰/۰۷۷ (۰/۰۰۹۹)	۱۰	
۰/۲۱۸ (۰/۰۰۳۹)	۰/۰۸۱ (۰/۰۰۳۵)	۰/۱۰۴ (۰/۰۰۸۴)	۲۵	
۰/۰۷۸ (۰/۰۰۸۵)	$\leq ۰/۰۴۳$ (۰/۰۰۵۰)	$\leq ۰/۰۲۵$ (۰/۰۰۹۱)	دیگر مشاهدات	
۰/۲۹۳ (۰/۰۰۹۲)	۰/۱۰۶ (۰/۰۰۴۱)	۰/۱۱۹ (۰/۰۰۷۱)	۵	۵۰
۰/۴۰۰ (۰/۰۰۶۱)	۰/۳۱۰ (۰/۰۰۵۷)	۰/۱۰۱ (۰/۰۰۱۹)	۱۰	
۰/۳۴۷ (۰/۰۰۳۱)	۰/۰۹۸ (۰/۰۰۷۱)	۰/۲۱۸ (۰/۰۰۷۸)	۲۵	
۰/۰۹۰ (۰/۰۰۷۱)	$\leq ۰/۰۶۵$ (۰/۰۰۸۸)	$\leq ۰/۰۴۳$ (۰/۰۰۶۶)	دیگر مشاهدات	
۰/۳۷۹ (۰/۰۰۱۱)	۰/۰۰۹ (۰/۰۰۱۳)	۰/۴۱۷ (۰/۰۰۲۱)	۵	۱۰۰
۰/۶۶۰ (۰/۰۰۲۳)	۰/۶۲۰ (۰/۰۰۱۱)	۰/۲۱۶ (۰/۰۰۵۳)	۱۰	
۰/۷۲۷ (۰/۰۰۱۱)	۰/۱۰۳ (۰/۰۰۳۱)	۰/۵۳۳ (۰/۰۰۱۲)	۲۵	
۰/۰۷۵ (۰/۰۰۰۱)	$\leq ۰/۰۴۳$ (۰/۰۰۱۷)	$\leq ۰/۰۶۴$ (۰/۰۰۱۰)	دیگر مشاهدات	

از جدول ۲ مشخص است که با افزایش حجم نمونه مشاهدات دورافتاده با توجه به اندازه‌ای که می‌گیرند آشکارتر از سایر مشاهدات ظاهر می‌شوند. این در حالی است که در نمونه‌های کوچک‌تر مشاهدات دورافتاده (مشاهده ۵ و ۱۰ در نمونه با اندازه ۲۵) به راحتی آشکار نمی‌شوند.

### ۵-۲- داده‌های سفال مصری

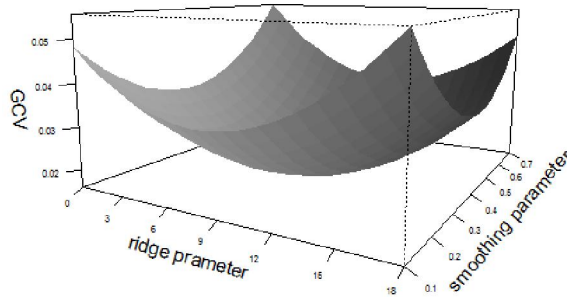
در این بخش برای انجام تحلیل عددی و نموداری مباحث تشخیصی مطرح شده در بخش‌های قبل از یک مجموعه داده شناخته‌شده به نام داده‌های سفال مصری استفاده می‌کنیم. به‌طور خلاصه این مجموعه داده مربوط به گزارش حاصل از بررسی‌های باستان‌شناسی گسترده‌ای از تولید و پخش سفال در شهر باستانی الامرنا مصر است. داده‌ها شامل اندازه‌گیری محتویات شیمیایی (عناصر معدنی) موجود در بسیاری از نمونه‌های سفال است که با استفاده از دو روش متفاوت یکی روش تجزیه و تحلیل نوترون ( $NAA$ ) و دیگری روش پلاسما القایی ( $ICP$ ) به دست آمده‌اند. برای توضیحات بیشتر روش اندازه‌گیری اسمیت و همکاران [۳۱] را ببینید. مجموعه سفال‌ها از مکان‌های مختلف اطراف شهر جمع‌آوری شده‌اند. در کل دو نوع خاک رس لای و مارن در ساخت سفال‌های مصر باستان استفاده می‌شده است و باستان‌شناسان برای جدا کردن قطعات سفال‌های مشترک (که از کشورهای آفریقایی شمالی وارد می‌شدند) از سفال لای و مارن مصری، آن‌ها را گروه‌بندی کرده‌اند. نوع گروه‌بندی ایشیا سفالی با توجه به دو منبع اصلی، یکی کد طرح و دیگری کد مکان سفال است که هر کدام از این دو نوع گروه‌بندی برای باستان‌شناسان حائز اهمیت است. نهایتاً با توجه به این نوع از گروه‌بندی قطعات سفال انتخابی به ۲۸ گروه تقسیم‌بندی شدند که گروه‌های ۶ و ۱۵ و ۱۸ و ۲۳ سفال‌های وارداتی بودند. در هر گروه تعداد متفاوتی از سفال‌ها دارای کد طرح و کد مکانی یکسانی وجود دارد که در اصل می‌تواند به‌عنوان مشاهدات تکراری در نظر گرفته شود. از میان تمام عناصر معدنی علاقه‌مند به بررسی رابطه بین عنصر  $Na$  به‌عنوان متغیر پاسخ که با روش  $NAA$  اندازه‌گیری شده است در برابر شش عنصر  $Cr, V, K, Al, Na$  و  $Mn$  به‌عنوان متغیرهای توضیحی که با روش  $NAA$  اندازه‌گیری شده‌اند هستیم. راسخ [۲۱] این داده‌ها را با برازش یک مدل خطی پارامتری کامل با خطا در اندازه‌گیری مورد مطالعه قرار داده است. او نشان داد که یک همخطی با اندازه متوسط بین متغیرهای توضیحی وجود دارد. به‌جای استفاده از انتخاب زیرمجموعه از متغیرها او برآورد ریح پارامترها را با  $k = 0.15$  به دست آورد و نشان داد که این نوع برآوردگر برآورد پارامترها را بهبود می‌بخشد. اخیراً قپانی و همکاران [۲۰] از نگاهی دیگر برآورد پارامترهای ریح وزنی متغیرهای فوق را با الگوی مدل خطی پارامتری محدود شده با اندازه‌گیری در خطا مورد بررسی و تحلیل قرار داده‌اند. از طرفی برای بررسی و شناسایی مشاهدات مؤثر راسخ [۲۱] و قپانی و همکاران [۲۲] از رویکرد مباحث تشخیصی به بررسی و مطالعه این داده‌ها پرداخته‌اند. آن‌ها این مجموعه داده را با استفاده

از روش‌های تأثیر موضعی و انتقال میانگین دورافتاده به ترتیب در مدل خطی پارامتری ریح با خطا در اندازه‌گیری و مدل‌های خطی پارامتری با اندازه‌گیری در خطا تحت محدودیت خطی تصادفی مورد تحلیل و بررسی قرار دادند و گروه‌های ۱۸، ۶ و ۱۲ (به ترتیب با بیشترین تأثیر) در مدل خطی پارامتری ریح با خطا در اندازه‌گیری و گروه ۱۳ در برازش مدل خطی پارامتری با اندازه‌گیری در خطا تحت محدودیت خطی تصادفی به‌عنوان گروه‌های مؤثر شناسایی کردند. از آنجایی که رابطه بین متغیر (پاسخ)  $Na$  و متغیر  $Mn$  را می‌توان به شکل غیرخطی فرض کرد (شکل ۳) و با توجه به وجود همخطی در ستون‌های ماتریس متغیرهای توضیحی با خطا در اندازه‌گیری (طبق پژوهش‌های پیشین) می‌توان مدل خطی نیمه‌پارامتری (۱) را با در نظر گرفتن به‌عنوان متغیر مربوط به مؤلفه‌های پارامتری و  $g(t) = Mn$  به‌عنوان متغیر مربوط به مؤلفه نا پارامتری مدل برای برازش به این مجموعه از داده‌ها در نظر گرفت. ابتدا از محاسبات مشخص می‌شود نسبت بزرگ‌ترین مقدار ویژه به کوچک‌ترین مقدار ویژه ماتریس  $\mathbf{X}'(\mathbf{I}-\mathbf{S})\mathbf{X}$  برابر  $\frac{\lambda_{max}}{\lambda_{min}} = 9871/5$  می‌باشد که نشان از یک همخطی شدید بین متغیرهای هموار شده است. لذا در اینجا برای برآورد پارامترها از روش رگرسیون ریح با یک  $k$  مناسب می‌توان استفاده نمود. با توجه به اینکه در داده‌ها تکرار وجود دارد یک برآورد ناریب برای  $\Lambda$  را می‌توان از روش لیانگ و همکاران [۳۲] به دست آورد. با استفاده از این روش برآورد ماتریس  $\Lambda$  در مدل ما برابر است با:

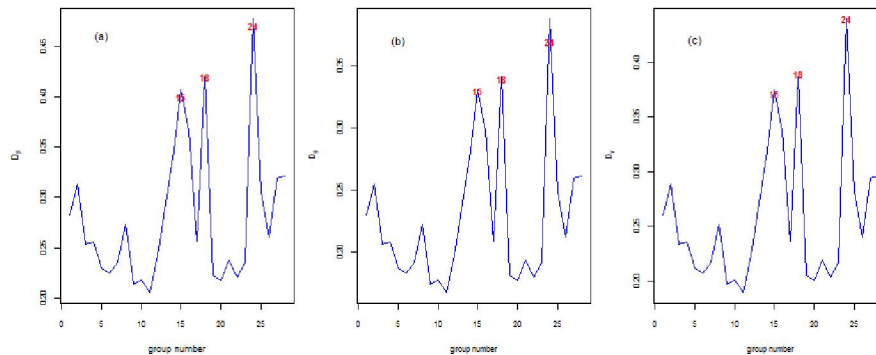
$$\hat{\Lambda} = \begin{bmatrix} 2/356 & 1/344 & 0/0190 & 0/0940 & 0/2291 \\ 17/724 & 18/880 & 0/138 & 1/357 & 0940/0 \\ 1/093 & -0/099 & 0/150 & 0/1380 & 0/0190 \\ 291/159 & 747/272 & -0/099 & 18/180 & 1/344 \\ 444/368 & 291/159 & 1/093 & 17/724 & 1/356 \end{bmatrix}$$

با توجه به شکل ۲ با مینیمم کردن معیار اعتبارسنجی متقابل هم‌زمان نسبت به  $k$  و  $\lambda$  برآورد هم‌زمان پارامتر ریح و پارامتر هموارساز برابر ۱۰/۲۳ و ۰/۳۵ به دست می‌آید. نخست با استفاده از رویکرد حذفی به ارزیابی مشاهدات می‌پردازیم. شکل ۳ مقادیر آماره‌های کوک به ترتیب از چپ به راست  $D_\beta$ ،  $D_g$  و  $D_y$  را در برابر شماره گروه‌های سفال نمایش می‌دهد. با توجه به شکل (۳.a) مشاهده ۲۴، ۱۸ و ۱۵ به ترتیب به‌عنوان مشاهدات مؤثر با بیشترین تأثیر روی بردار  $\hat{\beta}_k$  شناسایی می‌شوند. یادآور می‌شویم چنانکه در ابتدا بیان شد گروه‌های ۱۵ و ۱۸ جزو گروه‌های وارداتی می‌باشند. از شکل (۳.b) و (۳.c) روشن است که همین مشاهدات می‌توانند روی مقادیر برازشی مؤلفه نا پارامتری  $g$  و مقادیر برازشی کل  $\hat{y}_k$  مؤثر واقع شوند. این نتیجه زیاد دور از انتظار نیست زیرا که بخش اصلی آماره‌های  $D_\beta$ ،  $D_g$  و  $D_y$  به باقیمانده‌های کل  $e_i$  و مقادیر

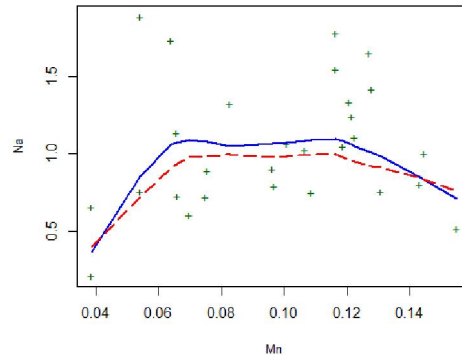
نافذ کل مدل یعنی  $h_{ii}$  بستگی دارند. یادآوری می‌شود که همیشه مشاهدات یکسانی روی تمامی اجزای مدل اثر ندارند و ممکن است مشاهدات با توجه به نوع داده متفاوت باشند. تأثیر این گروه‌ها در برازش مؤلفه نا پارامتری  $g$  در شکل ۴ نشان داده شده است در واقع این شکل مقادیر برازشی  $\hat{g}$  قبل از حذف سه مشاهده مؤثر ۱۵، ۱۸ (گروه‌های وارداتی) و ۲۴ را با خط ممتد و بعد از حذف مشاهدات مؤثر را با خط بریده به نمایش داده است. کاملاً مشهود است که وجود این مشاهدات تأثیر بسزایی در برآورد  $\hat{g}$  دارد.



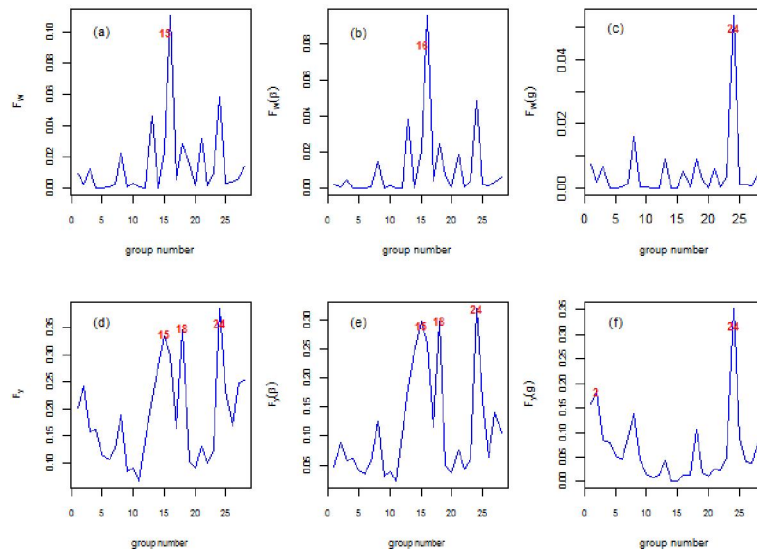
شکل (۲): نمودار  $GCV$  در مقابل پارامتر ریح  $k$  و پارامتر هموارساز  $\lambda$  برای داده‌های سفال مصری



شکل (۳): مقادیر (a) - آماره  $D_\beta$  (b) - آماره  $D_g$  (c) - آماره  $D_y$  در مقابل شماره گروه‌ها



شکل (۴): برازش مؤلفه نا پارامتری: خط پریده با حذف سه گروه مؤثر ۱۵، ۱۸ و ۲۴ - خط ممتد بدون گروه‌های مؤثر



شکل (۵): تحلیل تأثیر مکانی داده‌های سفال مصری: عناصر روی قطر اصلی ماتریس‌های (a)  $\ddot{\mathbf{F}}_w(\gamma)$ ، (b)  $\ddot{\mathbf{F}}_w(\beta)$ ، (c)  $\ddot{\mathbf{F}}_w(g)$ ، (d)  $\ddot{\mathbf{F}}(\gamma)$ ، (e)  $\ddot{\mathbf{F}}(\beta)$  و (f)  $\ddot{\mathbf{F}}(g)$  نمایانگر سهم هر یک از گروه‌ها در ماتریس‌های تأثیر می‌باشند.

## ۶- نتیجه‌گیری

شناسایی مشاهدات مؤثر در مدل‌های خطی نیمه پارامتری در حضور همخطی چندگانه و وجود خطا در اندازه‌گیری از دو رویکرد اصلی روش حذف موردی و تأثیر موضعی مورد مطالعه قرار گرفت. با توجه به روش حذف موردی اندازه‌های تأثیر توسط مقادیر نافذ و باقیمانده‌ها با استفاده از فرمول‌های به‌دست‌آمده در قضیه ۲ بخش ۴-۱ محاسبه می‌شوند. معتقدیم که فرمول‌های حذف موردی که در این مقاله به آن‌ها رسیدیم به‌عنوان بخشی از هر تحلیل داده‌ی مهم می‌توانند کمک‌کننده باشند. از طرفی هدف از شاخص‌های موضعی شناسایی و درک بهتر مشاهداتی است که ممکن است در تحلیل مرتبط با یک مدل خاص اهمیت داشته باشد. مقاله حاضر پژوهش کوک [۲۸] را به‌منظور سنجش تأثیر موضعی تحت انحراف وزن‌های نمونه، متغیر پاسخ و یا متغیرهای کمکی برای برآوردگرهای درست‌نمایی تصحیح‌شده توانانیده در مدل خطی نیمه پارامتری، بسط می‌دهد. بر این باوریم که شاخص‌های تأثیر موضعی که در اینجا به آن می‌رسیم می‌تواند به‌عنوان بخشی از هر تحلیل داده جدی مفید باشد. مخصوصاً که در برگیری یک تابع نا پارامتری می‌تواند تأثیر مهمی روی تحلیل تأثیر برای برآورد ضریب رگرسیون خطی داشته باشد. اندازه‌های تأثیر ارائه‌شده در این مقاله نقشی را در فهم اینکه چگونه هر مؤلفه برای برآورد دیگری کمک می‌کند، ایفا می‌کند. به ویژه ماتریس‌های نافذ شرطی و حاشیه‌ای برای دو مؤلفه می‌توانند به‌عنوان ماتریس‌های تأثیر جزئی تحت انحراف متغیر پاسخ در نظر گرفته شوند. آنالیز تأثیر تحت انحراف متغیرهای کمکی نقطه‌نظر متناوبی را برای مدل‌های خطی خطا در اندازه‌گیری ارائه می‌کند. انحراف وزنی مشاهدات نیز نظریه حذف موردی را تعمیم می‌دهد طوری که یک تقریب خوب به شاخص‌های حذف موردی بدون اجبار برای دوباره برآورد کردن پارامترهای هر حذف را فراهم می‌کند. ماتریس تأثیر متناظر می‌تواند در شناسایی مشاهدات مؤثر کمک‌کننده باشد. با توجه به اینکه اندازه تأثیر یک مشاهده تنها با افزایش اندازه نمونه گرایش به تقلیل دارد، شاخص‌های تأثیر احتمالاً در نمونه کوچک یا نسبتاً کم محبوب‌ترین می‌باشند.

## ضمیمه:

برای رابطه (۸) و (۹) با مشتق‌گیری از رابطه (۷) نسبت به  $\beta$  و  $g$  به‌سادگی خواهیم داشت:

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} - n\mathbf{I}_p & \mathbf{X}'\mathbf{N} \\ \mathbf{N}'\mathbf{X} & \mathbf{N}'\mathbf{N} + \lambda\mathbf{M} \end{pmatrix} \begin{pmatrix} \beta \\ g \end{pmatrix} = \begin{pmatrix} \mathbf{X}' \\ \mathbf{N}' \end{pmatrix} y \quad (1)$$

از معادله دوم ماتریس داریم:

$$\mathbf{N}'\mathbf{X}\beta + (\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})g = \mathbf{N}'y \rightarrow g = (\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1}\mathbf{N}'(y - \mathbf{X}\beta)$$

با قرار دادن  $g$  در معادله اول ماتریس و تعریف  $\mathbf{S} = \mathbf{N}(\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})^{-1}\mathbf{N}'$  روابط

$$\begin{aligned}(\mathbf{X}'\mathbf{X} - n\mathbf{\Lambda} + k\mathbf{I}_p)\beta + \mathbf{X}'\mathbf{N}g &= (\mathbf{X}'\mathbf{X} - n\mathbf{\Lambda} + k\mathbf{I}_p)\beta + \mathbf{X}'\mathbf{S}(y - \mathbf{X}\beta) = \mathbf{X}'y \\&\rightarrow (\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{S}\mathbf{X} - n\mathbf{\Lambda} + k\mathbf{I}_p)\beta = \mathbf{X}'y - \mathbf{X}'\mathbf{S}y \\&\rightarrow (\mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} - n\mathbf{\Lambda} + k\mathbf{I}_p)\beta = \mathbf{X}'(\mathbf{I} - \mathbf{S})y\end{aligned}$$

به دست می‌آیند. از رابطه اخیر خیلی راحت  $\hat{\beta}_k$  حاصل می‌شود که با قرار دادن آن در  $g$  مقدار  $\hat{g}_k$  نیز نتیجه می‌شود. از حل دو معادله فوق رابطه‌های (۸) و (۹) به دست می‌آیند.

**تعریف ۱:** برای دو ماتریس  $\mathbf{A}$  و  $\mathbf{B}$  گوییم  $\mathbf{A} \succ \mathbf{B}$  یا  $\mathbf{A} - \mathbf{B} \succ \mathbf{0}$  اگر  $\mathbf{A} - \mathbf{B}$  معین مثبت باشد.

**اثبات قضیه ۱:** برای اثبات ابتدا لم زیر تعریف می‌شود:

**لم ۱:** از فونگ و همکاران [۶] داریم:

$$\mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} = \mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z} + n\mathbf{\Lambda} + O_p\left(n^{-\frac{1}{2}}\right)$$

با توجه به لم ۱ رابطه‌های زیر به دست می‌آیند:

$$n^{-1}(\mathbf{X}'(\mathbf{I}_n - \mathbf{S})\mathbf{X} - n\mathbf{\Lambda} + k\mathbf{I}_p) = n^{-1}(\mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z} + k\mathbf{I}_p) + O_p\left(n^{-\frac{1}{2}}\right)$$

$$\begin{aligned}\sqrt{n}\hat{\beta}_k &= \left[ n^{-1}(\mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z} + k\mathbf{I}_p) + O_p\left(n^{-\frac{1}{2}}\right) \right]^{-1} n^{-\frac{1}{2}}\mathbf{X}'(\mathbf{I}_n - \mathbf{S})y \\&= \left[ \mathbf{I}_p + O_p\left(n^{-\frac{1}{2}}\right) \right]^{-1} \left[ n^{-1}(\mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z} + k\mathbf{I}_p) \right]^{-1} n^{-\frac{1}{2}}\mathbf{X}'(\mathbf{I}_n - \mathbf{S})y \\&= \left[ \mathbf{I}_p + O_p\left(n^{-\frac{1}{2}}\right) \right] \left[ n^{-1}(\mathbf{Z}'(\mathbf{I}_n - \mathbf{S})\mathbf{Z} + k\mathbf{I}_p) \right]^{-1} n^{-\frac{1}{2}}\mathbf{X}'(\mathbf{I}_n - \mathbf{S})y\end{aligned}$$

با توجه رابطه آخر  $\sqrt{n}\hat{\beta}_k$  به شکل زیر خلاصه می‌شود:

$$\sqrt{n}\hat{\beta}_k = \mathbf{C}^{-1}\xi + O_p\left(n^{-\frac{1}{2}}\right)$$



می‌توان نشان داد که در آن  $\xi = n^{-\frac{1}{2}} \mathbf{X}'(\mathbf{I}_n - \mathbf{S})y$  دارای توزیع مجانبی نرمال هست (به فونگ و همکاران [۶] مراجعه شود). از آنجایی که  $E(\xi) = n^{-\frac{1}{2}} \mathbf{C}_0 \beta + O_p\left(n^{-\frac{1}{2}}\right)$  در نتیجه خواهیم داشت:

$$\sqrt{n}(\hat{\beta}_k - \mathbf{C}_0 \mathbf{C}_k^{-1} \beta) = \mathbf{C}_k^{-1} [\xi - E(\xi)] + O_p\left(n^{-\frac{1}{2}}\right)$$

که نشان می‌دهد  $\sqrt{n}(\hat{\beta}_k - \mathbf{C}_0 \mathbf{C}_k^{-1} \beta)$  دارای توزیع مجانبی نرمال با میانگین صفر می‌باشد. از طرفی با استفاده از واریانس شرطی برای  $\xi$  داریم:

$$\begin{aligned} \text{var}(\xi) &= E(\text{var}(\xi|y)) + \text{var}(E(\xi|y)) \\ &= n^{-1} E(y'(\mathbf{I}_n - \mathbf{S})' y \Lambda) + n^{-1} \text{var}(\mathbf{X}'(\mathbf{I}_n - \mathbf{S})y) \\ &= n^{-1} \left[ (\mathbf{Z}\beta + \mathbf{N}g)' (\mathbf{I}_n - \mathbf{S})' (\mathbf{Z}\beta + \mathbf{N}g) + \sigma^2 \text{tr}(\mathbf{I}_n - \mathbf{S}) \right] + n^{-1} \mathbf{Z}' (\mathbf{I}_n - \mathbf{S})' \mathbf{Z} \sigma^2 \\ &= n^{-1} [\mathbf{D} + \mathbf{Z}' (\mathbf{I}_n - \mathbf{S})' \mathbf{Z} \sigma^2] = n^{-1} \Phi \end{aligned}$$

از آنجایی که  $\text{avar}(\sqrt{n} \hat{\beta}_k) = \mathbf{C}_k^{-1} \text{var}(\xi) \mathbf{C}_k^{-1}$  بنابراین داریم:

$$\text{avar}(\hat{\beta}_k) = \mathbf{C}_k^{-1} [\mathbf{D} + \mathbf{Z}' (\mathbf{I}_n - \mathbf{S})' \mathbf{Z} \sigma^2] \mathbf{C}_k^{-1}$$

و اثبات قضیه کامل است.

**لم ۲:** فرض کنید  $\mathbf{A}$  یک ماتریس معین نامنفی  $n \times n$ ،  $a$  یک بردار  $n \times 1$  باشد آنگاه  $\mathbf{A} - aa'$  معین نامنفی است اگر و تنها اگر  $a' \mathbf{A}^{-1} a \leq 1$  باشد که در آن  $\mathbf{A}^{-1}$  وارون تعمیم‌یافته ماتریس  $\mathbf{A}$  است. (رائو ۲۰۰۸).

از رابطه اول قضیه ۱ برای  $D_{\beta_i}$  می‌توان نوشت:

$$\begin{aligned}
D_{\beta_i} &= \frac{(\hat{\beta}_k - \hat{\beta}_{k(i)})' \mathbf{X}' \mathbf{X} (\hat{\beta}_k - \hat{\beta}_{k(i)})}{\sigma^2 \text{tr}(\tilde{\mathbf{H}}_k)} \\
&= \frac{\xi_i e_i (\mathbf{I}_n - \mathbf{S}) \mathbf{X} \{ \mathbf{X}' (\mathbf{I}_n - \mathbf{S}) \mathbf{X} + k \mathbf{I}_p - n \Lambda \}^{-1} \mathbf{X}' \mathbf{X} \{ \mathbf{X}' (\mathbf{I}_n - \mathbf{S}) \mathbf{X} + k \mathbf{I}_p - n \Lambda \}^{-1} \mathbf{X}' (\mathbf{I}_n - \mathbf{S}) e_i \xi_i}{\sigma^2 \text{tr}(\tilde{\mathbf{H}}_k) (1 - h_{ii})^2} \\
&= \frac{\xi_i e_i \tilde{\mathbf{H}}_k' \tilde{\mathbf{H}}_k \xi_i e_i}{\sigma^2 \text{tr}(\tilde{\mathbf{H}}_k) (1 - h_{ii})^2} = \frac{\left[ \sum_{j=1}^n \tilde{h}_{ij}^2 \right] e_i^2}{\sigma^2 \text{tr}(\tilde{\mathbf{H}}_k) (1 - h_{ii})^2}
\end{aligned}$$

محاسبات ماتریس تأثیر رابطه (۱۱) را برای همه ماتریس‌های تأثیر استفاده می‌کنیم. ماتریس  $\tilde{\mathbf{L}}$  برای همه موارد یکسان است که معکوس آن برابر است با

$$\begin{aligned}
\tilde{\mathbf{L}}^{-1} &= \begin{pmatrix} \mathbf{L}^{\text{I}} & \mathbf{L}^{\text{II}} \\ \mathbf{L}^{\text{III}} & \mathbf{L}^{\text{IV}} \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{X}'(\mathbf{I} - \mathbf{S})\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1} & -(\mathbf{X}'(\mathbf{I} - \mathbf{S})\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1} \mathbf{X}' \mathbf{N} \mathbf{T}^{-1} \\ -\mathbf{T}^{-1} \mathbf{N}' \mathbf{X} (\mathbf{X}'(\mathbf{I} - \mathbf{S})\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1} & \mathbf{T}^{-1} + \mathbf{T}^{-1} \mathbf{N}' \mathbf{X} (\mathbf{X}'(\mathbf{I} - \mathbf{S})\mathbf{X} - n\Lambda + k\mathbf{I}_p)^{-1} \mathbf{X}' \mathbf{N} \mathbf{T}^{-1} \end{pmatrix},
\end{aligned}$$

که  $\mathbf{T} = (\mathbf{N}'\mathbf{N} + \lambda\mathbf{M})$  است.

برای محاسبه  $\Delta$  متناظر به انحراف وزن نمونه در بخش ۴-۲-۱ داریم:

$$\begin{aligned}
\frac{\partial l^*(\theta, \omega)}{\partial \omega_i} &= \frac{-1}{2\sigma^2} [\|y - \mathbf{X}\beta - \mathbf{N}g\|_i^2] \rightarrow \frac{\partial l^*(\theta, \omega)}{\partial \omega_i \partial \beta} = \sigma^{-2} [\mathbf{X}'(y - \mathbf{X}\beta - \mathbf{N}g)]_i \\
\frac{\partial l^*(\theta, \omega)}{\partial \omega_i \partial g} &= \sigma^{-2} [\mathbf{N}'(y - \mathbf{X}\beta - \mathbf{N}g)]_i
\end{aligned}$$

که  $[\mathbf{A}]_i$  نشان‌دهنده  $i$ امین عضو بردار  $\mathbf{A}$  است. با تعریف  $\beta = \hat{\beta}$  و  $g = \hat{g}$  در مشتقات بالا و نمایش ماتریسی خواهیم داشت:

$$\Delta = \hat{\sigma}^{-2} (\mathbf{X}, \mathbf{N})' \text{diag}(y - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{g})$$

برای  $\Delta$  تحت انحراف  $i$ -امین متغیر کمکی، در بخش ۴-۲-۳ داریم:

$$\Delta = - \begin{pmatrix} l_i' \hat{\beta}_{ki} \\ \mathbf{N}' \end{pmatrix} \begin{pmatrix} \mathbf{X}' \\ \circ \end{pmatrix} + \begin{pmatrix} l_i e' \\ \circ \end{pmatrix}.$$

فرض کنید  $\hat{\beta}_{ki} = (l_i' \hat{\beta}_{ki})$ . بنابراین از رابطه (۱۱) می‌توان نوشت:

$$\begin{aligned} \ddot{F}_{x,i} = \Delta' \ddot{L}^{-1} \Delta = \hat{\beta}_{ki}^y (\mathbf{X} \ \mathbf{N}) \ddot{L}^{-1} (\mathbf{X} \ \mathbf{N})' - \hat{\beta}_{ki} (\mathbf{X} \ \mathbf{N}) \ddot{L}^{-1} (el_i' \ 0)' \\ - \hat{\beta}_{ki} (el_i' \ 0) \ddot{L}^{-1} (\mathbf{X} \ \mathbf{N})' + (el_i' \ 0) \ddot{L}^{-1} (el_i' \ 0). \end{aligned}$$

از تعریف  $\mathbf{H}$  و  $\ddot{L}^{-1}$  می‌توان محاسبات زیر را نتیجه گرفت:

$$\begin{aligned} \ddot{F}_{x,i} = \hat{\beta}_{ki}^y \mathbf{H} - \hat{\beta}_{ki} (\mathbf{X} \mathbf{L}' + \mathbf{N} \mathbf{L}') l_i e' - \hat{\beta}_{ki} el_i' (\mathbf{L}' \mathbf{X}' + \mathbf{L}' \mathbf{N}') + el_i' \mathbf{L}' l_i e' \\ = \hat{\beta}_{ki}^y (\mathbf{S} + (\mathbf{I} - \mathbf{S}) \mathbf{X} (\mathbf{X}' (\mathbf{I} - \mathbf{S}) \mathbf{X} - n\mathbf{A} + k\mathbf{I}_p)^{-1} \mathbf{X}' (\mathbf{I} - \mathbf{S})) \\ - el_i' (\mathbf{X}' (\mathbf{I} - \mathbf{S}) \mathbf{X} - n\mathbf{A} + k\mathbf{I}_p)^{-1} \mathbf{X}' (\mathbf{I} - \mathbf{S}) \hat{\beta}_{ki} \\ - \hat{\beta}_{ki} (\mathbf{I} - \mathbf{S}) \mathbf{X} (\mathbf{X}' (\mathbf{I} - \mathbf{S}) \mathbf{X} - n\mathbf{A} + k\mathbf{I}_p)^{-1} l_i e' + el_i' (\mathbf{X}' (\mathbf{I} - \mathbf{S}) \mathbf{X} - n\mathbf{A} + k\mathbf{I}_p)^{-1} l_i e' \\ = \hat{\beta}_{ki}^y \mathbf{S} + \mathbf{A}' (\mathbf{X}' (\mathbf{I} - \mathbf{S}) \mathbf{X} - n\mathbf{A} + k\mathbf{I}_p)^{-1} \mathbf{A}, \end{aligned}$$

که در آن  $\mathbf{A} = l_i e' - \hat{\beta}_{ki} \mathbf{X}' (\mathbf{I} - \mathbf{S})$ . محاسبه ماتریس‌های تأثیر دیگر به همین صورت به دست آمده است که به خاطر شباهت حذف شده است.

### تشکر و قدردانی

نویسندگان مقاله از داوران محترم و سردبیر محترم به خاطر نکته‌ها و رهنمودهای ارزشمندشان که موجب ارتقای کیفیت مقاله از لحاظ نگارشی و علمی شدند، بسیار سپاس‌گذار هستند.

### منابع

- [1] Cook, R. D. (1977). Detection of Influential Observations in Linear Regression. *Technometrics*, 19, 15-18.
- [2] Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall, New York
- [3] Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, New York.
- [4] Kim, C. (1996). Cook's Distance in Spline Smoothing, *Statist Proba Lett.*, 31, 139-144.
- [5] Kim, C., Park, B. U. and Kim, W. (2002). Influence Diagnostics in Semiparametric Regression Models, *Statist. Probab. Lett.*, 60, 49-58.
- [6] Fung, W. K., Zhu, Z. Y., Wei, B. C. and He, X. (2002). Influence Diagnostics and Outlier Tests for Semiparametric Mixed Models. *J. R. Stat. Soc. Ser. B*, 47, 332-341.

- 
- [7] Belsley, D. A. (1991). *Collinearity Diagnostics: Collinearity and Weak Data in Regression*. John Wiley & Sons, New York.
- [8] Hoerl, A. E. and Kennard R.W. (1970). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **12**: 69–82
- [9] Duran, E. A. and Akdeniz, F. (2012). Efficiency of the Modified Jackknifed Liu-type Estimator. *J. Stat Paper*. **53**(2): 265-280.
- [10] Roozbeh, M., and Arashi, M. (2013). Feasible Ridge Estimator in Partially Linear Models, *Multivariate Analysis*, **116**, 35-44.
- [11] Roozbeh, M. (2015). Shrinkage Ridge Estimator in Semiparametric Regression Models. *J. Multivariate Anal.* **136**, 56-74.
- [12] Emami, H. (2015). Influence Diagnostics in Ridge Semiparametric Regression Models. *J. Stat Prob Lett*, **105**, 106-115.
- [13] Emami, H. (2016). Local Influence for Liu Estimators in Semiparametric Linear Models. *Stat Pap*, Doi:10.1007/s00362-016-0775-6.
- [14] Fuller, W. A. (1987). *Measurements Error Models*. John Wiley & Sons, New York.
- [15] Hanfelt J. and Liang, K. Y. (1997). Approximate Likelihood for Generalized Linear Errors in Variables Models, *J. Roy. Statist. Soc. Ser. B*, **59**, 627-637.
- [16] Nakamura, T. (1990). Corrected score Functions for Error in Variables Models: Methodologie and Application to Generalized Linear Models. *Biometrika*, **77**, 127-137.
- [17] Stefanski, L. A. and Carrol, R. J. (1987). Conditional Scores and Optimal Scores for Generalized Linear Measurement reeor Models. *Biometrika*, **74**, 703-716.
- [18] Rasekh, A. R. (2001). Ridge Estimation in Functional Measurement Error Models. *Inst. Stat. Univ. Paris*, **45**, 47-59.
- [19] Saleh A.K. and Md. Shalabh. (2014). A ridge regression estimation approach to the measurement error model. *J. Mul. Anal.* **123**: 68-84.
- [20] Ghapani, F. Rasekh, A. R. and Babadi, B. (2016). The Weighted Ridge Estimator in Stochastic Restricted Linear Measurement Error Models. *J. Stat Paper*. Doi:10.1007/s00362-016-0786-3
- [21] Rasekh, A. R. (2006). Local Influence in Measurement Error Models with Ridge Estimate. *Computational Statistic and Data Analysis*, **50**, 2822-2834.

- [22] Ghapani, F., Rasekh, A. R. and Babadi, B. (2015). Mean Shift and Influence Diagnostics in linear Mixed Measurement Error Models. *J. Am Math Mange Sci*, **23**, 37-59
- [23] Hoerl, A. E. and Kennard, R. W. (1976). Ridge Regression: Iterative Estimation of the Biasing Parameter. *Communications in Statistics- theory and Methods*, **5**, 77-88.
- [24] Duran, E. A., Hardel, W. K. and Osipenko, M. (2012). Difference Based Ridge and Liu Type Estimators in Semiparametric Regression Models. *J. Multivariate Anal*, **105**, 164-175
- [25] Green, P.J., and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*, London: Chapman and Hall.
- [26] Roozbeh, M. (2017). Optimal QR-based Estimation in Partially Linear Regression Models With Correlated errors using GCV criterion. *Computational Statistics and Data Analysis*. DOI:<http://dx.doi.org/10.1016/j.csda.2017.08.002>
- [27] Arashi, M. and Roozbeh M. (2015). Some Improved Estimation Strategies in High-Dimensional Semiparametric Regression Models with Application to Riboflavin Production Data. *J. Stat. Paper*. DOI 10.1007/s00362-016-0843-y.
- [28] Cook, R. D. (1986). Assessement of Local Influence (with discussion). *J. Roy Statist Soc Ser. B* **48**, 133-169.
- [29] Thomas, W. and Cook, R. D. (1989). Assessing Influence On Regression Coefficients in Generalized Linear Models. *Biometrika* **76**, 741–749.
- [30] Kwan, C. W. and Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: application to factor analysis. *Psychometrika* **63**, 35–46.
- [31] Smith, D.M., Hart, F.A., Symond, R.D. and Walsh, J.N. (1988). Analysis of Roman Pottery From Colchester by Inductively Coupled Plasma Spectrometry. In: Slater, E.A., Tate, J.O. (Eds.), *Science and Archaeology Glasgow 1987*, vol. 196(I). B.A.R., Oxford, pp. 41–55.
- [32] Liang, H. Hardle, W. and Carroll, R.J. (1999). Estimation in a semiparametric partially linear errors-in-variables model, *Ann. Statist.* **27**: 1519–1535.

## **Influence Diagnostics in Ridge Semiparametric Linear Measurement Error Models**

Hadi Emami

Department of Statistics, University of Zanjan, Zanjan, Iran

### **Abstract**

In this paper influence diagnostics in the semi parametric linear measurement error is studied when the problem of multicollinearity is exist. First to combat multicollinearity, some ridge estimators are proposed based the penalized corrected likelihood approach, then using case deletion method the influence measures are defined to detect influential observations. In continue it is shown these measures are function of residuals and leverages. Furtherer more local influence approach is developed based the penalized likelihood function for assessing the influence of points. A simulation study is performed to illustrate the efficiency of proposed ridge estimators based on the asymptotical mean square error. Finally, the influence diagnostic approaches are applied to a real data.

**Keyword:** Generalized Cross Validation, Spline Smoothing, Measurement Error, Multicollinearity, Ridge Estimator, Semiparametric Linear Model

**Mathematics Subject Classification (2010):** 62J05, 62J0