گزیده‌مطالب آماری
سال ۱۹، شماره‌ی ۱، بهار و تابستان ۱۳۸۷، صص ۱۴-۱

براورد منحنی لورنتس و ضریب جینی به‌روش پارامتری

آزاده مجیری،[۱] غلامرضا محتشمی برزادران،[۲،*] یدا... واقعی[۱]

[۱] دانشگاه بیرجند
[۲] دانشگاه فردوسی مشهد

چکیده. منحنی لورنتس یک ابزار مهم برای اندازه‌گیری نابرابری درامد است، شاخص‌های بسیاری بر اساس منحنی لورنتس برای اندازه‌گیری میزان نابرابری تعریف می‌شوند، ضریب جینی یکی از مهم‌ترین این شاخص‌هاست. در این مقاله ابتدا به معرفی منحنی لورنتس و ضریب جینی می‌پردازیم، سپس پارامترهای توزیع‌های احتمال درامد را به‌روش ماکسیمم درستنمایی براورد می‌کنیم. فرم‌های تابعی لورنتس را به دو روش براورد کرده و بهترین فرم تابعی لورنتس را معرفی می‌کنیم. در نهایت بر اساس داده‌های حاصل از آمارگیری هزینه و درامد خانوار ایران در سال ۱۳۸۴ منحنی لورنتس و ضریب جینی را براورد می‌کنیم.

۱- مقدمه

همان‌طور که می‌دانیم آمار یکی از مهم‌ترین علوم کاربردی است که با سایر رشته‌های علمی از جمله اقتصاد مرتبط می‌باشد. در کشورهای عقب‌مانده و تابع نظام سرمایه‌داری، قسمت بیش‌تر درامدها بین اقلیت مردم و قسمت کمی از آن بین اکثریت مردم تقسیم می‌شود. بر اساس این نوع توزیع غیرعادلانه‌ی درامدها، منحنی‌ای حاصل می‌شود که برای اولین بار توسط ماکس اوتو لورنتس در سال ۱۹۰۵ معرفی شد [۷] و به منحنی لورنتس معروف است. در آمار معمولاً برای اندازه‌گیری میزان پراکندگی داده‌ها از واریانس،

واژگان کلیدی: براورد ؛کمترین توان‌های دوم؛ براورد ماکسیمم درستنمایی؛ ضریب جینی؛ فرم‌های تابعی؛ منحنی لورنتس.
دریافت: ۱۳۸۷/۱۰/۱، پذیرش: ۱۳۸۸/۱/۲۴
* نویسنده‌ی عهده‌دار مکاتبات

انحراف معیار یا ضریب تغییرات استفاده می‌شود، ولی در اقتصاد برای اندازه‌گیری میزان پراکندگی یا نابرابری درامد از شاخص دیگری به نام ضریب جینی استفاده می‌شود. اگرچه می‌توان منحنی لورنتس را به‌طور مستقیم از داده‌های تجربی محاسبه کرد اما براورد پارامتری منحنی لورنتس برای داده‌های درامد مفید و ارزنده است [۳].

حداقل دو راهبرد برای براورد پارامتری منحنی لورنتس وجود دارد؛

۱- یک تابع چگالی احتمال مناسب برای توزیع درامد یافته، پارامترهای آن را به‌روش مناسب براورد کرده و با جایگذاری براورد پارامترها در تابع منحنی لورنتس به براورد منحنی برسیم [۱].

۲- منحنی لورنتس داده‌ها را به شکل نقطه به نقطه براورد کرده و یک خانواده‌ی پارامتری لورنتس به آن برازش می‌دهیم [۱۰].

بنا بر این در بخش دوم منحنی لورنتس و ضریب جینی را معرفی کرده و در بخش سوم پارامترهای توزیع‌های احتمال را براورد می‌کنیم (روش اول). در بخش چهارم به براورد فرم‌های تابعی لورنتس پرداخته (روش دوم) و در پایان نیز نتایج تجربی داده‌های هزینه و درامد خانوار کل کشور را بیان می‌کنیم.

۲- منحنی لورنتس و ضریب جینی

فرض کنید درامد n فرد از جامعه یا نمونه را به‌صورت $x_1, x_2, \ldots, x_n$ نشان دهیم سپس درامدها را از کوچک به بزرگ مرتب می‌کنیم و مرتب‌شده‌ی آن را به‌صورت $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ نشان می‌دهیم. تابع لورنتس با نماد $L(\cdot)$ به‌صورت زیر تعریف می‌شود [۶].

$$L\left(\frac{k}{n}\right) = \frac{\sum_{i=1}^{k} x_{i:n}}{\sum_{i=1}^{n} x_{i:n}}, \qquad k = 1, \ldots, n \tag{1}$$

که $L(0) = 0$ و $x_{i:n}$ درامد iامین فرد مرتب‌شده است. اگر زوج نقاط $\left(\left(\frac{k}{n}\right), L\left(\frac{k}{n}\right)\right)$ را برای مقادیر مختلف k در یک شکل رسم کنیم نمودار محدبی به وجود می‌آید که به آن منحنی لورنتس می‌گویند.

یکی از مهم‌ترین شاخص‌هایی که توسط کورادو جینی [۴] برای اندازه‌گیری میزان نابرابری معرفی شد ضریب جینی است که به‌صورت دو برابر ناحیه‌ی بین منحنی لورنتس و خط نیم‌ساز ربع اول تعریف می‌شود که همواره نامنفی و مقادیری بین صفر و یک را می‌گیرد. برای یک جامعه با مقادیر اندیس $i = 1, 2, \ldots, n$ عبارت زیر معرف ضریب جینی است.

$$G = \frac{1}{n}\left(n + 1 - 2\frac{\sum_{i=1}^{n}(n+1-i)x_{i:n}}{\sum_{i=1}^{n} x_{i:n}} \right) \qquad (۲)$$

مقدار ضریب جینی تجربی با استفاده از نرم افزار SPlus و از طریق برنامه‌نویسی محاسبه می‌شود. از نرم‌افزارهای دیگری مانند Stata هم می‌توانیم استفاده کنیم.

اگر برای درامد، یک متغیر تصادفی X با تابع چگالی احتمال $f(x)$ در نظر بگیریم باز هم می‌توانیم تابع لورنتسی برای آن تعریف کنیم. در این حالت اگر $f(x)$ تابع چگالی احتمال درامد جامعه، تعریف‌شده برای متغیر تصادفی نامنفی X باشد و $F(x) = \Pr(X \le x) = \int_0^x f(y)dy$ تابع توزیع تجمعی، F^{-1} تابع چندکی و $E(X) = \int_0^\infty xf(x)dx$ موجود، متناهی و نامنفی باشد (که با تغییر متغیر $t = F(x)$ منجر به $E(X) = \int_0^1 F^{-1}(t)dt$ می‌شود) آن‌گاه $L(u)$ به‌صورت زیر تعریف می‌شود.

$$L(u) = \frac{1}{E(X)}\int_0^u F^{-1}(t)\,dt, \qquad u \in [0,1] \qquad (۳)$$

تابع لورنتس روی بازه‌ی $[0, 1]$ پیوسته، صعودی و محدب با $L(0) = 0$ و $L(1) = 1$ است. اگر متغیرتصادفی X دارای تابع لورنتس $L(u)$ باشد ضریب جینی آن به‌صورت زیر تعریف می‌شود.

$$G = 2\int_0^1 \left[u - L(u)\right]du = 1 - 2\int_0^1 L(u)du \qquad (۴)$$

۳- براورد منحنی لورنتس به‌روش ماکسیمم درستنمایی

برای یافتن توزیع مناسب داده‌های درامد می‌بایست چند توزیع که شکلی مشابه توزیع تجربی داده‌ها دارند به داده‌ها برازش داده شود و با مقایسه‌ی آن‌ها توزیع مناسب داده‌ها را انتخاب کرد. برای این منظور لازم است با استفاده از روش‌های براورد پارامترهای توزیع‌ها، مثل روش ماکسیمم درستنمایی، پارامترهای این توزیع‌ها را براورد کرد. در بیش‌تر توزیع‌های متداول یا کلاسیک براوردگر ماکسیمم درستنمایی پارامترها منحصر به فرد بوده و با مشتق‌گیری معمولی از تابع درستنمایی ($L(\theta)$) یا تابع لگاریتم درستنمایی ($\ln L(\theta)$) نسبت به پارامتر θ به دست می‌آید.

تابع چگالی احتمال توزیع پاراتو با پارامترهای $x_۰$ و α برابر

$$f(x)=\alpha x_۰^{\alpha} x^{-(\alpha+۱)}, \quad ۰<x_۰\leq x<\infty,\ \alpha>۰$$

است. براوردگر پارامتر α به‌روش ماکسیمم درستنمایی به‌صورت زیر می‌باشد.

$$\hat{\alpha}=n\left[\sum_{i=۱}^{n}\ln\left(\frac{X_i}{\hat{x}_۰}\right)\right]^{-۱}$$

با توجه به این‌که دامنه‌ی توزیع به $x_۰$ بستگی دارد و از طریق مشتق‌گیری معمولی مقدار $x_۰$ را نمی‌توان براورد کرد، بنا بر این $\hat{x}_۰=x_{۱:n}$ را در نظر می‌گیریم.

تابع چگالی احتمال توزیع لگ نرمال با پارامترهای μ و α برابر

$$f(x)=\frac{۱}{x\sqrt{۲\pi\sigma^۲}}\exp\left\{\frac{-۱}{۲\sigma^۲}(\ln x-\mu)^۲\right\}, \quad x>۰,\ \mu\in\Re,\ \sigma>۰$$

است. براوردگر پارامترهای μ و σ^2 به روش ماکسیمم درستنمایی به‌صورت زیر است.

$$\hat{\mu}=\frac{۱}{n}\sum_{i=۱}^{n}\ln X_i,$$

$$\hat{\sigma}^۲=\frac{۱}{n}\sum_{i=۱}^{n}(\ln X_i-\overline{\ln X})^۲.$$

تابع چگالی احتمال گامای تعمیم‌یافته با پارامترهای α و β و p برابر

$$f(x)=\frac{a}{\beta^{ap}\Gamma(p)}x^{ap-۱}e^{-\left(\frac{x}{\beta}\right)^{a}}, \qquad x>۰ \quad a,\beta,p>۰$$

است. اگر $a<0$ آن‌گاه توزیع گامای تعمیم‌یافته‌ی معکوس به دست می‌آید.

هرگاه $p=1$ توزیع گاما، $a=-1$ توزیع گامای معکوس و $a>0\ ,\ p=1$ توزیع وایبل نتیجه می‌شود.

تابع چگالی احتمال بتای تعمیم‌یافته‌ی نوع دوم برابر

$$f(x)=\frac{a x^{ap-1}}{b^{ap}B(p,q)\left[1+\left(\frac{x}{b}\right)^{a}\right]^{p+q}}\ ,\qquad x>0\quad a,b,p,q>0$$

است، که در آن $B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ و $\Gamma(\cdot)$ تابع گاماست.

هرگاه $p=1$ توزیع سینگ مادالا، $q=1$ توزیع داگم، $p=q=1$ توزیع فیسک، $p=a=1$ توزیع لوماکس و $q=a=1$ توزیع لوماکس معکوس به دست می‌آید.

برآورد پارامترهای خانواده‌ی توزیع گاما و بتا با استفاده از مشتق‌گیری معمولی قابل محاسبه نیست و باید با روش‌های عددی محاسبه شود، که از فرمان optim در نرم افزار R استفاده می‌کنیم.

با داشتن یک مجموعه‌ی داده، هر توزیعی را که قلمرو آن بیش‌تر از دامنه‌ی تغییرات داده‌ها باشد می‌توان به آن برازش داد، ولی باید توجه داشت که همیشه هر توزیعی نمی‌تواند برازنده‌ی داده‌ها باشد. لذا اگر چند توزیع مختلف را به داده‌ها برازش دهیم برای بررسی بهترین توزیعی که به داده‌ها برازش داده می‌شود داده‌ها را به K رده ($K=40$) رده‌بندی می‌کنیم. رده‌ها به‌گونه‌ای انتخاب شده است که حداقل ۲۷۰ نمونه در هر رده قرار گیرد. سپس از آماره‌ی خی‌دو (χ^2) به‌صورت زیر استفاده می‌کنیم.

$$\chi^2=\sum_{k=1}^{K}\frac{\left(n_k-Np_k\left(\hat{\theta}\right)\right)^2}{Np_k\left(\hat{\theta}\right)},$$

که n_k فراوانی مشاهده‌شده در رده‌ی kام، $Np_k(\hat{\theta})$ فراوانی مورد انتظار تحت توزیع مورد نظر، $p_k(\hat{\theta})=F(x_k,\theta)-F(x_{k-1},\theta)$، و x_k و x_{k-1} به‌ترتیب کران بالا و کران پایین رده‌هاست. جدول ۱ مقادیر آماره‌ی خی‌دو (χ^2)، p- مقدار (p-value) و

براورد پارامترهای توزیع‌های احتمال درامد را به‌طور خلاصه نشان می‌دهد.

جدول ۱– مقادیر آماره‌ی خی دو، p– مقدار و براورد پارامترهای توزیع‌های برازش داده‌شده به داده‌های هزینه و درامد خانوار کل کشور

نام توزیع	آماره‌ی خی‌دو $\chi^۲$	p-value	براورد پارامترها
پاراتو	۶۷۱۳۹/۱۳۳	۰	$\alpha = ۰/۲۷۹۳$ $x_۰ = ۰/۲۴$
لگ نرمال	۲۵۳/۰۴۵۴	۰	$\mu = ۲/۱۵۳۸$ $\sigma^۲ = ۰/۵۹۲۵$
گامای تعمیم‌یافته	۱۰۵۹/۹۶۲	۰	$a = ۰/۳۵۸۲,\ \beta = ۰/۰۰۸۱$ $p = ۱۲/۶۷۱۱$
گاما	۴۹۵۵/۹۹۴	۰	$\beta = ۶/۸۸۳۷$ $p = ۱/۷۲۰۱$
گامای معکوس	۸۷۷/۸۴۵۵	۰	$\beta = ۱۲/۳۶۷$ $p = ۱/۹۰۷۴$
وایبل	۵۴۲۴/۹۳	۰	$a = ۱/۲۱۲۴$ $\beta = ۱۲/۷۵۷۹$
بتای تعمیم‌یافته‌ی نوع دوم	۴۵/۹۷۵۳	۰/۱۰۱۵	$a = ۰/۴۹۹۷,\ b = ۰/۵۷۰۴$ $p = ۳۳/۴۸۶,\ q = ۸/۹۸۸۸$
فیسک	۲۹۸/۲۰۰۲	۰	$a = ۲/۲۹۷۲$ $b = ۸/۴۸۳۸$
لوماکس	۶۲۰۲/۲۷۹	۰	$b = ۱۳۴/۷۰۲۷$ $q = ۱۲/۳۰۳$
لوماکس معکوس	۵۰۲۲/۳۵۵	۰	$b = ۰/۲۵۴۳$ $p = ۲۶/۱۰۷۸$

با توجه به اطلاعات مربوط به جدول از میان تمامی توزیع‌هایی که به داده‌های هزینـه

خانوار کل کشور برازش داده شد توزیع بتای تعمیم‌یافته‌ی نوع دوم با p- مقدار ۰/۱۰۱۵ در سطح خطای ۰/۰۵ برازنده‌ترین توزیع برای این داده‌ها می‌باشد ولی این توزیع به‌علت نداشتن فرم بسته از تابع چندکی تابع لورنتس و ضریب جینی آن به‌طور صریح قابل محاسبه نمی‌باشد. با توجه به این‌که p- مقدار بقیه‌ی توزیع‌ها صفر است بقیه‌ی توزیع‌ها برازنده‌ی داده‌ها نمی‌باشند، به هر حال می‌توان سراغ توزیعی رفت که کمترین مقدار آماره‌ی خی‌دو را دارد. پس تابع لورنتس و ضریب جینی را برای توزیع لگ نرمال که بعد از توزیع بتای تعمیم‌یافته‌ی نوع دوم، کمترین مقدار آماره‌ی خی دو را دارد محاسبه می‌کنیم. لازم به ذکر است که در مواردی که p- مقدار کمتر از ۱×۱۰^{-۱۶} بوده، صفر گزارش شده است.

با استفاده از روابط (۳) و (۴) تابع لورنتس و ضریب جینی برای توزیع لگ نرمال به‌صورت زیر است.

$$L(u)=\Phi\left(\Phi^{-۱}(u)-\sigma\right), \qquad ۰<u<۱$$

$$G=۲\Phi\left(\frac{\sigma}{\sqrt{۲}}\right)-۱=۰/۴۶۵$$

که $\Phi(\cdot)$ نمایانگر تابع توزیع تجمعی لگ نرمال است.

۴- معرفی فرم‌های تابعی لورنتس و برآورد آن‌ها

اگر تابع چگالی احتمال توزیع درامد را داشته باشیم با استفاده از رابطه‌ی (۳) معمولاً می‌توانیم تابع لورنتس متناظر آن را به دست آوریم. ورای موضوع یادشده مدل‌های پارامتری زیادی برای تقریب منحنی‌های لورنتس تجربی پیشنهاد شده است. حال به معرفی مدل‌های پیشنهادی چوتیکاپانیچ [۲] (تابع $L_۱$)، ارتگا و همکاران [۸] (تابع $L_۲$)، راسچ و همکاران [۹] (تابع $L_۳$)، سارابیا و همکاران [۱۰] (تابع $L_۴$)، کاکوانی [۵] (تابع $L_۵$) و ضرایب جینی آن‌ها می‌پردازیم.

$$L_1(u;k)=\frac{e^{ku}-1}{e^{k}-1},\quad k>0$$
$$G_1=\frac{(k-2)e^{k}+(k+2)}{k(e^{k}-1)},$$

$$L_2(u;\alpha,k)=u^{\alpha}\left[1-(1-u)^{k}\right],\quad \alpha\geq 0, 0<k\leq 1$$
$$G_2=1-2\left[B(\alpha+1,1)-B(\alpha+1,k+1)\right],$$

$$L_3(u;k,\gamma)=\left[1-(1-u)^{k}\right]^{\gamma},\quad \gamma\geq 1, 0<k\leq 1$$
$$G_3=1-\frac{2}{k}B\left(\frac{1}{k},\gamma+1\right),$$

$$L_4(u;k,\alpha,\gamma)=u^{\alpha}\left[1-(1-u)^{k}\right]^{\gamma},\quad \alpha\geq 0, \gamma\geq 1, 0<k\leq 1$$
$$G_4=1-2\sum_{i=0}^{\infty}\left[\frac{\Gamma(i-\gamma)}{\Gamma(i+1)\Gamma(-\gamma)}B(\alpha+1,ki+1)\right],$$

$$L_5(u;\alpha,\beta,\delta)=u-\alpha u^{\delta}(1-u)^{\beta},\quad \alpha>0, 0<\delta\leq 1, 0<\beta\leq 1$$
$$G_5=2\alpha B(\delta+1,\beta+1).$$

هرگاه $\gamma=1$ تابع L_4، L_2 را نتیجه می‌دهد، هرگاه $\alpha=0$ تابع L_4، L_3 را نتیجه می‌دهد و هرگاه $\alpha=\delta=1$ تابع L_5، L_2 با $\alpha=1$ و $\beta=k$ را نتیجه می‌دهد. همچنین $B(\cdot,\cdot)$ تابع بتاست.

۱-۴- براورد پارامترهای فرم‌های تابعی لورنتس به‌روش کمترین توان‌های دوم

اگر برای منحنی لورنتس یک فرم پارامتری در نظر بگیریم، می‌توانیم آن را با استفاده از روش کمترین توان‌های دوم براورد کنیم، به این مفهوم که مجموعه‌ی مقادیر $(u_1,L(u_1)),\ldots,(u_n,L(u_n))$ را حساب کرده و یک تابع $L(u;\theta)$ به داده‌ها برازش

می‌دهیم، حال باید تابع $\sum_{i=1}^{n}\left(L\left(u_i\right)-L\left(u_i;\theta\right)\right)^2$ بر حسب پارامترهای θ مینیمم شود. اگر فرم پارامتری یک تابع خطی باشد پارامترهای آن به روش‌های کلاسیک و در غیر این صورت به روش‌های عددی با کامپیوتر برآورد می‌گردد. از آن‌جا که اغلب فرم‌های تابعی لورنتس پارامتری توابع غیر خطی از پارامترها است برآوردهای کمترین توان‌های دوم پارامترها با روش‌های عددی محاسبه می‌گردد. در این‌جا از فرمان nls در نرم‌افزار R استفاده می‌کنیم.

بعد از برآورد پارامترهای فرم‌های تابعی لورنتس نیاز به معیارهایی است که از روی آن‌ها بهترین فرم تابعی لورنتس را تشخیص دهیم، که از میانگین مربعات خطا (MSE) و میانگین قدرمطلق خطا (MAE) به‌صورت زیر استفاده می‌کنیم.

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(L\left(u_i\right)-L\left(u_i;\hat{\theta}\right)\right)^2,$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|L\left(u_i\right)-L\left(u_i;\hat{\theta}\right)\right|,$$

که $L\left(u_i\right)$ منحنی لورنتس تجربی و $L\left(u;\hat{\theta}\right)$ منحنی لورنتس برازش‌شده است. هرچه مقادیر MSE و MAE کمتر باشد نشان‌دهنده‌ی بهتر بودن فرم تابعی لورنتس است. جدول ۲ برآورد پارامترهای فرم‌های تابعی لورنتس به‌روش کمترین توان‌های دوم، میانگین مربعات خطا، میانگین قدر مطلق خطا و ضریب جینی را نشان می‌دهد.

با توجه به جدول ۲ فرم تابعی پیشنهادی راسچ و همکاران (ردیف سوم جدول) با داشتن کمترین مقدار میانگین مربعات خطا و میانگین قدر مطلق خطای به‌ترتیب ۰/۰۰۰۰۰۷ و ۰/۰۰۲۱، بهترین فرم تابعی است. لازم به ذکر است که فرم تابعی پیشنهادی سارابیا (ردیف چهارم جدول) با مقدار $\alpha=0$ فرم تابعی پیشنهادی راسچ را نتیجه می‌دهد. همچنین با مقایسه‌ی مقادیر ضرایب جینی فرم‌های تابعی لورنتس با مقدار ضریب جینی تجربی که توسط رابطه‌ی (۲) محاسبه شده و برابر ۰/۴۳۱۸ است مشاهده می‌شود که مقدار ضریب جینی برآورد شده‌ی فرم‌های تابعی مختلف تفاوت چندانی ندارند و حدود ۰/۴۳ هستند.

جدول ۲- براورد پارامترهای فرم‌های تابعی لورنتس به داده‌های هزینه و درامد خانوار کل کشور به‌روش حداقل مربعات

ردیف	براورد پارامترها	میانگین مربعات خطا	میانگین قدرمطلق خطا	ضریب جینی
۱	$k=۲/۹۱۷$	۰/۰۰۰۸۹	۰/۰۲۲۵	۰/۴۲۸۷
۲	$\alpha=۰/۵۰۶,\ k=۰/۵۲۸$	۰/۰۰۰۰۲	۰/۰۰۳۸	۰/۴۳۴۲
۳	$k=۰/۶۱۲,\ \gamma=۱/۴۴۴$	۰/۰۰۰۰۰۷	۰/۰۰۲۱	۰/۴۳۳۴
۴	$\alpha=۰,\ k=۰/۶۱۲,\ \gamma=۱/۴۴۴$	۰/۰۰۰۰۰۷	۰/۰۰۲۱	۰/۴۳۳۴
۵	$\alpha=۰/۸۱۲,\ \beta=۰/۵۰۲,\ \delta=۱$	۰/۰۰۰۰۰۹	۰/۰۰۲۲	۰/۴۳۲۱

۲-۴- براورد پارامترهای فرم‌های تابعی لورنتس به‌روش ماکسیمم درستنمایی

روش دیگر براورد پارامترهای فرم‌های تابعی لورنتس به این صورت است که ابتدا داده‌ها را به K رده، رده‌بندی می‌کنیم. بنا بر این توزیع درامد داده‌ها به شکل $\left(u_k, L\left(u_k\right)\right)$، $k=۱,\ldots,K$ می‌باشد. در نظر می‌گیریم سهم درامد برابر $Q_k=L\left(u_k,\theta\right)-L\left(u_{k-۱},\theta\right)$ متغیرهای تصادفی با میانگین زیر باشند.

$$E\left(Q_k\right)=L\left(u_k,\theta\right)-L\left(u_{k-۱},\theta\right), \qquad (۵)$$

که بردار $Q=(Q_۱,\ldots,Q_K)'$ از توزیع دریکله با تابع چگالی احتمال زیر تولید شده است.

$$Q\sim f\left(q\mid\alpha\right)=\frac{\Gamma\left(\alpha_۱+\cdots+\alpha_K\right)}{\Gamma\left(\alpha_۱\right)\cdots\Gamma\left(\alpha_K\right)}q_۱^{\alpha_۱-۱}\ldots\ q_K^{\alpha_K-۱}, \qquad (۶)$$

که $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_K)'$ پارامترهای تابع چگالی دریکله و $\Gamma(\cdot)$ تابع گاماست.

همچنین در نظر می‌گیریم α تابعی از پارامترهای منحنی است و $\alpha_i=\lambda\left[L(u_i;\theta)-L(u_{i-1};\theta)\right]$ که λ پارامتر اضافی است. این تعریف برای α_i قابل قبول است زیرا امید ریاضی تابع توزیع دریکله برابر است با

$$
\begin{aligned}
E(Q_k)&=\frac{\alpha_k}{\alpha_1+\cdots+\alpha_K}\\
&=\frac{\lambda\left[L(u_k,\theta)-L(u_{k-1},\theta)\right]}{\lambda\sum_{k=1}^{K}\left[L(u_k,\theta)-L(u_{k-1},\theta)\right]}\\
&=L(u_k,\theta)-L(u_{k-1},\theta),
\end{aligned}
$$

که این برابر رابطه‌ی (۵) است. با استفاده از روابط (۵) و (۶) تابع چگالی احتمال برای Q به‌صورت زیر است.

$$
f(q\,|\,\phi)=\Gamma(\lambda)\prod_{k=1}^{K}\frac{q_k^{\lambda\left[L(u_k,\theta)-L(u_{k-1},\theta)\right]-1}}{\Gamma\left(\lambda\left[L(u_k,\theta)-L(u_{k-1},\theta)\right]\right)}.
$$

برآورد ماکسیمم درستنمایی برای $\phi=(\theta',\lambda)'$، بر مبنای ماکسیمم ساختن تابع $\ln$ که به‌صورت زیر است به دست می‌آید [۳].

$$
\begin{aligned}
\ln\left[f(q\,|\,\phi)\right]&=\ln\Gamma(\lambda)+\sum_{k=1}^{K}\left(\lambda\left[L(u_k,\theta)-L(u_{k-1},\theta)\right]-1\right)\times\ln q_k\\
&\quad-\sum_{k=1}^{K}\ln\Gamma(\lambda[L(u_k,\theta)-L(u_{k-1},\theta)])
\end{aligned}
\tag{۷}
$$

با قرار دادن فرم‌های تابعی لورنتس در رابطه‌ی (۷) و استفاده از روش‌های عددی، برآورد ماکسیمم درستنمایی پارامترهای فرم‌های تابعی به دست می‌آید. جدول ۳ برآورد پارامترهای فرم‌های تابعی لورنتس به‌روش ماکسیمم درستنمایی، میانگین مربعات خطا، میانگین قدر مطلق خطا و ضریب جینی را نشان می‌دهد.

جدول ۳- براورد پارامترهای فرم‌های تابعی لورنتس به داده‌های هزینه و درامد خانوار کل کشور به‌روش ماکسیمم درستنمایی

ردیف	براورد پارامترها	میانگین مربعات خطا	میانگین قدرمطلق خطا	ضریب جینی
۱	$k=۰/۳۰۸$	۰/۰۴۲۷۶	۰/۱۹۰۳	۰/۰۵۱۳
۲	$\alpha=۰/۳۴۷,\ k=۰/۵۶$	۰/۰۰۰۹۶	۰/۰۲۷۱	۰/۳۷۷۹
۳	$k=۰/۶۱۶,\ \gamma=۱/۳۱۵$	۰/۰۰۰۷۵	۰/۰۲۴۳	۰/۳۸۳۳
۴	$\alpha=۰/۰۰۰۸,\ k=۰/۶,\ \gamma=۱/۲۵۳$	۰/۰۰۱۱۵	۰/۰۳۰۴	۰/۳۷۱۱
۵	$\alpha=۰/۷۳۲,\ \beta=۰/۵۲۶,\ \delta=۱$	۰/۰۰۰۸۲	۰/۰۲۶۱	۰/۳۷۹۸

با توجه به جدول ۳ فرم تابعی پیشنهادی راسچ و همکاران (ردیف سوم جدول (تابع $L_۳$)) با داشتن کمترین مقدار میانگین مربعات خطا و میانگین قدر مطلق خطا به‌ترتیب ۰/۰۰۰۷۵ و ۰/۰۲۴۳ بهترین فرم تابعی است.

۵- نتیجه‌گیری

با توجه به نتایج ذکرشده برازنده‌ترین توزیع برای داده‌ها توزیع بتای تعمیم‌یافته‌ی نوع دوم است و به‌طور کلی بهترین فرم تابعی لورنتس فرم تابعی پیشنهادی راسچ و همکاران (تابع $L_۳$) است.

اگر برای مقایسه‌ی دو روش براورد ذکرشده برای فرم‌های تابعی از معیار $I=\sum_{k=۱}^{K} q_k \ln\left(\frac{q_k}{\hat{q}_k}\right)$ استفاده کنیم نتایج به‌صورت زیر می‌باشد.

روش براورد پارامترهای فرم‌های تابعی	تابع $L_۱$	تابع $L_۲$	تابع $L_۳$	تابع $L_۴$	تابع $L_۵$
روش حداقل مربعات	۰/۰۴۰۶	۰/۰۰۳	۰/۰۰۱	۰/۰۰۱	۰/۰۰۲۱
روش ماکسیمم درستنمایی	۰/۲۷۵	۰/۰۰۶۴	۰/۰۰۴۷	۰/۰۰۷۶	۰/۰۰۶۳

بنا بر این روش براورد کمترین توان‌های دوم با داشتن کمترین مقدار I در کلیه‌ی فرم‌های تابعی لورنتس بهترین روش است. همچنین از این معیار برای بهتر بودن فرم

تابعی لورنتس می‌توان استفاده کرد، همان‌طور که ملاحظه می‌شود تابع L_3 در هر دو روش برآورد کمترین مقدار I را داراست و نتایج همانند نتایجی است که با معیار MSE به دست آمد.

با توجه به حجم بسیار زیاد نمونه و این‌که در کوتاه‌مدت (دو یا سه سال) شاخص‌های اقتصادی مانند ضریب جینی ممکن است تغییرات محسوسی داشته باشند در صورت موجود بودن داده‌های سال‌های ۱۳۸۵ تا ۱۳۸۷، می‌توان منحنی لورنتس را برای این سال‌ها نیز برآورد نموده وضریب جینی متناظر آن‌ها را برآورد کرد.

سپاس‌گزاری

نویسنده‌ی اول این مقاله از حمایت پژوهشکده‌ی آمار و نویسنده‌ی دوم از حمایت قطب داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد تشکر می‌نمایند.

مرجع‌ها

[1] Bandourian, R.; McDonald, J.B.; Turley, R.S. (2002). *A comparison of parametric models of income distribution across countries and over time*. Luxembourg income study working paper. Dept. of Economics, No. 305.

[2] Chotikapanich, D. (1993). *A comparison of alternative functional forms for the Lorenz curve*. Economics Letters, 41, 129-138.

[3] Chotikapanich, D.; Griffiths, W.E. (2002). *Estimating Lorenz curves a Dirichlet distribution*. Journal of Business and Economics Statistics, 20, 290-295.

[4] Gini, C. (1912). *Variabilita' e mutabilita, studio Economicogiuridici*. Universita di Cagliari Anno III, Parte 2a, reprinted in C. 211-382.

[5] Kakwani, N.C. (1980). *On a class of poverty measures*. Econometrica, 48, 437-446.

[6] Kleiber, C.; Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. John Wiley.

[7] Lorenz, M.O. (1905). *Method of measuring the concentration of wealth*. Journal of the American Statistical Association, 9, 209-219.

[8] Ortega, P.; Fernandez, M.A.; Lodoux, M.; Garcia, A. (1991). *A new functional form for estimating the Lorenz curve*. Review of Incom and Wealth, 37,

447-452.

[9] Rasche, R.H.; Gaffney, J.; Koo, A.; Obst, N. (1980). *Functional forms for estimating the Lorenz curve*. Econometrica. Vol. 48, No. 4, 1061-1062.

[10] Sarabia, J.M; Castillo, E.; Slottje, D.J. (1999). *An ordered family of Lorenz curves*. Journal of Econometrics, 91, 43-60.

آزاده مجیری

دانشجوی فوق لیسانس آمار

استان خراسان جنوبی، بیرجند، کیلومتر ۵ جاده‌ی زاهدان، دانشگاه بیرجند، دانشکده‌ی علوم، گروه آمار.

پیام‌نگار: a.mojiri2008@gmail.com

غلامرضا محتشمی برزادران

دانشیار آمار

مشهد، میدان آزادی، دانشگاه فردوسی مشهد، دانشکده‌ی علوم ریاضی، گروه آمار.

پیام‌نگار: gmb1334@yahoo.com

یدا... واقعی

استادیار آمار

استان خراسان جنوبی، بیرجند، کیلومتر ۵ جاده‌ی زاهدان، دانشگاه بیرجند، دانشکده‌ی علوم، گروه آمار.

پیام‌نگار: ywaghei@yahoo.com