دومین کارگاه
اندازه های اطلاعات و کاربردهای آن
8 و 9 بهمن 1393، قطب علمی داده های ترتیبی و فضایی
دانشگاه فردسی مشهد

شاخص تایل بر اساس توزیع لمبدای تعمیم‌یافته و کاربردی از آن

کدخدا، ا. [۱]، محتشمی برزادران، غ. ر.[۱]، امینی، م. [۱] و محمدی، م.[۲]

[۱] گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

[۲] گروه آمار، دانشکده علوم پایه، دانشگاه زابل

چکیده

هدف اصلی این مقاله محاسبه‌ی شاخص تایل برای توزیع لمبدای تعمیم‌یافته است. در این راستا ابتدا مروری خواهیم داشت بر توزیع لمبدای تعمیم‌یافته و پس از آن آنتروپی و شاخص تایل را معرفی خواهیم کرد. در انتها شاخص تایل متناظر با توزیع لمبدای تعمیم‌یافته را برای داده‌های هزینه، درآمد و سود شرکت‌های پذیرفته شده در بازار بورس اوراق بهادار تهران محاسبه می‌نماییم.

کلمات کلیدی: توزیع لمبدای تعمیم‌یافته، نظریه اطلاع، شاخص تایل.

۱ مقدمه

یکی از موضوعاتی که جوامع امروزی را با مشکل مواجه کرده، توزیع ناعادلانه درآمد و ثروت در بین اقشار مختلف جامعه و بروز نابرابری اقتصادی در جامعه است. از این‌رو محاسبه‌ی میزان این نابرابری توجه بسیاری از اقتصاددانان و سیاست‌گذاران را به خود جلب کرده است. با توجه به اهمیت این موضوع، معیارها و شاخص‌های متعددی برای اندازه‌گیری میزان نابرابری پیشنهاد شده‌اند. در سال ۱۹۰۵ ماکس اوتو لورنز منحنی‌ای را برای نمایش توزیع درآمد پیشنهاد کرد که به منحنی لورنتس مشهور است. معیارهای سنجش نابرابری زیادی بر اساس منحنی لورنتس تعریف شده است که از میان آنها می‌توان به ضریب جینی، ضریب بن فرونی و ضریب زنگا اشاره کرد. با افزایش حساسیت‌ها نسبت به توزیع درآمد و اندازه‌گیری نابرابری شاخص‌ها و معیارهای دیگری نظیر شاخص تایل، شاخص اتکینسون

[۱] Email: ekadkhoda@gmail.com

و شاخص هرفیندال، معرفی گردیدند که از این میان ما بر روی شاخص تایل تمرکز خواهیم کرد. در واقع در این شاخص از مفهوم آنتروپی و نظریه اطلاع برای اندازه‌گیری میزان نابرابری جامعه استفاده شده است.

یکی از روش‌های محاسبه‌ی این معیارهای نابرابری بدین ترتیب است که ابتدا یک توزیع درآمدی مناسب به داده‌ها برازش داده و پارامترهای توزیع را برآورد نماییم سپس معیارهای نابرابری را محاسبه کنیم برای این منظور از توزیع لمبدای تعمیم‌یافته که توزیعی پرکاربرد در زمینه‌ی داده‌های درآمد و دارایی می‌باشد، استفاده می‌نماییم. روش دیگر این است که بدون در نظر گرفتن توزیع خاصی برای داده‌ها به صورت مستقیم آنها را محاسبه نماییم.

در اینجا برای داده‌های سود حاصل از فروش اوراق بهادار کلیه شرکت‌های پذیرفته شده در بازار بورس تهران با استفاده از هر دو روش ذکر شده مقدار شاخص تایل را محاسبه خواهیم کرد.

۲ توزیع لمبدای تعمیم‌یافته و روش‌های برآورد پارامتر

توزیع لمبدای تعمیم‌یافته GLD تعمیمی از توزیع تک پارامتری توکی است که در سال ۱۹۶۰ توسط جان توکی معرفی شده است. اولین روش پارامتری کردن این تعمیم چهار پارامتری که به $RS-GLD$ مشهور است را در سال ۱۹۷۴ رمبرگ و اشمیزر پیشنهاد کردند و با تابع چندکی‌اش مشخص می‌شود که برابر با

$$Q_{RS}(u)=\lambda_1+\frac{u^{\lambda_3}-(1-u)^{\lambda_4}}{\lambda_2} \qquad (1)$$

است و تابع چگالی آن در نقطه‌ی $x=F^{-1}(u)$ از رابطه‌ی $f_{RS}(x)=\frac{\lambda_2}{\lambda_3 u^{\lambda_3-1}+\lambda_4(1-u)^{\lambda_4-1}}$ بدست می‌آید که $u\in[0,1]$ (کاریان و دودیکس ۲۰۰۰). رمبرگ و همکاران (۱۹۷۹) ذکر کردند که ترکیبات معلومی از λ_3 و λ_4 هستند که تابع توزیع فوق در آنها یک توزیع احتمالی معتبر نمی‌باشد. برای مطالعه دقیقتر خواننده را به کاریان و دودویکز[۳] (۲۰۰۰) ارجاع می‌دهیم. برای توسیع تکیه‌گاه توزیع روی همه‌ی صفحه‌ی (λ_3,λ_4)، فریمر و همکاران روش پارامتری کردن $FMKL$ را معرفی کردند که برابر است با

$$Q_{FMKL}(u)=\lambda_1+\frac{1}{\lambda_2}\left(\frac{u^{\lambda_3}-1}{\lambda_3}-\frac{(1-u)^{\lambda_4}-1}{\lambda_4}\right) \qquad (2)$$

و روی همه‌ی صفحه (λ_3,λ_4) تعریف می‌شود و تابع چگالی آن برابر است با:

$$f_{FMKL}(x)=\frac{\lambda_2}{u^{\lambda_3-1}+(1-u)^{\lambda_4-1}},\quad 0\le u\le 1.$$

قضیه ۱.۲. گشتاور مرتبه k، GLD وجود دارد اگر و تنها اگر:

$$min(\lambda_3,\lambda_4)>-\frac{1}{k}.$$

از طرفی امید ریاضی توزیع لمبدای تعمیم‌یافته برای هر دو روش پارامتری کردن آن به ترتیب از روابط

$$\mu_{RS}=\lambda_1+\frac{1}{\lambda_2}\left(\frac{1}{1+\lambda_3}-\frac{1}{1+\lambda_4}\right)$$

و

$$\mu_{FMKL} = \lambda_1 - \frac{1}{\lambda_2\lambda_3} + \frac{1}{\lambda_2\lambda_4} + \frac{1}{\lambda_2}\left(\frac{1}{\lambda_3(\lambda_3+1)} - \frac{1}{\lambda_4(\lambda_4+1)}\right)$$

بدست می‌آیند که با توجه به قضیه۱.۲ این مقادیر تنها در صورتی وجود دارند که

$$min(\lambda_3, \lambda_4) > -1.$$

برای مطالعه دقیقتر به لاخانی و موزر[۱] (۲۰۰۰) مراجعه کنید.

همانطور که اشاره شد هدف برازش GLD به یک مجموعه داده و سپس محاسبه شاخص تایل است. لذا باید در ابتدا پارامترهای توزیع را برآورد نماییم. برای برآورد پارامترهای این توزیع روشهای متفاوتی ارایه شده است که از میان آنها می‌توان به روش گشتاورها، صدکها، $starship$ و درستنمایی ماکسیمم اشاره کرد. ما در این تحقیق ار روش درستنمایی ماکسیمم استفاده می‌کنیم. در این روش که توسط سو[۴] (۲۰۰۷) ارائه شده است، از اعداد شبه تصادفی و روش صدکها برای $RS-GLD$ و روش گشتاورها برای $FMKL-GLD$ جهت یافتن مقادیر ورودی استفاده می‌کنیم. سپس این مقادیر ورودی را برای ماکسیمم کردن لگاریتم عددی درستنمایی بکار می‌بندیم تا پارامترهای GLD مناسب را بیابیم. الگوریتم روش درستنمایی ماکسیمم در ۴ گام خلاصه می‌شود:

گام ۱: محدوده‌ای از مقادیر ورودی برای λ_3 و λ_4 و همچنین تعداد مقادیر ورودی انتخاب شده را تعیین کنید. در نرم افزار R از دنباله سوبول برای تولید اعداد شبه تصادفی استفاده می‌شود. به طور پیش فرض تعداد مقادیر پیش فرض ۱۰۰۰۰ در نظر گرفته می‌شود.

گام ۲: λ_1 و λ_2 را برای هرکدام از λ_3 و λ_4 محاسبه کنید. مجموعه‌ی مقادیر ورودی که در پارامتری کردن GLD صدق نکنند یا همه‌ی فضای مجموعه داده ها را پوشش ندهند، غیرقابل قبول تلقی کرده و آنها را حذف کنید. در میان مجموعه نقاط شروع باقیمانده مجموعه‌ای را بیابید که کمترین مقدار $\sqrt{(\hat{\rho}_3-\rho_3)^2-(\hat{\rho}_4-\rho_4)^2}$ و $\sqrt{(\hat{\alpha}_3-\alpha_3)^2-(\hat{\alpha}_4-\alpha_4)^2}$ را تولید کند. حال این مجموعه مقادیر برای فرآیند بهینه‌سازی مورد استفاده قرار می‌گیرد.

گام ۳: با جایگذاری مقادیر ورودی در فرمول‌های (۱) و (۳) صدکهای u_i را برای $RS-GLD$ و $FMKL-GLD$ بیابید.

گام ۴: u_i بدست آمده در گام ۳ را در فرمولهای لگاریتم عددی درستنمایی (عبارات (۴) و (۳)) جایگذاری کنید.

$$ML_{RS} = \sum_{i=1}^{n} \log\left[\frac{\lambda_2}{\lambda_3 u_i^{\lambda_3-1} + \lambda_4(1-u_i)^{\lambda_4-1}}\right] \qquad (۳)$$

$$ML_{FMKL} = \sum_{i=1}^{n} \log\left[\frac{\lambda_2}{u_i^{\lambda_3-1} + (1-u_i)^{\lambda_4-1}}\right] \qquad (۴)$$

گام ۵: مقدار بهینه با استفاده از الگوریتم سیمپلکس نلدر-مید یا هر الگوریتم مینیمم سازی عددی دیگر بدست می‌آید.

برای برازش GLD و یافتن پارامترهای مناسب آن با استفاده از روش درستنمایی ماکسیمم در بسته $GLDEX$ در نرم افزار R از تابع $fun.data.fit.mm$ استفاده می‌کنیم.

۳ آنتروپی و نظریه اطلاع

فرض کنید $p_i, i = ۱, ۲, ..., n$ احتمال وقوع یک پیشامد گسسته باشد، در نظریه اطلاع تابع H مدنظر است که با احتمال وقوع پیشامد رابطه‌ی عکس دارد و میزان اطلاع موجود در یک توزیع احتمال را نشان می‌دهد. این تابع که آنتروپی نام دارد، نخستین بار توسط شانن [۲] ۱۹۴۸ معرفی شد و به صورت زیر تعریف می‌شود:

$$H = -\sum_{i=۱}^{n} p_i \log p_i \qquad (۵)$$

و در آن $\sum_{i=۱}^{n} p_i = ۱$ است. آنتروپی برای متغیر گسسته همواره نامنفی است و بیشترین مقدار آن $\log n$ می‌باشد که زمانی رخ میدهد که احتمال وقوع همه‌ی پیشامدها برابر باشد در واقع در این حالت بیشترین عدم حتمیت را داشته خواهیم داشت و عدم حتمیت از این جهت است که نمی‌دانیم کدام پیشامد رخ خواهد داد.

برای متغیر تصادفی مطلقاً پیوسته X با تابع چگالی احتمال $f(x)$، آنتروپی برابر است با

$$H = \int_{۰}^{\infty} f(x) \log f(x) dx \qquad (۶)$$

که میزان بی‌نظمی و اغتشاش تابع چگالی $f(x)$ را اندازه می‌گیرد، به عبارتی هر چه چگالی به تابع چگالی یکنواخت نزدیک‌تر باشد مقدار آنتروپی نیز کمتر می‌شود و تصمیم‌گیری براساس آن مشکل‌تر خواهد شد.

۴ شاخص تایل

در سال ۱۹۶۷ تایل[۵] یک شاخص جدید برای ارزیابی توزیع درآمد معرفی کرد که بر پایه مفهوم آنتروپی در نظریه اطلاع استوار است و به صورت تفاضل میان آنتروپی توزیع درآمد جامعه از آنتروپی درآمد عادلانه تعریف می‌شود. تایل در بحث کاربرد نظریه اطلاع در توزیع درآمد به جای استفاده از احتمال رخداد یک حادثه از سهم درآمد در جامعه استفاده کرد.

فرض کنید $x_۱$، $x_۲$ و ... و x_n درآمدهای افراد جامعه را نشان دهد که $x_i \geq ۰$. در این صورت سهم درآمد فرد iام برابر با $\frac{x_i}{n\overline{x}}$ می‌باشد که $\overline{x}$ میانگین درآمد کل جامعه را نشان می‌دهد. طبق تعریف شاخص تایل تفاضل میان آنتروپی

سهم درآمد جامعه از آنتروپی درآمد عادلانه یعنی زمانی که سهم درآمد هر فرد $\frac{۱}{n}$ باشد را اندازه می‌گیرد و برابر است با:

$$\begin{aligned} T &= \frac{۱}{n}\sum_{i=۱}^{n}\log\frac{۱}{n} - \sum_{i=۱}^{n}\frac{x_i}{n\overline{x}}\log\frac{n\overline{x}}{x_i} \\ &= \frac{۱}{n}\sum_{i=۱}^{n}\frac{x_i}{\overline{x}}\log\frac{x_i}{\overline{x}} \end{aligned} \quad (۷)$$

شاخص تایل بین صفر (حالت برابری کامل توزیع درآمد) و $\log n$ (حالت نابرابری کامل توزیع درآمد) تغییر می‌کند.

اگر برای افراد جامعه توزیع درآمد پیوسته در نظر گرفته شود در این صورت شاخص تایل برابر خواهد بود با:

$$T = \int_{۰}^{\infty}\frac{x}{\mu}\log\frac{x}{\mu}dF(x) = \int_{۰}^{۱}\frac{Q(u)}{\mu}\log\frac{Q(u)}{\mu}du. \quad (۸)$$

که در آن $F(x)$ تابع توزیع تجمعی با دامنه‌ی زیر مجموعه‌ی اعداد حقیقی نامنفی و میانگین متناهی μ و $Q(u)$ تابع چندکی متغیر تصادفی X می‌باشد.

۵ شاخص تایل بر اساس توزیع لمبدای تعمیم یافته

اگر از توزیع لمبدای تعمیم‌یافته برای توزیع داده‌های درآمد استفاده کنیم آنگاه با استفاده از فرمول (۸) شاخص تایل برای هر دو نوع پارامتری کردن این توزیع بدست می‌آید:

$$T_{RS} = \int_{۰}^{۱}\left(\frac{\lambda_۱\lambda_۲ + u^{\lambda_۳} - (۱-u)^{\lambda_۴}}{\mu_{RS}\lambda_۲}\right)\log\left(\frac{\lambda_۱\lambda_۲ + u^{\lambda_۳} - (۱-u)^{\lambda_۴}}{\mu_{RS}\lambda_۲}\right)du, \quad (۹)$$

$$\begin{aligned} T_{FMKL} = \int_{۰}^{۱}&\left(\frac{\lambda_۱}{\mu_{FMKL}} + \frac{۱}{\lambda_۲\mu_{FMKL}}\left(\frac{u^{\lambda_۳}-۱}{\lambda_۳} - \frac{(۱-u)^{\lambda_۴}-۱}{\lambda_۴}\right)\right) \\ &\log\left(\frac{\lambda_۱}{\mu_{FMKL}} + \frac{۱}{\lambda_۲\mu_{FMKL}}\left(\frac{u^{\lambda_۳}-۱}{\lambda_۳} - \frac{(۱-u)^{\lambda_۴}-۱}{\lambda_۴}\right)\right)du, \end{aligned} \quad (۱۰)$$

که μ_{FMKL} و μ_{RS} به ترتیب میانگین روش پارامتری $FMKL$ و RS می‌باشند. بدست آوردن انتگرال‌های (۹) و (۱۰) به روش‌های مستقیم تا حدی مشکل است لذا با شبیه‌سازی مونت‌کارلو آنها را بدست می‌آوریم.

۶ یافته‌های عددی

در این تحقیق هزینه، درآمد و سود حاصل از فروش کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران که از ابتدای سال ۱۳۹۰ تا پایان سال ۱۳۹۱ فعالیت مستمر داشته‌اند، را در نظر گرفته‌ایم. ابتدا جهت ارائه‌ی اطلاعات

کلی از داده‌های مورد استفاده، به آمار توصیفی متغیرهای هزینه، درآمد و سود پرداخته‌ایم که نتایج به شرح جدول۹ است.

جدول ۹: آمارتوصیفی داده‌ها

	سال	میانگین	انحراف معیار	چولگی	کشیدگی
هزینه	۱۳۹۰	۸۲۷۶۰	۵۱۱۴۸۳/۲	۱۴/۰۴۰۵	۲۱۹/۳۰۶۹
	۱۳۹۱	۴۶۵۱۰	۱۱۳۰۴۴/۸	۴/۳۹۵۷	۲۴/۲۰۹۵
درآمد	۱۳۹۰	۱۸۸۹۰۰۰	۱۰۶۷۱۲۳۱	۱۳/۰۰۵۷	۱۹۹/۰۲۴
	۱۳۹۱	۶۹۹۶۰۰	۲۴۴۶۲۹۶	۵/۳۵۰۱	۳۱/۴۵۴۴
سود	۱۳۹۰	۳۹۳۴۰۰	۱۳۸۸۵۱۸	۷/۸۶۶۶	۸۱/۳۱۸۶
	۱۳۹۱	۳۹۷۷۰۰	۱۵۸۳۰۳۹	۶/۵۹۰۶	۴۸/۱۰۵۹

همانطور که در جدول ۹ مشاهده می‌کنیم مقادیر چولگی و کشیدگی مربوط به داده‌ها زیاد می‌باشد که بیانگر این حقیقت است که داده‌ها انحراف زیادی از توزیع نرمال دارند.

سپس با استفاده از روش درستنمایی ماکسیمم پارامترهای GLD را برای هر دو روش پارامتری RS و $FMKL$ برای هرسه متغیر هزینه، درآمد و سود محاسبه نموده و توزیع GLD را به این داده‌ها برازش می‌دهیم و مقدار شاخص تایل را با استفاده از فرمول‌های(۷)، (۹) و (۱۰) برای آنها محاسبه می‌کنیم که نتایج به شرح جدول ۱۲ می‌باشد.

مقادیر مبهم گزارش شده در جدول ۱۲ به دلیل آن است که مقادیر λ_3 و λ_4 در شرط ذکر شده در قضیه ۱.۲ صدق نمی‌کنند و لذا در این موارد امیدریاضی توزیع تعریف نمی‌شود و نمی‌توان با استفاده از GLD شاخص تایل را برای آنها حساب کرد. همچنین مقادیر $*$ جدول نیز که در ستون T_{RS} دیده می‌شود به این دلیل است که پارامترهای حاصل از برازش داده‌ها در روش پارامتری RS یک توزیع معتبر را تولید نمی‌کند و نشان‌دهنده‌ی محدودیت برازش $RS - GLD$ می‌باشد. در حالت کلی $FMKL - GLD$ برازش بهتری را نتیجه می‌دهد و در اکثر موارد می‌توان شاخص تایل را بر اساس آن محاسبه کرد.

جدول ۱۰: مقادیر شاخص تایل برای میزان فروش و درآمد و سود شرکت‌های بورس

	سال	T	T_{RS}	T_{FMKL}
هزینه	۱۳۹۰	۲/۳۷۹۳	*	۴/۶۷۸۶
	۱۳۹۱	۱/۳۱۲۴	*	مبهم
درآمد	۱۳۹۰	۲/۳۴۶۳	مبهم	۳/۲۵۸۱
	۱۳۹۱	۱/۸۳۳۷	*	۲/۳۸۸۰
سود	۱۳۹۰	۱/۷۴۱۵	*	۲/۰۷۲۴
	۱۳۹۱	۲/۰۵۷۵	مبهم	۲/۲۳۸۴

۷ دست‌آوردهای پژوهش

در این پژوهش ضمن معرفی توزیع لمبدای تعمیم‌یافته و یکی از روش‌های برآورد پارامترهای آن، شاخص نابرابری تایل را بر اساس هر دو روش پارامتری کردن $RS-GLD$ و $FMKL-GLD$ برای داده‌های هزینه، درآمد و سود حاصل از فروش کلیه‌ی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران، محاسبه کردیم.

مراجع

[1] Lakhany, A. and Massuer, H. (2000). Estimating the parameters of the generalised lambda distribution. *Algo Research Quarterly*, 3(3), 47-58.

[2] Shannon, C. E. (1948), A mathematial theory of communiation. *Bell SystemTechnicalJournal*, 27, 370-432.

[3] Karian, Z.A. and Dudewicz. E.J. (2000). Fitting Statistical Distributions to Data: Generalized Lambda Distribution and Generalized Bootstrap Methods. *CRC Press, New York*.

[4] Su, S. (2007). Numerical maximum log likelihood estimation for generalized lambda distributions. *Journal of Computational statistics and data analysis* 51(8), 3983-3998.

[5] Theil, H. (1967). Economics and Information Theory. *Amsterdam: North Holland*.