اندیشه آماری، بهار و تابستان ۱۳۹۱، شماره پیاپی ۳۳
سال هفدهم شماره اول، ص ۱-۱۴

توزیع‌های آمیخته- مقیاس نرمال و کاربرد آن‌ها در برازش مدل رگرسیون پانلی

ریحانه شکل‌آبادی [۱] ، ایرج کاظمی [۲]

چکیده:

یک روش مناسب در برازش مدل‌های رگرسیونی با هدف افزایش انعطاف‌پذیری بیشتر مدل در تحلیل انواع داده‌ها با ساختار پخش غیر نرمال، استفاده از کلاس توزیع‌های آمیخته- مقیاس نرمال است. ساختار این خانواده بر پایه نمایش سلسله‌مراتبی تصادفی است که در آن توزیع آمیختگی نقش اساسی را در ویژگی‌های خاص آماری این توزیع‌های آمیخته ایفا می‌کند. این نمایش منجر به استفاده ساده‌تر و کاربرد بهتر این توزیع‌ها در برازش مدل‌های پیچیده توسط رهیافت بیز خواهد شد. ما در این مقاله خانواده توزیع آمیخته- مقیاس نرمال و توزیع‌های منتج از آن را در برازش مدل‌های پانلی به منظور انجام تحلیل‌های مقاوم‌تر داده‌های وابسته معرفی می‌کنیم. همچنین توزیع‌های پسین شرطی کامل مورد نیاز جهت برآورد پارامترها با روش الگوریتم نمونه‌گیری گیبز را به‌دست می‌آوریم. در ادامه، مدل‌های معرفی شده را در تحلیل داده‌های بورس اوراق بهادار تهران بکار برده و با معیارهای انتخاب مدل آن‌ها را مقایسه می‌کنیم.

واژه‌های کلیدی: الگوریتم نمونه‌گیری گیبز، پسین شرطی کامل، داده‌های پانلی، نمایش سلسله مراتبی.

[۱] گروه آمار دانشگاه اصفهان
[۲] گروه آمار دانشگاه اصفهان

۱ مقدمه

در مدل‌های خطی با اثرات تصادفی توزیع هریک از مولفه‌های خطا و اثرهای غیرقابل مشاهده نرمال درنظر گرفته می‌شود. با این فرض برآورد پارامترهای مدل توسط روش ماکسیمم درستنمایی از دیدگاه فراوانی‌گرا انجام می‌شود. در کاربردهای تجربی به دلایلی مانند کشیدگی در توزیع داده‌ها منطقی است که کلاس کلی‌تری از توزیع‌ها در مدل‌سازی به‌کار گرفته شود. در این صورت استفاده از توزیع‌های دم سنگین به عنوان جایگزینی برای نرمال می‌تواند مناسب‌تر باشد. یکی از آن‌ها توزیع تی-استیودنت می‌باشد که در مباحث مدل‌سازی انعطاف‌پذیرتر و مقاوم‌تر از نرمال تشخیص داده شده است. در این راستا، یک کلاس بزرگ‌تر از توزیع‌ها پیرسن نوع هفتم [۷] است که تی-استیودنت را به عنوان حالت خاص دربردارد. اگرچه این توزیع‌ها، نرمال را به عنوان یک حالت خاص دربرمی‌گیرند و در برازش مدل‌های مختلف آماری انعطاف‌پذیرتر هستند اما انجام استنباط آماری پارامترها مشکل‌تر خواهد شد به این منظور محققان با توسعه این توزیع‌ها کلاس بزرگی را به نام آمیخته-مقیاس معرفی و ویژگی آماری آن‌ها را بیان کرده‌اند. از جمله می‌توان به ساختار سلسله‌مراتبی اشاره نمود که در قالب یک نمایش تصادفی بیان می‌شود [۳]. این موضوع نقش مهمی را در استفاده ساده‌تر از روش‌های محاسباتی بیزی، به ویژه نمونه‌گیری گیبز، ایفا می‌کند.

یکی از مباحث مهم در تحقیقات کاربردی تحلیل داده‌های وابسته و یا اندازه‌های مکرر در طول زمان است. از مدل‌های متداول مرتبط با تحلیل این داده‌ها، مدل‌های پانلی با اثرهای تصادفی است. از ویژگی‌های مهم این مدل‌ها افراز تغییرپذیری کل به بین و درون گروه مشاهدات است. همچنین برای کنترل تغییرپذیری، اثرهای خاص افراد را که قابل مشاهده نیستند تصادفی درنظر می‌گیرد که از یک توزیع آماری پیروی می‌کند. استنباط آماری پارامترهای مدل بر اساس تابع درستنمایی حاشیه‌ای است که با انتگرال‌گیری نسبت به اثرهای تصادفی به‌دست می‌آید[۴]. این درستنمایی در صورتی شکل صریح خواهد داشت که توزیع مولفه‌های خطا و اثرهای تصادفی نرمال باشند. در غیر این صورت، محاسبه برآوردهای ماکسیمم درستنمایی نیازمند به استفاده از روش‌های عددی تکراری پیشرفته و یا تقریب‌هایی برای محاسبه انتگرال‌های چندگانه است. با توجه به آن‌که عدم یافتن تقریب مناسب در مدل‌های پیچیده امکان‌پذیر است، بنابراین کاربرد روش‌های جایگزینی مبتنی بر رهیافت شبیه‌سازی مونت کارلوی زنجیر مارکف ($McMC$) از دیدگاه بیز مورد توجه قرار گرفته‌است.

قابل توجه است که برای یافتن برآورد بیز پارامترها به محاسبه توزیع‌های پسین حاشیه‌ای نیاز داریم.

محاسبه این توزیع‌ها حتی با درنظر گرفتن فرض نرمال مستلزم حل انتگرال‌های پیچیده چندگانه است. ما در این مقاله برای مدل‌های پانلی توزیع مشاهده‌ها به شرط اثرهای تصادفی را آمیخته-مقیاس نرمال درنظر می‌گیریم. همچنین این توزیع‌ها را نیز برای اثرها فرض خواهیم کرد. با توجه به آن‌که توزیع‌های فوق نمایش سلسله مراتبی دارند، پس مدل حاصل یک مدل سلسله مراتبی بیزی خواهد شد که امکان به کار بردن الگوریتم نمونه‌گیری گیبز [۵] را که از روش‌های مبتنی بر $McMC$ است فراهم می‌کند. این الگوریتم بر اساس فرآیند مارکف نمونه‌هایی را از توزیع‌های شرطی کامل پارامترها تولید می‌کند که این منحصراً توزیع پسین توأم و همه‌ی توزیع‌های پسین حاشیه‌ای را تعیین می‌کند. پس استنباط بیزی بر اساس پسین‌های حاشیه‌ای خواهند بود.

ساختار این مقاله به این صورت است که در بخش دوم و سوم کلاس توزیع‌های آمیخته-مقیاس نرمال در حالت یک و چندمتغیره معرفی می‌شود. سپس چندین توزیع از این کلاس همراه با ویژگی‌های مرتبط با آن‌ها ارائه می‌شود. در بخش چهارم مدل رگرسیون پانلی را برای داده‌های وابسته با توزیع‌های آمیخته-مقیاس معرفی می‌کنیم. همچنین با درنظرگرفتن ساختار سلسله‌مراتبی بیزی توزیع‌های پسین شرطی کامل را جهت به‌کارگیری الگوریتم گیبز محاسبه می‌کنیم. در بخش پنجم انواع مدل‌های ارائه شده را بر داده‌های بازار بورس اوراق بهادار تهران برازش داده و با معیارهای متداول انتخاب مدل بهترین را که برازنده است برگزیده‌ایم.

۲ خانواده توزیع آمیخته-مقیاس نرمال یک متغیره

فرض کنید Z متغیر تصادفی استاندارد و λ یک متغیر تصادفی مثبت و مستقل از Z باشد. برای اولین بار آندره و مالوز (۱۹۷۴) با تعریف نمایش تصادفی $Y = Z\lambda$ توزیع آمیخته-مقیاس متغیر تصادفی نرمال [۳] (SMN) را معرفی کردند. در حالت کلی، این توزیع به صورت زیر است.

تعریف ۱.۲. *اگر Y یک متغیر تصادفی پیوسته با پارامتر مکان μ و پارامتر مقیاس σ باشد تابع چگالی آمیخته-مقیاس نرمال به صورت*

$$f(y|\mu,\sigma) = \int_0^{\infty} N(y|\mu, k(\lambda)\sigma^2)\pi(\lambda)d\lambda \quad (1)$$

تعریف می‌شود که در آن $N(y|.,.)$ تابع چگالی نرمال، $k(.)$ تابعی مثبت و $\pi(\lambda)$ تابع چگالی پارامتر آمیختگی λ روی فضای $\mathbb{R}^+$ است. این توزیع را با نماد $SMN(\mu, k(\lambda)\sigma^2)$ نشان می‌دهیم.

لم ۲.۲. *اگر Z دارای توزیع نرمال استاندارد و λ پارامتر آمیختگی باشد که مستقل از Z است*

[۳] Scale Mixture of Normal

آن‌گاه مقدار کشیدگی متغیر نرمال آمیخته-مقیاس $Y = Z\lambda$ کمتر از کشیدگی متغیر نرمال استاندارد Z نیست.

اثبات. با توجه به آن‌که امید ریاضی Y صفر است میزان کشیدگی را به صورت زیر داریم:

$$\begin{aligned} kur(Y) &= \frac{E(Y^4)}{[E(Y^2)]^2} = \frac{E(Z^4\lambda^4)}{[E(Z^2\lambda^2)]^2} \\ &= \frac{E(Z^4)E(\lambda^4)}{[E(Z^2)]^2[E(\lambda^2)]^2} \\ &\geq \frac{E(Z^4)}{[E(Z^2)]^2} = 3 \end{aligned} \qquad (۲)$$

از آنجا که واریانس λ^2 همواره نامنفی است، کسر $\frac{E(\lambda^4)}{[E(\lambda^2)]^2}$ همواره بزرگتر و مساوی یک است. بنابراین نامساوی فوق برقرار است. □

یکی دیگر از خواص توزیع‌های SMN متقارن بودن آن‌ها حول μ است، همچنین دم‌های سنگین‌تری نسبت به نرمال دارند. به عبارت دیگر در توزیع‌های SMN تجمع احتمال در دم‌ها بیشتر از توزیع نرمال است، یعنی در این توزیع‌ها احتمال مقادیر بزرگتر از ۳، بیشتر از این احتمال در توزیع نرمال استاندارد است (شکل ۱ را ببینید). این ویژگی خاص امکان مدل‌سازی مشاهدات با نقاط غایی را بهتر فراهم می‌کند. علاوه بر آن، اگر تابع $k(\lambda)$ به ازای تمام مقادیر $\lambda > 0$ از یک بیشتر باشد آن‌گاه واریانس در این توزیع‌ها بیشتر از توزیع نرمال است.

لم ۳.۲. *اگر متغیر تصادفی Y دارای توزیع $SMN(\mu, k(\lambda)\sigma^2)$ باشد آن‌گاه در صورت وجود داریم*

$$E(Y) = \mu\ ;\ \ var(Y) = \sigma^2 E(k(\lambda)) \qquad (۳)$$

اثبات. با استفاده از قوانین امید ریاضی مکرر و تفکیک واریانس به سادگی اثبات می‌شود. □

حال تعدادی از توزیع‌های عضو این کلاس را با ویژگی آن‌ها معرفی می‌کنیم.

مدل ۲-۱: توزیع تی-استیودنت. اگر در رابطه (۱) قرار دهیم

$$\begin{aligned} k(\lambda) &= \frac{1}{\lambda}\,, \\ \lambda &\sim Gam(\lambda|\frac{a}{2}, \frac{a}{2}) \end{aligned}$$

آن‌گاه توزیع تی-استیودنت با پارامتر مکان μ و پارامتر مقیاس σ و درجه آزادی a به صورت

$$f(y|\mu,\sigma,a) = \int_0^\infty N(y|\mu, \frac{\sigma^2}{\lambda})Gam(\lambda|\frac{a}{2},\frac{a}{2})d\lambda \qquad (۴)$$

به‌دست می‌آید که در آن $Gam(.|a,b)$ تابع چگالی گاما با هسته $\lambda^{a-1}e^{-b\lambda}$ برای $\lambda, a, b > 0$ است. حاصل انتگرال فوق به صورت

$$f(y,\mu,\sigma,a) = \frac{1}{Beta(\frac{1}{2},\frac{a}{2})\sqrt{a}\sigma} \times \left[1+\frac{1}{a}\left(\frac{x-a}{\sigma}\right)^2\right]^{\frac{a+1}{2}} \qquad (۵)$$

است که تابع چگالی تی-استیودنت می‌باشد. چگالی آمیخته (۱) بیان‌گر آن است که توزیع تی را می‌توان به صورت سلسله مراتبی

$$Y|\lambda \sim N\left(\mu, \frac{\sigma^2}{\lambda}\right),$$
$$\lambda \sim Gam(\lambda|\frac{a}{2}, \frac{a}{2}) \qquad (۶)$$

نوشت. میانگین و واریانس این توزیع با استفاده از رابطه (۳) به ترتیب μ و $\frac{\sigma^2 a}{a-2}$ به ازای $a > 2$ محاسبه می‌شود. همچنین با استفاده از قوانین امید ریاضی مکرر میزان کشیدگی این توزیع $\frac{3(a-2)}{(a-4)}$ به ازای $a > 4$ به‌دست می‌آید که این مقدار از میزان کشیدگی نرمال که برابر با ۳ است، بیشتر است.

مدل ۲-۲: خانواده توزیع پیرسن نوع هفتم. اگر در رابطه (۱) قرار دهیم

$$k(\lambda) = \frac{1}{\lambda},$$
$$\lambda \sim Gam(\lambda|\frac{a}{2}, \frac{a}{2})$$

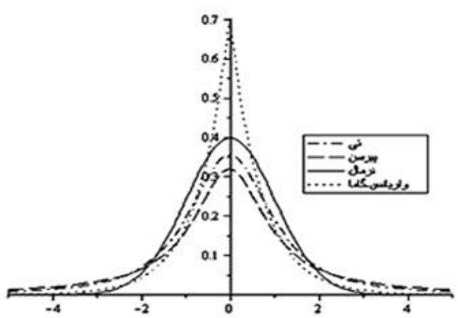
آن‌گاه توزیع پیرسن نوع هفتم به‌دست می‌آید[۷]. میانگین و واریانس این توزیع با استفاده از رابطه (۳) به ترتیب μ و $\frac{\sigma^2 b}{a-2}$ با قید $a > 2$ محاسبه می‌شود. با انجام چند محاسبه جبری می‌توان نشان داد که کشیدگی این توزیع با قید $a > 4$ برابر با $\frac{3(a-2)}{a-4}$ است. اگر $a = b$ آن‌گاه توزیع تی-استیودنت با درجه آزادی a حاصل می‌شود. یک حالت خاص آن توزیع کوشی است که a و b را یک قرار دهیم.

مدل ۲-۳: توزیع واریانس-گاما. اگر در رابطه (۱) قرار دهیم

$$k(\lambda) = \frac{1}{\lambda},$$
$$\lambda \sim InvGam(\lambda|\frac{a}{2}, \frac{a}{2})$$

آن‌گاه توزیع واریانس-گاما (VG) حاصل می‌شود. با انجام چند محاسبه جبری میانگین، واریانس و کشیدگی این توزیع به ترتیب برابر با μ، σ^2 و $\frac{a+2}{a}$ به‌دست می‌آیند. اگر در توزیع VG قرار دهیم $a = 2$ آن‌گاه توزیع لاپلاس حاصل می‌شود.

به منظور مقایسه با توزیع نرمال چند تابع چگالی



شکل ۱: مقایسه چند تابع چگالی متعلق به خانواده SMN با چگالی نرمال

متعلق به خانواده SMN در شکل (۱) آورده شده است. ملاحظه می‌شود که میزان کشیدگی و دم سنگین بودن این توزیع‌ها متفاوت است.

۳ خانواده توزیع آمیخته-مقیاس نرمال چند متغیره

توزیع‌های آمیخته-مقیاس را به حالت چندمتغیره به صورت زیر می‌توان تعمیم داد.

تعریف ۱.۳. *فرض کنید بردار تصادفی X دارای توزیع نرمال T متغیره با بردار میانگین صفر و ماتریس کواریانس معین مثبت Σ است و متغیر تصادفی مثبت λ مستقل از X با چگالی آمیختگی $\pi(\lambda)$ است. در این صورت، بردار تصادفی $Y = \mu + k(\lambda)^{1/2}$ دارای توزیع آمیخته-مقیاس نرمال T متغیره است. نمایش سلسله مراتبی*

$$Y|\mu, \Sigma, \lambda \sim N_T(y|\mu, k(\lambda)\Sigma)$$

را که در آن $\lambda \sim \pi(\lambda)$ برای این توزیع به کار می‌بریم [۸].

مشابه با حالت یک متغیره، اگر $k(\lambda) = 1$ و $\pi(\lambda)$ تباهیده در صفر باشد، آن‌گاه توزیع نرمال چندمتغیره و اگر $k(\lambda) = \frac{1}{\lambda}$ و $\lambda \sim Ga(\lambda|\frac{a}{2}, \frac{a}{2})$ آن‌گاه توزیع تی چندمتغیره با درجه آزادی a به‌دست می‌آید.

به دلیل یکتایی توابع مشخصه بسیاری از خواص توزیع‌ها توسط این تابع ثابت می‌شوند. در قضیه زیر تابع مشخصه توزیع‌های آمیخته-مقیاس نرمال چندمتغیره را به‌دست آورده‌ایم.

قضیه ۲.۳. *در صورتی که بردار تصادفی T بعدی Y دارای توزیع آمیخته-مقیاس نرمال چندمتغیره‌ی ($SMMN$) [۴] باشد آن‌گاه تابع مشخصه Y به صورت زیر است:*

$$\psi_Y(t) = exp\{it'\mu\} \times E_\lambda(exp\{-\frac{1}{2}t'k(\lambda)\Sigma t\}) \quad (۷)$$

اثبات. بردار تصادفی Y به شرط λ دارای توزیع نرمال $N_T(\mu, k(\lambda)\Sigma)$ است پس داریم

$$\psi_{Y|\lambda}(t) = exp(it'\mu - \frac{1}{2}t'k(\lambda)\Sigma t) \quad (۸)$$

با استفاده از قانون امید ریاضی مکرر تابع مشخصه Y به صورت زیر محاسبه می‌شود:

$$\begin{aligned}\psi_Y(t) &= E(exp\{it'Y\}) \\ &= E_\lambda[E_Y(exp\{it'Y\}|\lambda)] \\ &= exp\{it'\mu\} \\ &\quad \times E_\lambda\left(exp\left\{-\tfrac{1}{2}t'k(\lambda)\Sigma t\right\}\right)\end{aligned}$$

برای اثبات دیگر این قضیه به [۸] مراجعه شود.

□

قضیه ۳.۳. *اگر بردار تصادفی Y دارای توزیع آمیخته-مقیاس نرمال T متغیره با پارامتر مکان μ و پارامتر مقیاس Σ و متغیر آمیختگی λ باشد آن‌گاه $AY + b$ دارای توزیع $SMMN$ با پارامتر مکان $A\mu+b$ و پارامتر مقیاس $A\Sigma A'$ و متغیر آمیختگی λ است، که در آن A ماتریس $m \times T$ با رتبه $m \leq T$ و b برداری m بعدی است. به عبارت دیگر توزیع $SMMN$ نسبت به ترکیب خطی پایاست.*

[۴] Scale Mixture of Multivariate Normal

اثبات. چون Y دارای توزیع $SMMN$ است پس $Y|\lambda \sim N_T(\mu, k(\lambda)\Sigma)$. با استفاده از قاعده امید ریاضی مکرر و تعریف تابع مشخصه داریم

$$\begin{aligned}
\psi_{AY+b}(t) &= exp\{it'(AY+b)\} \\
&= E_\lambda[E_Y(exp\{it'(AY+b)\}|\lambda)] \\
&= exp\{it'b\}E_\lambda[E_Y(exp\{it'AY\}|\lambda)] \\
&= exp\{it'(A\mu+b)\} \\
&\quad E_\lambda\left(exp\left\{-\tfrac{1}{2}t'Ak(\lambda)\Sigma A't\right\}\right)
\end{aligned}$$

با استفاده از قضیه ۲.۳ و خاصیت یکتایی توابع مشخصه قضیه اثبات می‌شود. □

۴ مدل رگرسیونی پانلی

مجموعه داده‌های پانلی شامل مشاهداتی برای چندین واحد آزمایشی می‌باشند که در طی زمان‌های مختلف جمع‌آوری شده‌اند. تحلیل داده‌های پانلی را می‌توان در مواردی که مشاهدات به صورت ترکیبی از سری‌های زمانی و مقطعی هستند به کار برد. یک مدل خطی برای داده‌های پانلی به صورت

$$\begin{aligned}
Y_{it} &= X'_{it}\beta + \alpha_i + \epsilon_{it}, \qquad (۹) \\
& i=1,\dots,n; \;\; t=1,\dots,T
\end{aligned}$$

معرفی می‌شود[۴]، که در آن Y_{it} متغیر پاسخ برای واحد i-ام در زمان t، β بردار $p \times 1$ از ضرایب رگرسیون، X_{it} بردار $p \times 1$ از مشاهدات مربوط به متغیرهای توضیحی با $X_{it1} = 1$ و ϵ_{it} مانده‌های تصادفی مستقل و هم‌توزیع می‌شوند. α_iها اثرات خاص واحدهای آزمایشی هستند که تصادفی و مستقل از متغیرهای توضیحی و مولفه‌های خطا فرض می‌شوند. وجود اثرهای تصادفی در مدل‌های خطی باعث تبیین تغییرپذیری در مشاهدات و تعدیل ساختار همبستگی بین مشاهدات مربوط به هر واحد آزمایشی شده و این امر موجب افزایش دقت و کارایی برآورد پارامترها می‌گردد. پس مدل با اثرهای تصادفی در بسیاری از تحلیل داده‌های پانلی، بر اثر ثابت ارجحیت دارد. این موضوع به خصوص وقتی n بزرگ است و برآورد پارامترها در اثر ثابت سازگار نخواهند شد مشخص است. زیرا با افزایش n تعداد پارامترهای α_i نیز افزایش می‌یابند. در مدل با اثر تصادفی لازم است که توزیعی مناسب برای α_iها درنظر گرفته شود تا برآورد پارامترها معقول و پیش‌بینی اثرهای α_i به درستی انجام گیرد. با درنظر گرفتن Y_i به صورت

$$Y_i = (Y_{i1}, Y_{i2}, ..., Y_{it})'$$

می‌توان مدل (۹) را برای آزمودنی i-ام به صورت

$$\begin{aligned}
Y_i &= 7X_i\beta + \alpha_i e_T + \epsilon_i\,, \qquad (۱۰) \\
& i=1,\dots,n, \;\; t=1,\dots,T
\end{aligned}$$

بازنویسی کرد، که در آن e_T بردار یکه با بعد T است. در این مدل فرض معمول آن است که α_iها مستقل و دارای توزیع نرمال به صورت

$$\alpha_i \overset{\text{iid}}{\sim} N(0, \sigma^2_\alpha) \quad , \quad \epsilon_i \overset{\text{iid}}{\sim} N_T(0, \sigma^2_\epsilon I)$$

باشند. در تحلیل داده‌های واقعی ممکن است تعدادی از مانده‌ها و یا اثرهای تصادفی نسبت به مرکزشان دورافتاده باشند و فرض نرمال بودن آن‌ها منجر به نادیده گرفتن آن‌ها شود. این در صورتی است که یکی از اهداف اساسی کاربرد مدل‌های پانلی شناسایی عملکردهای بالقوه شدید است که این منجر به شناسایی واحدهایی می‌شود که از نظر عملکرد در رده‌های خیلی بالا و یا خیلی پایین قرار دارند. شناسایی این واحدها در تحقیقات کاربردی از اهمیت ویژه‌ای برخوردار است. ما در این مقاله استنباط بیزی مدل‌های پانلی را با استفاده از توزیع‌های آمیخته-مقیاس نرمال (چندمتغیره) که در برابر مشاهدات غایی مقاوم‌تر از نرمال هستند، پیشنهاد می‌کنیم.

۱.۴ تحلیل بیزی مدل‌های پانلی با اثرهای تصادفی و مانده‌های آمیخته-مقیاس نرمال چندمتغیره

مدل (۱۰) را با فرض اینکه مانده‌ها و اثرهای تصادفی دارای توزیع آمیخته-مقیاس نرمال (چندمتغیره) هستند، درنظر بگیرید. با توجه به اینکه خانواده توزیع $SMMN$ دارای ساختار سلسله مراتبی به صورت

$$\begin{aligned} Y|\mu,\Sigma,\lambda &\sim N_T(y|\mu,k(\lambda)\Sigma)\,, \\ \lambda &\sim \pi(\lambda) \end{aligned} \qquad (۱۱)$$

است، مدل پانلی را می‌توان به صورت نمایش سلسله مراتبی

$$\begin{aligned} &\alpha_i \overset{\text{ind}}{\sim} N(0,\sigma_\alpha^2 k(\lambda_i)), \quad \lambda_i \sim \pi(\lambda_i) \\ &\epsilon_i \overset{\text{ind}}{\sim} N_T(0,\sigma_\epsilon^2 k(w_i)I), \quad w_i \sim \pi(w_i) \\ &Y_i|\alpha_i \overset{\text{ind}}{\sim} N_T(X_i\beta+\alpha_i e_T,\sigma_\epsilon^2 k(w_i)I) \end{aligned}$$

درنظر گرفت. این ساختار در روش‌های محاسباتی پیشرفته برای انجام استنباط بیزی پارامترهای مدل که مبتنی بر رهیافت $McMC$ از جمله الگوریتم نمونه‌گیری گیبز و متروپولیس-هستینگز هستند، مفید می‌باشد. با انتخاب چگالی‌های پیشین مناسب برای پارامترهای مدل می‌توان برآوردهای بیزی پارامترها را با استفاده از نمونه‌گیری گیبز به‌دست آورد. معمولاً درنظر گرفتن توزیع‌های پیشین مزدوج محاسبات بیزی را راحت‌تر می‌کند[۶]. پیشین‌های مزدوج زیر را برای اثرهای ثابت β و مولفه‌های واریانس مانده‌ها و واریانس اثرهای تصادفی درنظر می‌گیریم

$$\begin{aligned} \beta &\sim N_P(\beta_., \Sigma_.)\,, \\ \sigma_\epsilon^2 &\sim InvGam\left(\frac{v_\epsilon}{2}\frac{v_\epsilon}{2}\right), \\ \sigma_\alpha^2 &\sim InvGam\left(\frac{v_\alpha}{2}\frac{v_\alpha}{2}\right) \end{aligned}$$

که در آن پارامترهای β_0، Σ_0، v_α و v_ϵ معلوم هستند. چگالی پسین توأم مناسب با حاصل ضرب تابع درستنمایی و چگالی پیشین توأم است. لذا با فرض مستقل بودن پارامترها و بکارگیری پیشین‌های

فوق توزیع پسین توأم به صورت

$$\begin{aligned} & f(\alpha,\beta,\sigma^2_\epsilon,\sigma^2_\alpha,w,\lambda|y,x) \\ \propto\; & f(y|\alpha,\beta,\sigma^2_\epsilon,w)\times\pi(w) \\ \times\; & f(\alpha|\sigma^2_\alpha,\lambda)\times\pi(\lambda) \\ \times\; & \pi(\beta)\times\pi(\sigma^2_\epsilon)\times\pi(\sigma^2_\alpha) \end{aligned} \quad (۱۲)$$

محاسبه می‌شود که در آن $\alpha'=(\alpha_1,...,\alpha_n)$، $w'=(w_1,...,w_k)$، $\lambda'=(\lambda_1,...,\lambda_k)$ و y مجموعه داده‌ها است. برای به‌دست آوردن برآورد بیزی پارامترها توسط روش نمونه‌گیری گیبز به محاسبه توزیع‌های پسین کامل هر پارامتر به شرط سایر پارامترها نیاز است. با انجام محاسبات جبری این توزیع‌هارا به صورت زیر به‌دست می‌آوریم.

1. $\beta|\alpha,\sigma^2_\epsilon,\sigma^2_\alpha,w,y\sim N_P(\beta|\hat{\beta},\Omega)$

2. $\sigma^2_\epsilon|\beta,\alpha,\sigma^2_\alpha,w,y\sim$
$$InvGam(\sigma^2_\epsilon\left|\frac{N+v_\epsilon}{2}\right.,\frac{1}{2}(\sum_i(\frac{\epsilon'_i\epsilon_i}{k(w_i)})+v_\epsilon))$$

3. $\sigma^2_\alpha|\beta,\alpha,\sigma^2_\epsilon,\lambda,y\sim$
$$InvGam(\sigma^2_\alpha\left|\frac{v_\alpha+n}{2}\right.,\frac{1}{2}(\sum_i(\frac{\alpha_i^2}{k(\lambda_i)})+v_\alpha))$$

4. $\alpha_i|\alpha_{(-i)},\beta,\sigma^2_\epsilon,\sigma^2_\alpha,\lambda_i,w_i,y\sim$
$$N(\alpha_i|V_i\bar{r}_i,(1-V_i)\sigma^2_\alpha k(\lambda_i))$$

5. $f(\lambda_i|\lambda_{(-i)},\alpha_i,\sigma^2_\alpha,y)\propto$
$$\frac{1}{\sqrt{k(\lambda_i)}}exp[-\frac{1}{2\sigma^2_\alpha k(\lambda_i)}\alpha_i^2]\times\pi(\lambda_i)$$

6. $f(w_i|w_{(-i)},\alpha_i,\beta,\sigma^2_\epsilon,y)\propto$
$$(\frac{1}{k(w_i)})^{\frac{T}{2}}exp[-\frac{1}{2}\sigma^2_\epsilon k(w_i)\epsilon'_i\epsilon_i]\times\pi(w_i)$$

که در آن $\alpha_{(i)}$، $\lambda_{(-i)}$ و $w_{(i)}$ بردارهای مرتبط با حذف i - امین عضو هستند و

$$\begin{aligned} \Omega &= (\sum_i\left(\frac{x'_ix_i}{\sigma^2_\epsilon k(w_i)}\right)+\Sigma_0^{-1})^{-1}, \\ \hat{\beta} &= \Omega(\sum_i\left(\frac{x'_iu_i}{\sigma^2_\epsilon k(w_i)}\right)+\Sigma_0^{-1}\beta_0), \\ r_i &= y_i-X_i\beta, \\ \bar{r}_i &= \frac{1}{T}r'_ie_T, \\ V_i &= \frac{T\sigma^2_\alpha k(\lambda_i)}{T\sigma^2_\alpha k(\lambda_i)+\sigma^2_\epsilon k(w_i)}, \\ \epsilon_i &= (Y_i-X_i\beta-\alpha_ie_T). \end{aligned}$$

الگوریتم گیبز برای یافتن برآورد بیز پارمترهای مدل به صورت زیر انجام می‌شود.
ابتدا بردار $(\alpha^{(0)},\beta^{(0)},\sigma^{2^{(0)}}_\epsilon,\sigma^{2^{(0)}}_\alpha,w^{(0)},\lambda^{(0)})$ را به عنوان مقادیر اولیه پارامترهای مدل درنظر می‌گیریم.
سپس مقادیر $(\alpha^{(l)},\beta^{(l)},\sigma^{2^{(l)}}_\epsilon,\sigma^{2^{(l)}}_\alpha,w^{(l)},\lambda^{(l)})$، $l=1,...,L$ را از توزیع‌های پسین شرطی کامل تا مرحله همگرایی تولید می‌کنیم، که در آن L یک مقدار بزرگ است. در مرحله آخر میانگین مقادیر شبیه‌سازی شده، به عنوان برآورد بیز پارامترها خواهند بود.

برای توزیع‌های مختلف عضو خانواده $SMMN$ با توجه به نوع تابع $k(\lambda)$ و $k(w)$ و نیز چگالی‌های آمیختگی $\pi(\lambda)$ و $\pi(w)$ عبارت‌های ۵ و ۶ را می‌توان به ترتیب بر حسب λ و w ساده نمود. اگر

شکل شناخته شده‌ای برای توابع چگالی حاصل شود آن‌گاه از الگوریتم گیبز برای برآورد پارامترها استفاده می‌نماییم در غیر اینصورت الگوریتم متروپلیس-هستینگز جایگزین خواهد شد.

به عنوان مثال فرض کنید مانده‌ها دارای توزیع واریانس-گاما و اثرهای تصادفی پیرسن نوع هفتم باشند. در این صورت داریم

$$\lambda_i \sim Gam\left(\frac{a}{2},\frac{b}{2}\right), k(\lambda_i)=\frac{1}{\lambda_i},$$

$$w_i \sim InvGam\left(\frac{v}{2},\frac{v}{2}\right), k(w_i)=\frac{1}{w_i}$$

که با انجام چند محاسبه جبری پیچیده چگالی‌های پسین شرطی کامل به صورت صریح زیر خواهند بود:

1. $\beta|\alpha,\sigma^2_\epsilon,\sigma^2_\alpha,w,y,x \sim N_P(\beta|\hat{\beta},\Omega)$

2. $\sigma^2_\epsilon|\beta,\alpha,\sigma^2_\alpha,w,y \sim$
$InvGam(\sigma^2_\epsilon \left| \frac{N+v_\epsilon}{2}\right., \frac{1}{2}(\sum_i(\epsilon'_i\epsilon_i w_i)+v_\epsilon))$

3. $\sigma^2_\alpha|\beta,\alpha,\sigma^2_\epsilon,\lambda,y \sim$
$InvGam(\sigma^2_\alpha \left| \frac{v_\alpha+n}{2}\right., \frac{1}{2}(\sum_i \lambda_i\alpha_i^2+v_\alpha))$

4. $\alpha_i|\alpha_{(-i)},\beta,\sigma^2_\epsilon,\sigma^2_\alpha,\lambda_i,w_i,y \sim$
$N(\alpha_i|V_i\bar{r}_i,(1-V_i)\frac{\sigma^2_\alpha}{\lambda_i})$

5. $\lambda_i|\lambda_{(-i)},\alpha_i,\sigma^2_\alpha,y \sim$
$Gam(\lambda_i\left|\frac{a+1}{2}\right., \frac{1}{2}(\frac{\alpha_i^2}{\sigma^2_\alpha}+b))$

6. $w_i|w_{(-i)},\alpha_i,\beta,\sigma^2_\epsilon,y \sim$
$GIG(w_i\left|\frac{T-v}{2}\right., \frac{\epsilon'_i\epsilon_i}{2\sigma^2_\epsilon},\frac{v}{2})$

که در آن $\alpha_{(-i)}$، $\lambda_{(-i)}$ و $w_{(-i)}$ بردارهای مرتبط با حذف i-امین عضو هستند و

$$\begin{aligned}
\Omega &= (\sum_i(\frac{w_i x'_i x_i}{\sigma^2_\epsilon})+\Sigma_0^{-1})^{-1}\\
\hat{\beta} &= \Omega(\sum_i(\frac{w_i x'_i u_i}{\sigma^2_\epsilon})+\Sigma_0^{-1}\beta_0)\\
\epsilon_i &= (Y_i - X_i\beta - \alpha_i e_T)\\
\bar{r}_i &= \frac{1}{T}r'_i e_T\\
V_i &= \frac{T\sigma^2_\alpha}{T\sigma^2_\alpha+\sigma^2_\epsilon w_i^{-1}\lambda_i}
\end{aligned}$$

و GIG معرف توزیع وارون-گوسین تعمیم‌یافته با تابع چگالی

$$\begin{aligned}
f(x|p,a,b) &= \frac{\left(\frac{a}{b}\right)^{p/2}}{2Bes(p,\sqrt{ab})}x^{(p-1)}\\
&\quad\times exp[-\frac{1}{2}(ax+\frac{b}{x})],\\
&\quad x>0, a,b>0, p\in\mathbb{R}
\end{aligned}$$

است، که در آن $Bes(p,\sqrt{ab})$ تابع بسل تعدیل‌یافته نوع سوم [5] است.

حال کاربرد مدل‌های مذکور را در تحلیل داده‌های بازار بورس تهران نشان می‌دهیم.

۵ تحلیل قیمت سهام در بازار بورس تهران

در این بخش مدل‌های معرفی شده در این مقاله را در تحلیل داده‌های پانلی مربوط به قیمت سهام در

[5] Modified Bessel Function of the Third Kind

بازار بورس تهران به کار می‌بریم. برای انتخاب بهترین مدل برازش‌شده از معیار اطلاع بیزی شوارتز (BIC) که به صورت

$$BIC(\theta) = -2logL(data|\theta) + pln(N)$$

تعریف می‌شود، استفاده می‌کنیم، که در آن p تعداد پارامترهای مدل و N تعداد کل مشاهدات است. مدل با BIC کمتر به عنوان مدل بهتر انتخاب می‌شود. داده‌های این مثال بر اساس اطلاعات بازار بورس اوراق بهادار تهران در بازه زمانی سال‌های ۱۳۸۰ تا ۱۳۸۶ تهیه شده است[۲]. هدف از انجام این تحقیق بررسی تأثیر متغیرهای حسابداری و ویژگی‌های شرکت بر قیمت سهام ۹۲ شرکت پذیرفته شده در بازار بورس اوراق بهادار تهران است. متغیر پاسخ قیمت سهام در دوره جاری (P_t) و متغیرهای توضیحی قیمت سهام در دوره قبل (P_{t-1})، جریان نقد عملیاتی هر سهم ($OCFS$)، ارزش دفتری سهام (BVS)، سود هر سهام عادی (EPS)، مدت فعالیت شرکت (A) و اندازه شرکت (S) هستند.

داده‌ها به صورت پانلی می‌باشند چون مشاهدات شرکت‌ها در سال‌های مختلف وابسته‌اند. بنابراین از یک اثر تصادفی در مدل جهت تعدیل همبستگی بین مشاهدات مربوط به هر شرکت و کنترل تغییرپذیری بین آن‌ها استفاده می‌شود. مدل پانلی با اثر تصادفی شرکت‌ها، α_i، را به صورت

$$p_{it} = \beta_0 + \beta_1 OCFS_{it} + \beta_2 p_{i,t-1}$$
$$+\beta_3 BVS_{it} + \beta_4 EPS_{it}$$
$$+\beta_5 S_{it} + \beta_6 A_{it} + \alpha_i + \epsilon_{it}$$

درنظر می‌گیریم که در آن $i = 1, 2, ..., 92$ و $t = 1, 2, ..., 7$. ابتدا مدل پانلی با اثرهای تصادفی را با فرض مانده‌ها و اثرهای تصادفی نرمال برازش داده‌ایم. با انجام آزمون کلموگروف-اسمیرنوف برای مانده‌ها p-مقدار برابر با $0/02$ محاسبه شد. بنابراین توزیع نرمال برای مانده‌ها انتخاب مناسبی نیست. مدل‌های متنوعی را بر اساس خانواده توزیع SMN برای ϵ_{it} و α_i برازش داده و با معیاز انتخاب مدل BIC مقایسه نمودیم.

برای محاسبه برآورد بیز پارامترهای مدل، توزیع پیشین برای β_iها را $N(0, 10^{-3})$، برای ابرپارامترهای مؤلفه‌های واریانس توزیع $InvGam(1, 0/1)$ و برای درجه آزادی توزیع تی، یکنواخت پیوسته در بازه $(2, 50)$ را درنظر گرفتیم.

نتایج تحلیل برای پارامترهای مدل را، با ۱۰۰۰۰۰ نمونه شبیه‌سازی شده و دورریز ۱۰۰۰ مقدار اولیه با استفاده از الگوریتم نمونه‌گیر گیبز توسط نرم‌افزار اُپن باگز [۹] در جدول ۱ آورده‌ایم.

نتایج نشان می‌دهند که مدل تی-نرمال کمترین مقدار BIC را بین مدل‌های برازش‌یافته دارد. لذا بهترین مدل برازش شده آن است که مانده‌ها تی-استیودنت و اثرهای تصادفی نرمال باشند. هم‌چنین نتایج نشان می‌دهند که متغیرهای قیمت سهام در دوره قبل،

سود هر سهم عادی و مدت فعالیت شرکت بر قیمت سهام در دوره جاری مؤثرند. علامت مثبت برآورد ضرایب A و EPS نشان می‌دهد که با افزایش مدت فعالیت شرکت و سود سهم عادی شرکت‌ها، قیمت سهام در دوره جاری افزایش می‌یابد و از علامت منفی برآورد ضریب p_{t-1} استنباط می‌شود که قیمت سهام در دوره قبل اثر معکوس بر قیمت سهام در دوره جاری دارد.

۶ نتیجه‌گیری

در این مقاله خانواده توزیع‌های آمیخته-مقیاس نرمال معرفی گردید و کاربرد این توزیع‌ها در مدل پانلی با اثرهای تصادفی، با توجه به موضوع تغییرپذیری در مشاهدات و همبستگی بین مشاهدات مربوط به هر واحد آزمایشی بررسی گردید. با توجه به این‌که در نظر گرفتن توزیع‌های غیرنرمال برای اثرهای تصادفی و مانده‌ها منجر به پیچیدگی محاسبات در برآوردیابی پارامترها می‌شود، ساختار سلسله مراتبی جهت انجام استنباط بیزی درنظر گرفته شد. در آخر با تحلیل داده‌های وابسته قیمت سهام در بازار بورس تهران مشاهده شد که فرض نرمال برای مانده‌ها لزوماً مناسب نیست.

این مباحث می‌توانند در پیش‌بینی‌های اقتصادی و تصمیم‌گیری‌های صحیح مؤثر باشند.

۷ تشکر و قدردانی

نویسندگان این مقاله، از معاونت محترم تحصیلات تکمیلی دانشگاه اصفهان به خاطر فراهم نمودن تسهیلات پژوهشی و تامین نمودن هزینه‌ی اجرای این طرح در قالب پایان‌نامه، کمال تشکر و قدردانی را دارند. هم‌چنین از مدیر محترم، سردبیر و هیئت داوران که با بیان اصلاحات موجب ارتقا کیفیت این مقاله شدند، سپاسگزاری می‌شود.

جدول ۱: نتایج برازش مدل‌های مختلف برای داده‌های بورس اوراق بهادار تهران

	نرمال-نرمال	تی-نرمال	VG-تی	نرمال-VG	تی-پیرسن	تی-VG
عرض از مبدا	7/322	6/897	1/097	20/12	7/041	9/936
	(0/75)	(3/988)	(0/568)	(25/21)	(3/281)	(5/108)
$OCFS$	0/57	0/5312	1/148	0/521	0/522	0/473
	(0/004)	(0/056)	(0/332)	(0/097)	(0/048)	(0/079)
p_{t-1}	4/968	−5/16	−1/56	−9/58	−5/372	−1/185
	(1/21)	(1/761)	(2/509)	(3/519)	(1/727)	(1/912)
BVS	−3/77	6/61	0/002	6/243	5/236	2/409
	(6/48)	(3/551)	(0/002)	(1/461)	(3/05)	(3/864)
EPS	2/974	3/807	−0/002	2/189	3/84	3/716
	(7/94)	(4/839)	(0/002)	(1/887)	(4/282)	(4/742)
S	−0/529	−0/295	−8/42	−1/703	−0/288	−0/5
	(0/13)	(0/331)	(0/583)	(2/826)	(0/245)	(0/381)
A	0/025	8/541	0/313	0/0918	−7/638	0/002
	(0/009)	(0/008)	(0/083)	(0/199)	(0/003)	(0/006)
σ^2	0/205	0/107	5/5	0/318	0/108	0/103
	(0/016)	(0/0142)	(7/048)	(0/394)	(0/014)	(0/137)
τ^2	74/75	0/099	6	23/38	0/057	0/131
	(46/25)	(0/199)	(13/1)	(13/05)	(0/063)	(0/014)
BIC	1423/28	663/08	708/78	783/08	786/28	711/48

اعداد داخل پرانتز انحراف استاندارد بیزی را نشان می‌دهند.

مراجع

[1] Andrew, D.F. and Mallows, C.L. (1974), Scale mixtures of normal distributions, *J. Roy. Statist. Soc.*, 36B, 99-102.

[2] Ebrahimi, M. (2009), *The effects of accounting variables and firm's characteristics on stock prices of companies listed on TSE.* M.A. thesis.Uni. of Isfahan.

[3] Choy, S.T., and Chan, J.S.K. (2008), Scale mixtures distributions in statistical modeling, *Aust. N. Z. J. Stat.*, 50(2), 135-146.

[4] Congdon, P. D. (2006). *Bayesian statistical modeling*, John Wiley & Sons Ltd,.

[5] Gelfand, A.E. (2000), Gibbs sampling, *Journal of the American Statistical Association*, 95, 1300-1304.

[6] Gelman, A. (2006), Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis*, 3(1), 515-533.

[7] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995), *Continuous univariate distributions*, 2nd ed., Wiley, New York.

[8] Kim, H.M. and Genton, M.G. (2011)m Characteristic functions of scale mixtures of multivariate skew-normal distributions, *Journal of Multivariate Analysis*, 102(7), 1105-1117.

[9] Spiegelhalter, D. J. Thomas, A. Best, N. G. and Lunn, D. (2010), *OpenBugs user manual*, version 3.1.1.MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, and Department of Mathematics and Statistics, University of St Andrews, Scotland, and Department of Epidemiology and Public Health, Imperial College School of Medicine, London,(www.mrc-bsu.cam.ac.uk/bugs).