

مجله علوم آماری، پاییز و زمستان ۱۳۸۶

جلد ۱، شماره ۲، ص ۹۵-۱۰۸

## مقایسه مخاطره انواع برآوردگرها در مدل رگرسیون چندگانه با خطای $t$ چندگانه

محمد آرشی، سید محمد مهدی طباطبایی

گروه آمار، دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۸۵/۱۰/۱۳ تاریخ آخرین بازنگری: ۱۳۸۶/۸/۱

**چکیده:** در این مقاله، با فرض اینکه در مدل رگرسیونی خطی چندگانه، بردار خطای تصادفی دارای توزیع  $t$  چند متغیره است، برآوردگرهای کمترین توانهای دوم تعمیم یافته، کمترین توانهای دوم تعمیم یافته مقید و تورنجش را برای بردار پارامتر مجهول مدل رگرسیونی بدست می آوریم. سپس با استفاده از تابع زیانهای مربعی و مربعی موزون، مخاطره برآوردگرهای بدست آمده را با یکدیگر مقایسه می کنیم و نشان می دهیم در شرایطی خاص کدامیک از برآوردگرها بر دیگری برتری دارند.

**واژه‌های کلیدی:** کمترین توانهای دوم تعمیم یافته، کمترین توانهای دوم تعمیم یافته مقید، تورنجش، توزیع ویشارت معکوس.

۱ مقدمه

مدل خطی

$$y = X\beta + e \quad (1)$$

آدرس الکترونیک مسئول مقاله: محمد آرشی، m\_arashi\_stat@yahoo.com

۹۶..... مقایسه مخاطره انواع برآوردگرها در مدل رگرسیون چندگانه با خطای  $t$  چندگانه

را در نظر بگیرید، که در آن  $y$  بردار تصادفی  $n$  بعدی مقادیر پاسخ،  $X$  یک ماتریس غیر تصادفی  $(n \times p)$  با رتبه کامل  $p$ ،  $\beta = (\beta_1, \dots, \beta_p)'$  بردار ضرایب رگرسیونی و  $e$  بردار خطای تصادفی  $n$  بعدی می باشد. معمولاً وقتی رگرسیون کمترین توانهای دوم را با استفاده از  $n$  مشاهده، برای  $p$  متغیر تحت مدل خطی (۱) در نظر می گیریم، فرض می کنیم توزیع خطاها نرمال است، در حالی که در عمل ممکن است این چنین نباشد. اگر تحلیل ما بر پایه این فرض باشد که ممکن است توزیع خطاها غیر نرمال باشد، آن گاه باید از روشهای رگرسیونی استوار<sup>۱</sup> و دقیقتری استفاده بکنیم. فیشر (۱۹۵۶) برای اولین بار به نتایج نامناسب استفاده از توزیع نرمال به عنوان توزیع خطای تصادفی در مدل‌های خطی آماری اشاره کرد. در این روش با قبول متقارن بودن توزیع خطاها، سعی می شود از توزیعهایی که منحنی تابع چگالی آنها دنباله‌هایی پهن تر از نرمال دارند استفاده بشود. یعنی از توزیعهایی استفاده می کنیم که احتمال در دنباله‌های توزیع بیشتر از حالت نرمال است تا بتوان نقاط فرین<sup>۲</sup> و نقاط پرت<sup>۳</sup> را تحت پوشش قرار داد. زلنر (۱۹۷۶) از توزیع  $t$  به عنوان مدل خطای تصادفی در روش‌های کلاسیک و بییزی استفاده کرد و نشان داد که توزیع نرمال حالت خاصی از توزیع  $t$  است. همچنین اولو و والش (۱۹۸۴) در بهبود انواع متفاوتی از آزمون‌هایی که در مطالعات اقتصادی با استفاده از مدل  $t$  بکار می روند، تحقیقات زیادی انجام داده‌اند. برای آگاهی بیشتر در خصوص اهمیت استفاده از توزیع  $t$  در روش‌های آماری دقیق و مدل‌ها می توان به لانگ و همکاران (۱۹۸۹) و تیکو و همکاران (۱۹۹۲) مراجعه کرد.

در این مقاله فرض می کنیم که بردار خطای تصادفی در مدل خطی (۱) دارای توزیع  $t$  چند متغیره (Mt) است و بر این اساس برآوردگر کمترین توانهای دوم تعمیم یافته<sup>۴</sup> (GLS) و برآوردگر کمترین توانهای دوم تعمیم یافته مقید<sup>۵</sup> (RGLS)

<sup>۱</sup> Robust regression methods

<sup>۲</sup> Extreme values

<sup>۳</sup> Outliers

<sup>۴</sup> Generalized least square

<sup>۵</sup> Restricted generalized least square

تحت یک فرضیه آماری در مورد بردار پارامتر  $\beta$  و برآوردگر تورنچس<sup>۶</sup> (S)، ترکیبی از دو برآوردگر GLS و RGLS را برای  $\beta$  بدست می آوریم و در نهایت مخاطره این سه برآوردگر را تحت تابع زیان مربعی خطا با یکدیگر مقایسه می کنیم. صالح (۲۰۰۶) ترکیبات مختلفی از برآوردگرهای فوق را تحت تئوری نرمال و در حالات مختلف ناپارامتری مورد بررسی قرار داده است.

توزیع Mt عضوی از کلاس توزیع های نرمال مرکب بیضی گون<sup>۷</sup> (ECND) است که دارای میانگین صفر و ماتریس واریانس-کوواریانس همیشگی مثبت  $\Sigma$  (p.d.) می باشد به عبارت دیگر داریم  $E(e) = 0$  و  $E(ee') = \Sigma$ . کلاس ECND زیر کلاسی از خانواده توزیع ها با منحنی های تراز بیضی گون<sup>۸</sup> (ECD) است که می توان آن را به عنوان آمیزه واریانس توزیع های نرمال<sup>۹</sup> به صورت

$$f(e) = \int \dots \int_{\Sigma > 0} f_n(e|\Sigma)g(\Sigma) d\Sigma, \quad (2)$$

بیان کرد، که در آن  $f(e)$  تابع چگالی احتمال (pdf) بردار  $e$ ،  $f_n(e|\Sigma)$  pdf توزیع نرمال چند متغیره با میانگین صفر و ماتریس واریانس-کوواریانس همیشگی مثبت  $\Sigma$  به صورت  $N_n(0, \Sigma)$  و  $g(\Sigma)$  تابع چگالی احتمال  $\Sigma$  برای  $\Sigma$  های p.d. می باشد. معروفترین عضوهای کلاس ECD، توزیع نرمال چند متغیره و توزیع Mt می باشند. برای آگاهی بیشتر در این زمینه و کاربرد این روش به موایرهد (۱۹۸۲) و گوپتا و وارگا (۱۹۹۳) مراجعه کنید.

به منظور ارائه دلیل استفاده از صورت شرطی فوق به جای استفاده مستقیم از توزیع Mt، دقت کنید به طور کلی تر، در شرایطی که لازم است برآوردگرهایی برای پارامترهای توزیع های در ECD بدست آوریم ممکن است در عمل نتوان مشاهداتی از توزیع های در ECD به طور مستقیم تولید کرد و یا اغلب در مطالعه صفات و پدیده های طبیعی با مواردی مواجه می شویم که توزیع مشاهدات بدست آمده نرمال

<sup>۶</sup> Shrinkage  
<sup>۷</sup> Elliptical compound normal distributions  
<sup>۸</sup> Elliptically contoured distributions  
<sup>۹</sup> Variance mixture of normal distributions

۹۸. . . . . مقایسه مخاطره انواع برآوردگرها در مدل رگرسیون چندگانه با خطای  $t$  چندگانه

می باشد. در این موارد ممکن است ساختار ماتریس واریانس-کوواریانس بیان کننده نوعی اطلاع بیشتر نسبت به مشاهدات بدست آمده باشد، که با استفاده از تکنیک شرطی در (۲) می توان با دو دسته مشاهدات، یکی از توزیع شرطی و دیگری از توزیع پیشین است بتوان پارامترها را برآورد کرد. همچنین محاسبه توابع مخاطره برآوردگرهای ذکر شده نیز ساده تر است. طباطبایی و همکاران (۲۰۰۴a و ۲۰۰۴b) در حالت ساده تری از روش شرطی در (۲)، با فرض این که  $e | \tau \sim N_n(0, \tau^2 I_n)$  و  $\tau^2$  دارای توزیع پیشین گامای معکوس با پارامتر  $\sigma^2$  است، که توزیع  $Mt$  با ماتریس واریانس-کوواریانس  $\sigma^2 I_n$  را نتیجه می دهد، برآوردگرهایی از نوع کمترین توانهای دوم معمولی، کمترین توانهای دوم معمولی مقید و انواع مختلفی از برآوردگرهای تورنجیده را بدست آورده اند. ما در این مقاله با تعمیم گامای معکوس به ویشارت معکوس برآوردگرهای موردنظر را می یابیم. برای این منظور با استفاده از (۲)، در لم اساسی زیر نشان می دهیم که اگر  $e | \Sigma \sim N_n(0, \Sigma)$  و همچنین  $\Sigma$  دارای توزیع پیشین ویشارت معکوس ( $W^{-1}$ ) باشد آنگاه بردار خطای تصادفی  $e$  دارای توزیع  $Mt$  می باشد. می گوئیم  $\Sigma$  دارای توزیع ویشارت معکوس با پارامتر مقیاس  $\Psi^{-1}$  و  $m$  درجه آزادی است و با  $\Sigma \sim W_n^{-1}(\Psi^{-1}, m)$  نشان می دهیم اگر pdf آن به صورت

$$g(\Sigma) = \frac{|\Psi|^{m/2} |\Sigma|^{(m+n+1)/2} e^{-tr(\Psi\Sigma^{-1})/2}}{2^{mn/2} \Gamma_n(m/2)}, \quad (3)$$

باشد، که در آن  $\Gamma_n(t) = \pi^{\frac{n(n-1)}{2}} \prod_{i=1}^n \Gamma[t - \frac{1}{2}(i-1)]$  تابع گامای چند متغیره است. همچنین می گوئیم  $e$  دارای توزیع  $Mt$  با پارامتر مقیاس  $\Psi$  و  $\nu$  درجه آزادی است و با  $e \sim Mt(\Psi, \nu)$  نشان می دهیم اگر pdf آن به صورت زیر باشد.

$$f(e) = \frac{\Gamma(\frac{\nu+n}{2}) |\Psi|^{-1/2}}{(\nu\pi)^{n/2} \Gamma(\frac{\nu}{2})} \left(1 + \frac{e' \Psi^{-1} e}{\nu}\right)^{-\frac{(\nu+n)}{2}}. \quad (4)$$

برای جزئیات بیشتر در رابطه با این توزیع و کاربردهای آن با استفاده از روش شرطی در (۲)، در حالاتی ساده تر، ناداراجا و کتز (۲۰۰۵) و یا کیبیریا و جردر (۲۰۰۶) را ببینید.

لم ۱: فرض کنید  $e | \Sigma \sim N_n(0, \nu\Sigma)$  و  $\Sigma \sim W_n^{-1}(\Psi^{-1}, \nu + n - 1)$ ؛ در این صورت بردار تصادفی  $e$ ، دارای توزیع حاشیه‌ای  $Mt$  با پارامترهای  $\Psi$  و  $\nu$  درجه آزادی است. برای اثبات به بخش ضمایم مراجعه کنید.

## ۲ برآورد بردار پارامتر $\beta$

در این بخش با فرض اینکه در مدل خطی (۱)، بردار خطای تصادفی دارای توزیع  $Mt$  است، ابتدا برآوردگر کمترین توانهای دوم تعمیم یافته برای بردار پارامتر  $\beta$  را بدست می‌آوریم. سپس تحت قید  $H\beta = h$ ، که در آن  $H$  ماتریس معلوم  $(q \times p)$  و رتبه کامل  $q$  و  $h$  بردار مقادیر از پیش تعیین شده با بعد  $q$  است، سعی می‌کنیم با استفاده از روش می‌نیمم سازی با استفاده از ضریب لاگرانژ، برآوردگر کمترین توانهای دوم تعمیم یافته مقید برای بردار پارامتر  $\beta$  بیابیم. در نهایت ترکیبی از دو برآوردگر ذکر شده را به عنوان برآوردگر انقباض معرفی می‌کنیم. این قیود خطی در اغلب مسائل کاربردی به صورت مقابله‌ها<sup>۱۰</sup> مورد استفاده قرار می‌گیرند. امروزه این گونه قیدهای خطی در انواع تصادفی و غیر تصادفی، در بهبود برآوردگرهای رگرسیون استفاده می‌شوند. با استفاده از برآوردگرهای محدود شده تحت قیود خطی می‌توان برآوردگرهای تورنجیده تعریف کرد که با وجود اریبی، دارای ریسک کمتری نسبت به برآوردگرهای محدود نشده هستند. برای آگاهی بیشتر در این زمینه به طباطبایی (۱۹۹۵)، کبیریا و صالح (۲۰۰۴) و صالح (۲۰۰۶) مراجعه کنید. برای این منظور، ابتدا فرض کنید دو شرط اول کلاسیک برای ماتریس مشاهدات  $X$  در (۱) برقرار باشد:

A1.  $X$  غیرتصادفی است.

A2.  $y$  برداری تصادفی است به طوری که با شرط  $\beta_0 = \beta$  و  $\Sigma_0 > 0$  داریم

$$E(y) = X\beta_0 \text{ و } Var(y) = \Sigma_0.$$

در این مقاله فرض می‌کنیم  $\Sigma \sim W_n^{-1}(\Psi^{-1}, m)$  و ابرپارامترهای توزیع پیشین  $(\Psi$  و  $m)$ ، همگی معلوم هستند. توجه به این نکته ضروری است که در مسائل

<sup>۱۰</sup> Contrasts

۱۰۰. . . . . مقایسه مخاطره انواع برآوردگرها در مدل رگرسیون چندگانه با خطای  $t$  چندگانه

کاربردی به خصوص روش‌های بیزی، در صورتی که  $\lambda$  یک توزیع پیشین معلوم داشته باشد و به اشتباه آن را در نظر نگیریم، ممکن است با مشکل بیش‌پراکنش<sup>۱۱</sup> در محاسبه انحرافات استاندارد برآوردگرها یا انجام آزمون فرضیه‌های آماری روبرو شویم. لازم به ذکر است در شرایطی که ممکن است پارامترهای توزیع پیشین مجهول باشند، اگر  $m$  معلوم و  $\Psi$  مجهول، همان‌طور که در مقدمه ذکر شد با تکرار نمونه‌گیری، می‌توان  $\Psi$  را برآورد کرد و در حالتی که هر دو ابرپارامتر  $m$  و  $\Psi$  مجهول هستند، با استفاده از روشهای ECM و ECME، می‌توان برآوردگرهایی برای پارامترهای توزیع پیشین بدست آورد. برای آگاهی بیشتر در این زمینه، لیو و روبین (۱۹۹۵) را ببینید.

حال برای بدست آوردن برآوردگر کمترین توانهای دوم  $\beta$ ، عبارت زیر را نسبت به  $\beta$  مینیمم می‌کنیم.

$$S_1(\beta, \lambda) = (y - X\beta)' \Psi^{-1} (y - X\beta).$$

که با حل معادله  $-2X'\Psi^{-1}(y - X\beta) = 0$  بر حسب  $\beta$ ، برآوردگر کمترین توانهای دوم تعمیم‌یافته  $\beta$  به صورت زیر بدست می‌آید.

$$\hat{\beta} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y. \quad (5)$$

حال فرض کنید اطلاعاتی بیشتر در مورد بردار پارامتر  $\beta$  به صورت  $H\beta = h$  که در آن  $H$  ماتریس معلوم  $(q \times p)$  و رتبه کامل  $q$  و  $h$  بردار مقادیر از پیش تعیین شده با  $q$  مولفه است، داریم. در این صورت برای یافتن برآوردگر کمترین توانهای دوم  $\beta$  تحت قید  $H\beta = h$  با استفاده از روش ضریب لاگرانژ عبارت زیر را نسبت به  $\beta$  و  $\lambda$  مینیمم می‌کنیم.

$$S_2(\beta, \lambda) = (y - X\beta)' \Psi^{-1} (y - X\beta) + 2\lambda'(H\beta - h).$$

با حل همزمان معادلات نرمال زیر

$$\begin{cases} -2X'\Psi^{-1}(y - X\beta) + 2H'\lambda = 0 \\ 2(H\beta - h) = 0 \end{cases}$$

<sup>۱۱</sup> Overdispersion

برآوردگر کمترین توانهای دوم تعمیم یافته مقید  $\beta$  به صورت زیر

$$\tilde{\beta} = \hat{\beta} - (X'\Psi^{-1}X)^{-1}H'[H(X'\Psi^{-1}X)^{-1}H']^{-1}(H\hat{\beta} - h). \quad (6)$$

بدست می آید. اکنون برآوردگر تورنچش  $\beta$  را با  $\beta^*$  نشان داده، به صورت

$$\beta^* = \tau\hat{\beta} + (1 - \tau)\tilde{\beta} \quad (7)$$

می نویسیم، که در آن  $\tau \in \mathbb{R}$ . توجه کنید در صورتی که  $\tau \in [0, 1]$ ،  $\beta^*$  ترکیب محدبی از دو برآوردگر  $\hat{\beta}$  و  $\tilde{\beta}$  است؛ و در غیر این صورت  $\beta^*$  ترکیب محدبی از دو برآوردگر  $\hat{\beta}$  و  $\tilde{\beta}$  نمی باشد و همواره داریم

$$\beta^* |_{(\tau=0)} = \tilde{\beta} \quad \beta^* |_{(\tau=1)} = \hat{\beta}$$

اهمیت برآوردگر تورنچش در برتری آن نسبت به دو برآوردگر دیگر می باشد که در حالت های مختلفی در این مقاله در مورد آن بحث شده است. انواع دیگری از برآوردگر تورنچش را می توان به طور جامع تر در کارهای طباطبایی (۱۹۹۵) و صالح (۲۰۰۶) دید. در تمامی انواع مختلف برآوردگرهای تورنچش، همواره ترکیبی از دو برآوردگر کمترین توانهای دوم و کمترین توانهای دوم مقید در نظر گرفته شده است، لذا یکی از دلایل بدست آوردن برآوردگر کمترین توانهای دوم مقید، استفاده از آن در ساخت برآوردگر تورنچش است. در اکثر موارد از برآوردگر کمترین توانهای دوم مقید استفاده می شود تا برآوردگری با مخاطره ای کمتر از مخاطره برآوردگر کمترین توانهای دوم به دست آید.

### ۳ محاسبه مخاطره برآوردگرها

در این بخش با استفاده از تابع زیان درجه دوم و تابع زیان درجه دوم موزون، مخاطره برآوردگرهای GLS، RGLS و S بدست آمده در بخش قبل را محاسبه می کنیم. حال تابع زیان درجه دوم

$$L(\bar{\beta}; \beta) = (\bar{\beta} - \beta)'(\bar{\beta} - \beta), \quad (8)$$

۱۰۲. . . . . مقایسه مخاطره انواع برآوردگرها در مدل رگرسیون چندگانه با خطای t چندگانه

را در نظر بگیرید، که در آن  $\bar{\beta}$  یک برآوردگر بردار پارامتر  $\beta$  می باشد. در این صورت تابع مخاطره درجه دوم با استفاده از (۸) به صورت زیر تعریف می شود.

$$R(\bar{\beta}; \beta) = E[(\bar{\beta} - \beta)'(\bar{\beta} - \beta)]. \quad (9)$$

حال با استفاده از (۹)، مخاطره برآوردگرهای بدست آمده در بخش قبل را محاسبه می کنیم. برای سهولت در نوشتن فرض کنید

$$G_1 = (X' \Psi^{-1} X)^{-1}, \quad G_2 = (H G_1 H')^{-1}.$$

با توجه به اینکه  $y | \Sigma \sim N_n(X\beta, \Sigma)$  می توان نتیجه گرفت

$$(\hat{\beta} - \beta) | \Sigma \sim N_p(0, \nu G_1 X' \Psi^{-1} \Sigma \Psi^{-1} X G_1). \quad (10)$$

با استفاده از  $E(\Sigma) = \Psi/(\nu - 2)$  (اندرسن، ۲۰۰۳)، (۵) و (۱۰) داریم

$$\begin{aligned} R(\hat{\beta}; \beta) &= EE[(\hat{\beta} - \beta)'(\hat{\beta} - \beta) | \Sigma] \\ &= \nu Etr(G_1 X' \Psi^{-1} \Sigma \Psi^{-1} X G_1) \\ &= \frac{\nu}{\nu - 2} tr(G_1). \end{aligned} \quad (11)$$

همچنین با استفاده از (۶)، (۱۱) و تغییر متغیر  $\delta = G_1 H' G_2 (H\beta - h)$  داریم

$$\begin{aligned} R(\tilde{\beta}; \beta) &= EE[(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta) | \Sigma] \\ &= EE\{[(I_p - G_1 H' G_2 H)(\hat{\beta} - \beta) \\ &\quad - \delta]'[(I_p - G_1 H' G_2 H)(\hat{\beta} - \beta) - \delta] | \Sigma\} \\ &= EE[(\hat{\beta} - \beta)'(I_p - G_1 H' G_2 H)'(I_p \\ &\quad - G_1 H' G_2 H)(\hat{\beta} - \beta) | \Sigma] + \delta' \delta \\ &= R(\hat{\beta}; \beta) - \frac{\nu}{\nu - 2} tr(G_1 H' G_2 H G_1) + \delta' \delta. \end{aligned} \quad (12)$$

برای بدست آوردن  $R(\beta^*; \beta)$  با استفاده از (۷) می توان نوشت

$$R(\beta^*; \beta) = EE[(\beta^* - \beta)'(\beta^* - \beta) | \Sigma]$$

$$\begin{aligned}
 &= EE[(\tau\hat{\beta} + (\lambda - \tau)\tilde{\beta} - \beta)'(\tau\hat{\beta} + (\lambda - \tau)\tilde{\beta} - \beta) | \Sigma] \\
 &= EE\{[(I_p - (\lambda - \tau)G_\lambda H'G_\lambda H)\hat{\beta} + (\lambda - \tau)G_\lambda H'G_\lambda h \\
 &\quad - \beta]'[(I_p - (\lambda - \tau)G_\lambda H'G_\lambda H)\hat{\beta} \\
 &\quad + (\lambda - \tau)G_\lambda H'G_\lambda h - \beta] | \Sigma\} \\
 &= R(\hat{\beta}; \beta) - \frac{\nu(\lambda - \tau)^\lambda}{\nu - \lambda} tr(G_\lambda H'G_\lambda HG_\lambda) + (\lambda - \tau)^\lambda \delta' \delta.
 \end{aligned}
 \tag{۱۳}$$

حال فرض کنید تابع زیان در (۸)، تابع زیان درجه دوم موزون بصورت

$$L(\bar{\beta}; \beta) = (\bar{\beta} - \beta)'W(\bar{\beta} - \beta), \tag{۱۴}$$

باشد، که در آن  $W$  ماتریس وزن درجه دوم معلوم و ناویژه<sup>۱۲</sup> با بعد  $p$  است. با قراردادن  $W = G_\lambda^{-1} = X'\Psi^{-1}X$  و استفاده از (۱۱)، (۱۲) و (۱۳) می توان نتیجه گرفت

$$R(\hat{\beta}; \beta) = \frac{\nu}{\nu - \lambda} tr(WG_\lambda) = \frac{\nu}{\nu - \lambda} tr(I_p) = \frac{\nu p}{\nu - \lambda}, \tag{۱۵}$$

$$\begin{aligned}
 R(\tilde{\beta}; \beta) &= R(\hat{\beta}; \beta) - \frac{\nu}{\nu - \lambda} tr(WG_\lambda H'G_\lambda HG_\lambda) + \delta'W\delta \\
 &= \frac{\nu(p - q)}{\nu - \lambda} + (H\beta - h)'G_\lambda(H\beta - h),
 \end{aligned}
 \tag{۱۶}$$

$$\begin{aligned}
 R(\beta^*; \beta) &= R(\hat{\beta}; \beta) - \frac{\nu(\lambda - \tau)^\lambda}{\nu - \lambda} tr(WG_\lambda H'G_\lambda HG_\lambda) \\
 &\quad + (\lambda - \tau)^\lambda \delta'W\delta \\
 &= \frac{\nu[p - q(\lambda - \tau)^\lambda]}{\nu - \lambda} \\
 &\quad + (\lambda - \tau)^\lambda (H\beta - h)'G_\lambda(H\beta - h).
 \end{aligned}
 \tag{۱۷}$$

<sup>۱۲</sup> Non-singular

### بحث و نتیجه گیری

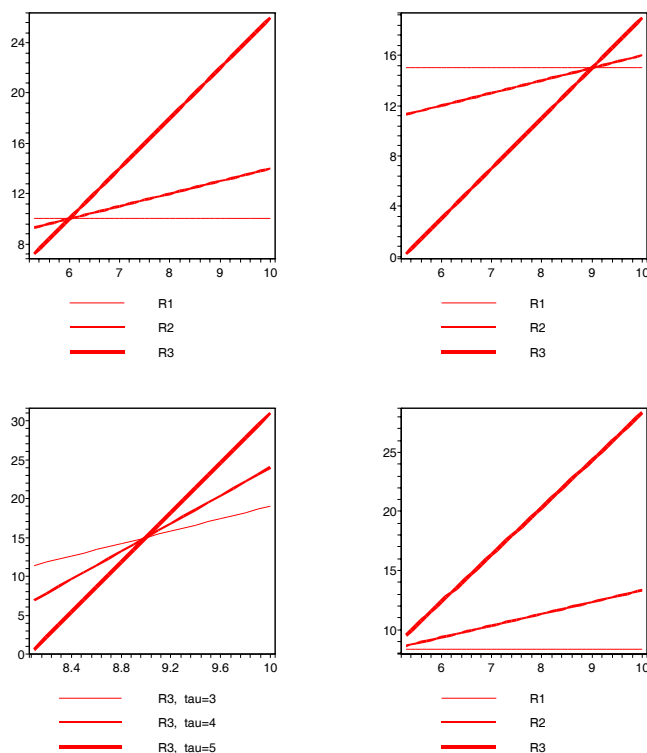
در این مقاله با فرض این که در مدل رگرسیون چندگانه، بردار خطای تصادفی دارای توزیع  $t$  چندمتغیره است، برآوردگرهایی از نوع کمترین توان‌های دوم تعمیم یافته بدست آوردیم. همچنین با استفاده از توابع زیان درجه دوم و زیان درجه دوم موزون مخاطره آن‌ها را محاسبه کردیم. برآوردگر تورنجش ترکیبی از دو برآوردگر محدود نشده و محدود شده تحت یک مجموعه قیود خطی است. معمولاً در مسائل کاربردی، در زمینه برآورد، همواره به دنبال برآوردگری هستیم که مخاطره آن کمترین مقدار ممکن را در بین بقیه برآوردگرها داشته باشد، لذا انتخاب نوع برآوردگر از اهمیت بسزایی برخوردار است. با توجه به ضریب تورنجش  $\tau$  در رابطه (۷)، می توان نوع برآوردگر  $\beta^*$  را طوری تعیین کرد که بهترین کارایی را نسبت به دو برآوردگر  $\hat{\beta}$  و  $\tilde{\beta}$  داشته باشد. در زیر نشان می دهیم که حتی در شرایطی انتخاب مقدار منفی برای  $\tau$  باعث برتری  $\beta^*$  نسبت به دو برآوردگر دیگر می شود. بدیهی است که تحت قید  $H\beta = h$  برای مقادیر  $\nu > 2$  براساس (۱۱)، (۱۲) و (۱۳) داریم

$$R(\tilde{\beta}; \beta) < R(\hat{\beta}; \beta),$$

$$R(\beta^*; \beta) < R(\hat{\beta}; \beta),$$

$$R(\beta^*; \beta) < R(\tilde{\beta}; \beta), \quad \forall \tau \in (-\infty, 0) \cup (2, \infty).$$

یعنی تحت قید  $H\beta = h$ ، دو برآوردگر  $\tilde{\beta}$  و  $\beta^*$  بر  $\hat{\beta}$  برتری دارند، که با  $\tilde{\beta} > \hat{\beta}$  و  $\beta^* > \hat{\beta}$  نشان می دهیم. به علاوه اگر  $\tau \in (-\infty, 0) \cup (2, \infty)$ ، آنگاه  $\beta^* > \tilde{\beta}$ . تعبیر این نتیجه در مسائل کاربردی بدین صورت است، در شرایطی که مجموعه قید خطی  $H\beta = h$  بر روی فضای پارامتر مدل رگرسیون چندگانه اعمال شده است، به ازای مقادیر منفی  $\tau$  یا مقادیر  $\tau \in (2, \infty)$ ، انتخاب برآوردگر تورنجش در (۷) نتیجه بهتری نسبت به انتخاب یکی از دو برآوردگر  $\hat{\beta}$  یا  $\tilde{\beta}$  دارد. با توجه به این که فرمول مخاطره برآوردگرها صورت بسته ای دارد، در عین حال به منظور نشان دادن صحت نتایج بدست آمده فوق (تحت زیان درجه دوم موزون)، توابع مخاطره را برای مقادیر مختلف  $p, q, \tau$  و  $\nu$  به روش نموداری با یکدیگر مقایسه شده، که در



شکل ۱:  $p$ ،  $q$  و  $\tau$  ثابت است و  $\nu$  افزایش می یابد.

نمودارهای شکل ۱ آمده است. در این نمودارها محور افقی نشان دهنده مقادیر  $\theta = (H\beta - h)'G\tau(H\beta - h)$  و محور عمودی مقادیر مخاطره

$$R_1 = R(\hat{\beta}; \beta), \quad R_2 = R(\tilde{\beta}; \beta), \quad R_3 = R(\beta^*; \beta)$$

است. بدیهی است که  $\theta$  یک صورت درجه دوم است و  $\theta \in [0, \infty)$ . انتخاب  $\theta$  به صورت فوق به این دلیل است که در بررسی روند برتری برآوردها نسبت به یکدیگر تنها فاصله  $H\beta$  از مقادیر از پیش تعیین شده  $h$  اهمیت دارد نه علامت آن، و همچنین به دلیل انتخاب زیان درجه دوم موزون، وزن  $G\tau$  در این فاصله تأثیر دارد. با توجه به نتایج بدست آمده از نمودارهای حاصل می توان نتیجه گرفت:

۱. با افزایش مقدار  $\tau$ ، ابتدا یک روند نزولی و سپس صعودی دارد. و هر چه

۱۰۶. . . . . مقایسه مخاطره انواع برآوردگرها در مدل رگرسیون چندگانه با خطای  $t$  چندگانه

مقدار  $\tau$  زیاد شود مینیمم نسبی  $R^3$  در مقادیر بزرگتر  $\theta$  رخ می دهد.  
 ۲. ابتدا روند برتری سه برآوردگر به صورت  $\hat{\beta} > \tilde{\beta} > \beta^*$  است و هر چه به مقدار  $\nu$  افزوده شود روند برتری به به  $\hat{\beta} > \tilde{\beta} > \beta^*$  گرایش پیدا می کند. یعنی انتخاب مقادیر کوچک  $\nu$  نتیجه بهتری در استفاده از برآوردگر تورنجش می دهد.

### ضمایم

اثبات لم ۱: از آنجایی که انتگرال pdf،  $\Sigma$  برای  $\Sigma > 0$  برابر ۱ است، می توان نتیجه گرفت

$$\int \dots \int_{\Sigma > 0} |\Sigma|^{-(\nu+2n)/2} \exp\left[\frac{-tr(\Psi\Sigma^{-1})}{2}\right] d\Sigma = (2^{-n}|\Psi|)^{-(\nu+n-1)/2} \times \Gamma_n\left(\frac{\nu+n-1}{2}\right).$$

با استفاده از فرع A.3.1 در اندرسن (۲۰۰۳) داریم

$$|ee' + \nu\Psi| = \nu^n |\Psi| \left(1 + \frac{e'\Psi^{-1}e}{\nu}\right),$$

با استفاده از (۲)، (۳) و (۴)، pdf بردار تصادفی  $e$  عبارتست از

$$\begin{aligned} f(e) &= \int \dots \int_{\Sigma > 0} f(e|\Sigma)g(\Sigma) d\Sigma \\ &= \frac{(2\pi)^{-n/2} \nu^{-n/2} |\Psi|^{(\nu+n-1)/2} \nu^{-n/2}}{2^n (\nu+n-1)^{n/2} \Gamma_n\left(\frac{\nu+n-1}{2}\right)} \\ &\quad \times \int \dots \int_{\Sigma > 0} |\Sigma|^{-(\nu+2n)/2} \exp\left\{\frac{-tr\left[\left(\frac{ee'}{\nu} + \Psi\right)\Sigma^{-1}\right]}{2}\right\} d\Sigma \\ &= \frac{(2\pi)^{-n/2} \nu^{-n/2} 2^n (\nu+n)^{n/2} \Gamma_n\left(\frac{\nu+n}{2}\right)}{2^n (\nu+n-1)^{n/2} \Gamma_n\left(\frac{\nu+n-1}{2}\right)} \\ &\quad \times |ee' + \nu\Psi|^{-(\nu+n)/2} |\Psi|^{(\nu+n-1)/2} \\ &= \frac{\Gamma\left(\frac{\nu+n}{2}\right) |\Psi|^{-1/2}}{[\nu\pi]^{n/2} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{e'\Psi^{-1}e}{\nu}\right)^{-(\nu+n)/2} \end{aligned}$$

□ بنابراین  $e$  دارای توزیع Mt با پارامتر  $\Psi$  و  $\nu$  درجه آزادی است.

### تقدیر و تشکر

نویسندگان مقاله از پیشنهادات و نظرات ارزنده داوران محترم که باعث اصلاحات سازنده‌ای در محتوا و ارائه بهتر مقاله شده‌است، کمال تشکر و قدردانی را دارند. از حمایت مالی قطب علمی داده‌های ترتیبی و فضایی نیز قدردانی می‌گردد.

### مراجع

- Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, 3rd Ed., John Wiley, New York.
- Fisher, R. A. (1956), *Statistical Methods in Scientific Inference*. Oliver and Boyd, London.
- Gupta, A. K. and Varga, T. (1993), *Elliptically Contoured Models in Statistics*, Kluwer Academic Pres London.
- Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989), *Robust Statistical Modeling using the t Distribution*, Journal of American Statistical Association, **84**, 881-896.
- Liu, Ch. and Rubin, D. B. (1995), ML Estimation of the t Distribution using EM and its Exrensions, ECM and ECME, *Statistica Sinica*, **5**, 19-39.
- Kibria, B. M. G. and Joarder, A. H. (2006), A Short Review of Multivariate t Distribution, *J. Statist. Res.*, **40**(1), 59-72.
- Kibria, B. M. Golam and Saleh, A. K. Md. E. (2004), Preliminary Test Ridge Regression Estimators with Student's t Errors and Conflicting Test Statistics, *Metrika*, **59**, 105-124.

۱۰۸. . . . . مقایسه مخاطره انواع برآوردگرها در مدل رگرسیون چندگانه با خطای  $t$  چندگانه

Muirhead, R. J. (1982), *Aspect of Multivariate Statistical Theory*, John Wiley, New York.

Nadarajah, S. and Kotz, S. (2005), *Mathematical Properties of the Multivariate  $t$  Distribution*, Acta Appl. Math. **89**, 53-84.

Saleh, A. K. Md. E. (2006), *Theory of Preliminary Test and Stein-type Estimation with Applications*, John Wiley, New York.

Tabatabaey, S. M. M., (1995), *Preliminary Test Approach Estimation: Regression Model with Spherically Symmetric Errors*, Ph.D. Thesis, Carleton University, Canada.

Tabatabaey, S. M. M., Saleh, A. K. Md. E. and Kibria, B. M. Golam, (2004a), *Estimation Strategies for Parameters of the Linear Regression Models with Spherically Symmetric Distributions*, J. Statist. Res., **38**(1), 13-31.

Tabatabaey, S. M. M., Saleh, A. K. Md. E. and Kibria, B. M. Golam, (2004b), *Simultaneous Estimation of Regression Parameters with Spherically Symmetric Errors under Possible Stochastic Constraints*, Int. J. Statist. Sci., **3**, 1-20.

Tiku, M. L., Tan, W. Y. and Balakrishnan, N. (1992), *Robust Inference*, Marcel Dekker, New York.

Ullah, A. and Walsh, V. Z. (1984), *On the Robustness of LM, LR and W Tests in Regression Models*, Econometrica, **52**, 1055-1066.

Zellner, A. (1976), *Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student- $t$  Error Term*, Journal of American Statistical Association, **66**, 601-616.