فصلنامه مدلسازی اقتصادی (سال هشتم، شماره ۲ «پیاپی ۲۶»، تابستان ۱۳۹۳، صفحات ۲۰-۱)

مدلسازی توزیع درآمد برای ایران: مقایسه الگوی داگوم با چند مدل منتخب

صادق بختیاری،[*] سجاد محموداوغلی[**]

تاریخ دریافت: ۹۲/۱۱/۲۹ تاریخ پذیرش: ۹۳/۰۳/۲۶

چکیده

هدف این پژوهش برآورد مدل‌های دوپارامتری وایبل و سه پارامتری بتا، لگ نرمال، گاما و داگوم به روش حداکثر درست‌نمایی (MLE) به صورت سالانه با استفاده از اطلاعات مربوط به هزینه و درآمد خانوارهای ایرانی برای سال‌های ۱۳۹۰-۱۳۶۱ به وسیله زیربرنامه محاسبه‌گر بسته‌ی VGAM در نرم افزار R بوده است. مقایسه برازش مدل‌های یاد شده به وسیله معیار اطلاعات آکائیک (AIC) صورت گرفته است. نتایج نشان می‌دهد. در دوره‌ی ۱۳۹۰-۱۳۶۱ با وجود فراز و نشیب‌های مقادیر این شاخص در میان خانوارهای کشور، میزان ضریب جینی روندی کاهنده داشته یعنی در واقع شدت نسبی نابرابری درآمد در کشور کاهنده، ولی میزان کاهش آن بسیار محدود بود. هم‌چنین براساس معیار اطلاعات آکائیک و هم‌چنین نمودارهای حاصل از توابع چگالی احتمال مشخص شد که تابع توزیع داگوم یک برازش‌گر خوب است. مقادیر برآورد شده پارامترهای بتا و دلتا در طول این سال‌ها روند صعودی و پارامتر آلفا روند نزولی دارد.

طبقه‌بندی JEL: O15, C16, D63

واژگان کلیدی: توزیع درآمد، توزیع داگوم، حداکثر درست‌نمایی.

[*] استاد دانشگاه آزاد اسلامی واحد خوراسگان، گروه اقتصاد، اصفهان، ایران(نویسنده‌ی مسئول)، پست الکترونیکی:
bakhtiari_sadegh@yahoo.com

[**] کارشناس ارشد اقتصاد، پست الکترونیکی:
soghli13@gmail.com

۱. مقدمه

در سال ۱۹۷۰ کامیلو داگوم[۱] در جستجوی یک مدل آماری دقیق و درخور از توزیع درآمد تجربی و توزیع ثروت، کار خود را آغاز کرد. استفاده از توزیع‌های کلاسیک توزیع پارتو (توسعه یافته توسط اقتصاددان و جامعه شناس ایتالیایی ویلفردو پارتو[۲] در اواخر قرن ۱۹ (پارتو، ۱۸۹۵، ۱۸۹۶، ۱۸۹۷) و توزیع لگ نرمال (توسط مهندس فرانسوی رابرت گیبرات[۳] (۱۹۳۱) برای چنین داده‌هایی کافی نبود. داگوم در جستجوی یک مدل انعطاف‌پذیر با دم سنگین در توزیع‌های ثروت و درآمد تجربی بود. پژوهش‌هایی با تغییر توزیع لگ- لجستیک (داگوم، ۱۹۹۰) و تعمیم توزیع قبلی به وسیله فیسک[۴] (۱۹۶۱) انجام شده بود، داگوم متوجه شد که یک پارامتر اضافی مورد نیاز است. این موضوع به توزیع داگوم نوع I، یک توزیع سه پارامتری و تعمیم چهار پارامتری منجر شد (داگوم، ۱۹۹۰ و ۱۹۸۰). داگوم در سال ۱۹۸۹، برای واحدهای اقتصادی با ثروت خالص منفی، مدل تک پارامتری لاپلاس را مورد بررسی قرار داد. از تلفیق (ترکیب محدب) مدل نوع دوم داگوم و مدل تک پارامتری لاپلاس، داگوم مدلی به دست آورد که قابلیت کاربرد برای توزیع ثروت خالص با برد $R = (-\infty, \infty)$ را دارد. این خصوصیتی است که به مدل داگوم برجستگی خاصی می‌دهد زیرا دیگر مدل‌ها عموماً از این قابلیت محروم می‌باشند. در پژوهش حاضر، همانند بسیاری از بررسی‌های توزیع درآمد در کشورهای در حال توسعه، از هزینه‌های مصرفی به جای درآمد استفاده شده است بنابراین با توجه به این موارد این پژوهش درصدد است تا مدل داگوم را با مدل‌های دوپارامتری وایبل و سه پارامتری بتا، لگ‌نرمال، گاما بر اساس معیار اطلاعات آکائیک (AIC) مقایسه و پارامترهای آن را به روش حداکثر درست‌نمایی برآورد کند.

۲. ادبیات موضوع

چوتیکی پانیچ[۵] و همکارانش (۲۰۱۰) و سورو[۶] (۱۹۷۰) از مدل‌های سه پارامتری برای بررسی توزیع درآمدی رایج استفاده کردند نتایج نشان داد که توزیع سینگ- مادالا نیز نسبت به توزیع‌های

---

[1] Camilo Dagum
[2] Vilfredo Pareto
[3] Gibrat
[4] Fisk
[5] Chotikapanich
[6] Thurow

دو پارامتری بهتر بوده و مورد استفاده قرار می‌گیرد. یافته‌های گرتل[1] و همکاران (۲۰۰۰) نشان داد که تنها ریشه کن کردن بیکاری نمی‌تواند نسبت جینی را بهبود ببخشد، چون تاثیر آن به نسبت ضعیف است. تحقیقات آنها نشان داد که در طول دوره ۲۰۰۰-۱۹۹۲ پارامتر آلفا به میزان ٥٤ درصد افزایش یافته است و پارامتر بتا به میزان ۷۸ درصد کاهش داشته اما پارامتر دلتا ۲۳ درصد افزایش داشته است. در نتیجه نسبت ضریب جینی به میزان ٤/۳ درصد افزایش خالص داشته است. در ایران نیز بررسی اهداف مرتبط با توزیع درآمد بعد از انقلاب اسلامی حاکی از آن است که سیاست‌گذاران و برنامه‌ریزان اقتصادی، در راستای سیاست‌های کلی نظام، سند چشم‌انداز بیست ساله کشور، قوانین برنامه‌های پنج‌ساله توسعه و قوانین سنواتی بودجه، مبنای فعالیت‌های خود را بر پایه رویکرد عمل‌گرایی و در جهت توسعه عدالت اجتماعی و توزیع درآمد و ثروت برای اقشار محروم قرار داده‌اند. از طرفی بیش‌تر مطالعات انجام شده در زمینه‌ی توزیع درآمد در ایران، بر اساس اطلاعات آماری گروه‌بندی شده است. به عبارتی دیگر، با طبقه‌بندی داده‌های حاصل از نمونه‌گیری در مناطق شهری و روستایی در گروه‌های چندگانه، شاخص‌های نابرابری متفاوتی محاسبه شده است.

بختیاری و همکاران (۱۳۸۰) با ارزیابی وضعیت هزینه‌های مصرفی مناطق شهری و روستایی استان اصفهان در دوره ۷۲-۱۳٦۸ بیان کردند که انواع هزینه‌های مصرفی در طول این دوره افزایش یافته است که این افزایش هزینه، در مناطق شهری با بدتر شدن وضعیت تغذیه‌ای همراه بوده است. نتایج مطالعه ابونوری، خوشکار و حیدری (۱۳۸۵) و خسروی نژاد (۱۳۹۱) حاکی از آن است که نابرابری در مناطق روستایی از مناطق شهری بیش‌تر بوده است. نابرابری در مناطق شهری و روستایی شهرستان بندر لنگه در سال انتهایی اجرای برنامه سوم توسعه اقتصادی- اجتماعی (۱۳۸۳) نسبت به سال ابتدای اجرای این برنامه (۱۳۷۹) افزایش یافته است. هم‌چنین ابونوری (۱۹۸۷) در رساله دکتری خود در دانشگاه کنت انگلستان، تحت عنوان تجزیه و تحلیل ریاضی‌آماری توزیع درآمد و اثر نفت بر نابرابری‌های اقتصادی در کشورهای عضو اوپک به نتایج مشابهی دست یافتند.

[1] Gertel

۳. روش‌شناسی

برای یافتن توزیع مناسب داده‌های درآمد 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 برآورد پارامترهای توزیع‌ها، مثل روش ماکسیمم درست‌نمایی، پارامترهای این توزیع‌ها را برآورد کرد. در بیش‌تر توزیع‌های متداول یا کلاسیک برآوردگر 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. برای تحلیل داده‌ها از نرم افزار R استفاده شده است. نرم‌افزار R به واسطه قابلیت‌های بالا، رایگان بودن آن و امکان دسترسی به توابع کتابخانه‌ای سایر محققان اقصی نقاط دنیا، در بین سایر نرم افزارهای برنامه‌نویسی از جایگاه خاصی برخوردار است.(هژبرکیانی و مرادی، ۱۳۸۷) در این قسمت به بررسی پنج تابع توزیع درآمدی داگوم، بتا، گاما، لگ نرمال و وایبل پرداخته می‌شود، و سپس به نحوه برآورد پارامترهای هریک از توابع به روش حداکثر درست‌نمایی[1] اشاره می‌شود.

داگوم با ارایه یک توصیف نظری و براساس ویژگی‌های مشاهده شده و برگیری از کشش درآمدی، یک تابع توزیع خاصی معرفی نمود. به طور کلی، کشش درآمدی از تابع توزیع تجمعی(CDF) با توجه به نقطه آلفا (α) از F(x)، به عنوان یک تابع کاهشی یک‌نواخت از F(x) است. نمایش ریاضی این استدلال، معادله دیفرانسیلی با سه یا چهار پارامتر به صورت زیر است:

$$\frac{d\ln[F(x)-\alpha]}{d\ln x} = \beta_1\left[1-\left(\frac{F(x)-\alpha}{1-\alpha}\right)^{\beta_2}\right], x>0, (\beta_1,\beta_2)>0 \qquad (۱)$$

با حل این رابطه برای CDF خواهیم داشت

$$F(x) = \alpha + \frac{(1-\alpha)}{(1+\lambda x^{-\delta})^{\beta}} \;, (\beta,\delta,\lambda)>0, \beta = 1/\beta_2 \;, \delta = \beta_1\beta_2 \qquad (۲)$$

که با مشتق‌گیری از F(x) ، تابع چگالی احتمال (Pdf)، f(x) به دست می‌آید:

$$F'(x) = f(x)$$

که معادله (۲) ، برای δ های بزگتر از صفر تعریف شده است (داگوم، ۱۹۷۷).

---

[1] Maximum Likelihood Method

سه نسخه از مدل داگوم برای محاسبه مفروضات خاص در مورد جمعیت دریافت کننده درآمد وجود دارد.

داگوم رابطه (۱) از تابع توزیع تجمعی (CDF) مربوط به مبداء زمانی که α=0، که این حالت نشان دهنده بهترین وجه رفتار دستمزد بگیران (افراد شاغل) است.

$$F(x) = \frac{1}{(1+\lambda x^{-\delta})^{\beta}} \tag{۳}$$

از آنجا که α=0 است داگوم نوع اول شامل سه پارامتر است که توصیف و توزیع با دریافت کنندگان درآمدی که درآمد صفر دارند (هیچ درآمدی دریافت نمی‌کنند) شروع می‌شود. پارامترهای β و δ مقیاس‌های آزاد هستند که مقدار برابری را توضیح می‌دهد (داگوم، ۱۹۸۳).

داگوم نوع (۲) و (۳) که شامل پارامتر چهارم α است که معنای اقتصادی ویژه‌ای دارد، پس معادله زیر

$$F(x) = \alpha + \frac{(1-\alpha)}{(1+\lambda x^{-\delta})^{\beta}} \tag{٤}$$

شامل چهار پارامتر تابع توزیع تجمعی است. برای داگوم نوع دوم ۱> α<۰ تعریف می‌شود که واحد سنجش درآمدی α با درآمد صفر یا منفی است. بنابراین α پارامتر نابرابری است، در حالی که λ پارامتر مقیاس است که به زمان و یا مقایسه فاصله توزیع درآمد در واحدهای مختلف پولی بیان شده است. داگوم نوع سوم مختص توزیع درآمد نمونه از دریافت کنندگان کل درآمد (دستمزد بگیران به علاوه کسانی که از اموال خود درآمد به دست می‌آورند) است. به این دلیل که درآمد جمع آوری شده از جمعیت، با درآمد اولیه مثبت آغاز می‌شود. بنابراین نسبت جینی مرتبط با مدل داگوم از طریق فرمول زیر محاسبه می‌شود:

$$G = (2\alpha - 1) + (1-\alpha)\frac{\Gamma(\beta)\Gamma\left(2\beta+\frac{1}{\delta}\right)}{\Gamma(2\beta)\Gamma\left(\beta+\frac{1}{\delta}\right)} \tag{٥}$$

که در آن $\Gamma(0)$ همان تابع گامای کامل مشخص شده در تابع داگوم است. نسبت جینی یک تابع فزاینده از α است بنابراین نسبت جینی مربوط به مدل داگوم با افزایش ارزش β و δ به صفر میل می‌کند که نشان دهنده برابری کامل است و با کاهش β و δ، تمایل به یک دارد که نشان‌دهنده نابرابری کامل است. به این معناست زمانی که ارزش β و δ افزایش می‌یابد، توزیع درآمد بهبود

می‌یابد. (دانچلی، ۱۹۸٦) به طور خاص، با حرکت از قسمت مرکزی نمودار به سمت راست و به سمت انتهای بالایی نمودار، به این مفهوم است که جمعی از افراد طبقه متوسط، درآمد بالاتری را دریافت می‌کنند، که این با افزایش ارزش δ، منعکس می‌شود [$\frac{\partial Gini}{\partial \delta}<0$]. به طور مشابه افزایش ارزش β، موجب بهبود رفاه مردم کم درآمد میشود [$\frac{\partial Gini}{\partial \beta}<0$]. پارامترهای بتا (β) و دلتا (δ) در معادله(٥)، به طورخلاصه اطلاعاتی در مورد چگالی، یا توزیع فراوانی افراد در نیروی کار در سطوح مختلف درآمدی را نشان می‌دهد. در حقیقت با توجه به معادله (۲)، با تغییر بتا (β) و دلتا (δ) و ثابت نگه داشتن پارامتر دیگری مشخص می‌شود که: i) هرگونه افزایش (کاهش) در بتا (β) با ثابت نگه داشتن دلتا (δ)، میزان برابری افزایش (کاهش) می‌یابد. هم‌چنین هرگونه افزایش (کاهش) در دلتا (δ) با ثابت نگه داشتن پارامتر بتا (β)، میزان برابری افزایش (کاهش) می‌یابد. ii) زمانی که بتا (β) و دلتا مخالف علامت هم هستند، اثر مشترک بتا (β) و دلتا (δ) به طور خلاصه موجب بهبود نسبی (زوال) در میزان برابری خواهد شد وقتی که بتا (β) افزایش (کاهش) می‌یابد و موجب زوال جزیی (بهبود) می‌شود زمانی که دلتا (δ) کاهش (افزایش) می‌یابد. بنابراین پارمترهای موجود در ضریب جینی به رفتار پارامترهای مهم برابری در اقتصاد، بستگی دارند.

توزیع وایبل یکی از مهم‌ترین توزیع‌ها در مسایل برنامه‌ریزی‌های اقتصادی و هم‌چنین در رشته مهندسی است. تابع تجمعی توزیع وایبل به صورت زیر است:

$$F(x,\mu,\sigma,p) = 1 - \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^p\right] \quad (٦)$$

که درآنμ پارامتر مکان، σ پارامتر مقیاس و p پارامتر شکل می‌باشد. هم‌چنین تابع چگالی احتمال توزیع وایبل به صورت زیر می‌باشد:

$$F(x,\mu,\sigma,p) = \frac{p}{\sigma^p}(x-\mu)^{p-1}\exp\left[-\left(\frac{x-\mu}{\sigma}\right)^p\right] \quad (٧)$$

متغیر تصادفی x دارای توزیع لگ نرمال است اگر log(x) دارای توزیع نرمال باشد. تابع چگالی احتمال برای متغیر تصادفی از توزیع سه پارامتری لگ نرمال به صورت زیر است

$$F(x,\mu,\sigma,p) = \frac{1}{(x-\mu).\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(\ln(x-y)-\mu)^2\right] \quad (٨)$$

که در آن $\pm > 0$ ، $-\infty < \mu < \infty$ ، $0 \leq ¥ < x$. پارامترهای σ، μ و γ پارامترهای توزیع هستند. وقتی مقدار پارامتر γ صفر باشد، توزیع سه پارامتری به توزیع دو پارامتری لگ نرمال تبدیل می‌شود. توزیع سه پارامتری لگ نرمال به وسیله یوان (۱۹۳۳)، کوهن (۱۹۵۱)، هیل (۱۹۶۳)، هارتر و moور (۱۹٦٦)، وینگو (۱۹۸٦) و ... مورد مطالعه قرار گرفته است.

تابع چگالی احتمال گاما با پارامترهای α ، β و γ به صورت زیر است:

$$F(x,\mu,\sigma,p) = \frac{1}{\gamma\Gamma(\beta)}\left(\frac{x-\alpha}{\gamma}\right)^{\beta-1}\exp\left[-\left(\frac{x-\alpha}{\gamma}\right)\right], (x \geq \alpha, \beta > 0, \gamma > 0) \qquad (۹)$$

که در آن α, β و γ به ترتیب پارامترهای مکان، شکل و مقیاس هستند. در صورتی که تابع چگالی احتمال سه پارامتری گاما به صورت اریب منفی باشد در این صورت تابع چگالی احتمال را برای این توزیع می‌توان به صورت زیر نوشت:

$$f(x:\mu,\sigma,\lambda) = \frac{1}{\alpha\lambda\Gamma(\lambda^{-2})}\left[\lambda^{-2}\left\{1+\lambda\left(\frac{x-\mu}{\sigma}\right)\right\}\right]^{\lambda^{-2}-1}.\exp\left[-\lambda^{-2}\left\{1+\lambda\left(\frac{x-\mu}{\sigma}\right)\right\}\right], (\sigma > 0, \lambda \neq 0) \qquad (۱۰)$$

که در آن $\gamma = \sigma|\lambda|, \alpha = \mu - \sigma\lambda^{-1}, \beta = \lambda^{-2}$ می‌باشد. بنابراین پارامترهای σ، μ و λ به ترتیب پارامتر مقیاس، مکان و شکل هستند.

و در نهایت تابع چگالی احتمال توزیع بتای نوع دوم به صورت زیر می‌باشد:

$$f(y:a,b,a_1,a_2) = \frac{|a|y^{a_1a_2-1}(1-(y/b)^{a})^{a_2-1}}{b^{a_1a_2}B(a_1,a_2)} \qquad (۱۱)$$

که در آن $b, a_1, a_2 > 0$ می‌باشد.

برآورد به روش حداکثر درست‌نمایی

داگوم (۱۹۷۷) در یک دوره زمانی وقتی اطلاعات فردی به ندرت در دسترس بودند، به صورت زیر حداقل سازی کرد:

$$\sum_{i=1}^{n}\{F_n(x_i) - [1+(x_i/b)^{-a}]^{-p}\}^2 \qquad (۱۲)$$

معیار حداقل مربعات غیرخطی در فاصله بین تابع توزیع تجمعی (CDF) F_n تجربی و CDF تقریب زده شده از توزیع داگوم مبنا قرار گرفت. علاوه بر این نوع رگرسیون برآورد شده با استفاده

از کشش رابطه (۱) به بعد، توسط استوپا (۱۹۹۵) بررسی شد. امروزه اکثر محققان از برآورد حداکثر احتمال (ML) استفاده می‌کنند. دو مورد برجسته، داده های گروه‌بندی شده و داده‌های فردی مورد نیاز است. تا همین اواخر، تنها داده‌های گروه بندی شده در دسترس بود و احتمال $L(\theta)$ که در آن $\theta = (a,b,p)$ یک احتمال چندجمله‌ای است با (فرض داده‌های مستقل):

$$L(\theta) = \coprod_{j=1}^{m}\{F(x_j) - F(x_{j-1})\},\ x_0 = 0, x_m = \infty \tag{۱۳}$$

بنا به ساختار این احتمال همیشه از بالا کراندار است.

در نگاهی به سی‌امین سالگرد کمک‌های داگوم به نظر می‌رسد او دوباره یکی از نمونه‌های اولیه تجربی خود را برای درآمد خانوارهای آمریکایی در سال ۱۹٦۹ به خود اختصاص می‌دهد.

نمودار هیستوگرام استخراجی کریستین کلیبر از طریق برآورد تقریبی داگوم نوع یک به وسیله گروه‌بندی حداکثر احتمال را نشان می‌دهد. نتایج برآورد$\hat{b} =$ ۱٤/٤۸، $\hat{a} =$ ٤/۲۷۳ و $\hat{p} =$۰/۳٦ است و این مقادیر سازگار با مقادیر برآورد شده به وسیله داگوم از طریق حداقل مربعات غیر خطی است.

با افزایش ریز داده‌های در دسترس، برآورد احتمال برای مشاهدات فردی توجه بیش‌تری را جلب می‌کند، و این وضعیت بیش‌تر مستلزم: لگ احتمال $l(\theta) = Log\ L(\theta)$ یک نمونه کامل تصادفی از اندازه n هست:

ست:

$$l(a,b,p) = n\log a + n\log p + (ap - 1)\sum_{i=1}^{n}\log x_i - nap\log b - (p + 1)\sum_{i=1}^{n}\log\{1 + (x_i/b)^a\} \tag{۱٤}$$

بازده معادلات احتمال

$$\frac{n}{a} + p\sum_{i=1}^{n}\log(x_i/b) = (p+1)\sum_{i=1}^{n}\frac{\log(x_i/b)}{1+(x_i/b)^a} \tag{۱۵}$$

$$np = (p+1)\sum_{i=1}^{n}\frac{1}{1+(b/x_i)^a} \tag{۱٦}$$

$$\frac{n}{a} + a\sum_{i=1}^{n}\log(x_i/b) = \sum_{i=1}^{n}\log\{1 + (x_i/b)^a\} \tag{۱۷}$$

که باید عددی محاسبه شوند.

با این حال، برآورد احتمال این خانواده بدون مشکلات نیست: با توجه به توزیع logX، توزیع لجستیک تعمیم یافته، شائو[1] (۲۰۰۲) نشان می‌دهد که ممکن است MLE وجود داشته باشد و اگر نداشته باشد، به اصطلاح مشکل مدل تعبیه شده رخ می‌دهد. اجازه دادن به پارامترهای خاص، که به مقادیر مرزی خود تمایل دارند، یک توزیع با پارامتر جزیی را به وجود می‌آورد. مفاهیمی هستند که رفتار احتمال باید برای آن‌ها به دقت در کار تجربی بررسی شوند. این امر می‌تواند جالب باشد که برای تعیین این که تا چه حد این اشکال، در برنامه‌های کاربردی از داده های درآمد که در آن انعطاف پذیری کامل خانواده داگوم مورد نیاز نیست، به وجود می‌آید.

ظاهراً غافل از این مشکلات، دومانسکی و یرژیچاک[2] (۱۹۹۸) یک مطالعه مشابه عملکرد MLE را ارایه کردند. به نظر می رسد که نمونه نسبتا بزرگ برای برآورد پارامترهای شکل a,p مورد نیاز بود، در حالی که برآورد قابل اعتماد از پارامتر مقیاس به نظر می‌رسد نیاز به نمونه‌های حتی بزرگ‌تر دارد.

روش‌های برآورد حداکثر درست‌نمایی برای تابع توزیع سه پارامتری گاما به وسیله جانسون و کوتز (۱۹۷۰)، کوهن و نورگارد (۱۹۷۷)، هارتر و موور (۱۹۶۵)، بوومن و شنتون (۱۹۸۸) و ... بیان شده است. بنابراین برای معادله احتمال گاما داریم:

$$
\begin{aligned}
&\frac{\partial \log L}{\partial \sigma} = -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{1-z_i^2}{1+\lambda z_i} = 0\\
&\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n}\frac{\lambda-z_i}{1+\lambda z_i} = 0 \qquad (۱۸)\\
&\frac{\partial \log L}{\partial \lambda} = n\left\{\frac{2}{\lambda^3}\Psi\left(\frac{1}{\lambda}\right) + 2\log\lambda\right\} + \sum_{i=1}^{n}\left\{-\frac{2}{\lambda^3}\log(1+\lambda z_i) + \frac{\lambda(1+z_i^2+2z_i)}{\lambda^2(1+\lambda z_i)}\right\} = 0
\end{aligned}
$$

که در آن $z_i = (x_i - \mu)/\sigma$ و ψ تابع سای است.

اگر $Y_1 \dots Y_n$ متغیرهای تصادفی از توزیع بتا باشند در این صورت تابع لگاریتم احتمال برای این مشاهدات به صورت زیر خواهد بود:

---
[1] Shao
[2] Domanski and Jedrzejczak

$$
\begin{aligned}
\ln L(\alpha_1, \alpha_2, a, b) &= \sum_{i=1}^{N} \ln L_i(\alpha_1, \alpha_2, a, b) = \sum_{i=1}^{N} \ln f(Y : \alpha_1, \alpha_2, a, b) \\
&= \sum_{i=1}^{N} \ln \frac{(Y_i - a)^{\alpha_1 - 1} (b - Y_i)^{\alpha_2 - 1}}{(b - a)^{\alpha_1 + \alpha_2 - 1} B(\alpha_1, \alpha_2)} \\
&= (\alpha_1 - 1) \sum_{i=1}^{N} \ln(Y_i - a) + (\alpha_2 - 1) \sum_{i=1}^{N} \ln(b - Y_i) \\
&\quad - N \ln B(\alpha_1, \alpha_2) - N (\alpha_1 + \alpha_2 - 1) \ln(b - a)
\end{aligned}
\tag{۱۹}
$$

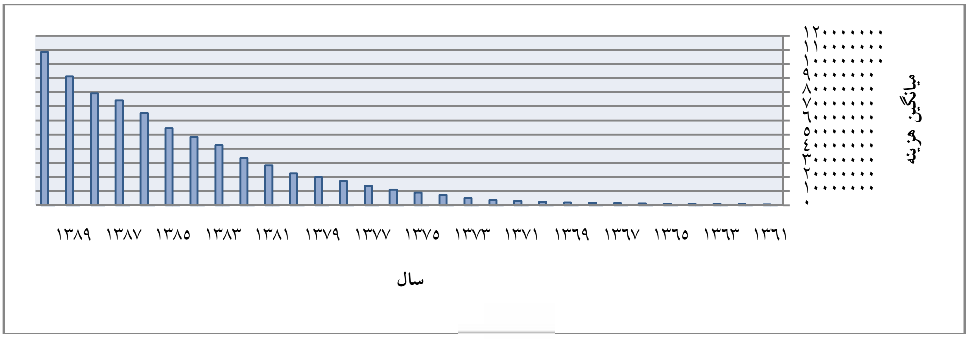
با مشتق‌گیری نسبت به پارمترهای توزیع و برابر صفر قرار دادن این روابط خواهیم داشت:

$$
\begin{aligned}
\frac{\partial \ln L(\alpha_1, \alpha_2, a, b|Y)}{\partial \alpha_1} &= \sum_{i=1}^{N} \ln(Y_i - a) - N(-\Psi(\alpha_1 + \alpha_2) + \Psi(\alpha_1)) - N \ln(b - a) = 0 \\
\frac{\partial \ln L(\alpha_1, \alpha_2, a, b|Y)}{\partial \alpha_2} &= \sum_{i=1}^{N} \ln(b - Y_i) - N(-\Psi(\alpha_1 + \alpha_2) + \Psi(\alpha_2)) - N \ln(b - a) = 0 \\
\frac{\partial \ln L(\alpha_1, \alpha_2, a, b|Y)}{\partial a} &= -(\alpha_1 - 1) \sum_{i=1}^{N} \frac{1}{(Y_i - \alpha)} + N(\alpha_1 + \alpha_2 - 1) \frac{1}{b - a} = 0 \\
\frac{\partial \ln L(\alpha_1, \alpha_2, a, b|Y)}{\partial b} &= -(\alpha_1 - 1) \sum_{i=1}^{N} \frac{1}{(b - Y_i)} - N(\alpha_1 + \alpha_2 - 1) \frac{1}{b - a} \\
&= 0
\end{aligned}
\tag{۲۰}
$$

٤. یافته‌های تحقیق

در سال‌های ۱۳٦۱ تا ۱۳۹۰ بررسی هزینه و درآمد خانوارهای ایرانی در شهرهای مختلف و با مراجعه به خانوارهای نمونه‌گیری شده، انجام شده است. در سال ۱۳۹۰، متوسط هزینه ناخالص سالانه یک خانوار ایرانی برابر ۱۰۸۳٤٤٥۹٤ ریال گزارش شده است که نسبت به سال مبدا این پژوهش یعنی سال ۱۳٦۱ (متوسط هزینه ناخالص سالانه یک خانوار ایرانی برای این سال برابر ٦۹٤٤۲۸ ریال است) به میزان ۱۰۷٦٥۰۱٦٦ ریال افزایش داشته است هم‌چنین متوسط تعداد افراد خانوار برای سال ابتدا این تحقیق کاهش ۱/٦۹ نفری نسبت به سال ۱۳۹۰ داشته است. شکل زیر روند حرکت مقادیر متوسط هزینه ناخالص سالانه یک خانوار ایرانی از سال ۱۳٦۱ تا سال ۱۳۹۰ را نشان می‌دهد.

شکل ۱. میانگین هزینه متوسط هر خانوار



در مدل توزیع داگوم مقادیر پارامترهای (α,β,δ) از روش حداکثر درست‌نمایی با استفاده از زیربرنامه محاسبه‌گر بسته‌ی VGAM[۱] در نرم افزار R[۲] به دست آمده است.

جدول (۲) برآورد مقادیر آلفا، بتا و دلتا را با استفاده از معادله (٤) برای نمونه‌ای از هزینه خانوار ایرانی برای سال‌های ۱۳۹۰–۱۳۶۱ و با استفاده از معادله پارامتری (۵)، نسبت جینی به روش ماکسیمم درست‌نمایی را نشان می‌دهد.

در جدول (۲) مقادیر مربوط به پارامترها و ضریب جینی برآورد شده که مقادیر مربوط به پارامتر آلفا بیانگر میزان نابرابری در بین سال‌های مورد مطالعه دارد، که در بین سال‌های ۱۳۶۱ و ۱۳۹۰ کم‌ترین میزان نابرابری مربوط به سال ۱۳۹۰ و بیش‌ترین میزان نابرابری مربوط به سال ۱۳۶۱ می‌باشد، همان طور که مشاهده می‌شود برای این سال‌ها ضریب جینی نیز به ترتیب کم‌ترین و بیش‌ترین مقدار را داراست و این حاکی از وجود رابطه مستقیم بین پارامتر آلفا و ضریب جینی دارد. هم‌چنین مقادیر مربوط به بتا که نشان دهنده میزان برابری است برای سال ۱۳۶۱ کم‌ترین و برای سال ۱۳۹۰ بیش‌ترین مقدار را داراست بنابراین از برآورد پارامترهای یاد شده مشخص می‌شود که توزیع درآمد در سال ۱۳۶۱ بسیار نابرابرتر از سال ۱۳۹۰ است که این نابرابری تقریباً به میزان ۱۰ درصد در فاصله سال‌های مذکور کاهش داشته است. همچنین مقادیر حاصل از $\delta\beta$، چون از عدد یک بزرگ‌تر هستند $\delta\beta > 1$، پس یک تابع توزیع حاصل تک نمایی خواهد بود.

---
[1] Vector Generalized Linear and Additive Models (VGAM)

[2] بنیان‌گذاران پروژه R آقایان Ross Ihaka و Robert Gentelman بودند. وجه تسمیه این زبان نیز ابتدای نام این دو نفر است.

جدول ۲. برآورد پارامترهای مدل داگوم نوع دوم و ضریب جینی به روش ماکسیمم درست‌نمایی برای ایران در سال‌های ۱۳۹۰–۱۳۶۱

سال	برآورد پارامترها			βδ	نسبت جینی (Gini ratio)	مجموع مربعات خطا SSE(Pdf)	آماره Kolmogrov-smirnov	
	Alpha	Beta	Delta				K-S	P-Value
۱۳۶۱	۰/۰۵۸۸۱	۰/۵۸۹۵۳	۵/۳۴۶۷۰	۲/۹۶۴۶	۰/۴۹۲۱	۰/۰۶۷	۰/۰۰۴	۰/۹۲۸
۱۳۶۲	۰/۰۵۱۱۹۲	۰/۸۱۴۳۸	۳/۴۹۳۹۹	۲/۸۴۵۴	۰/۴۸۰۷	۰/۱۷۸	۰/۰۳۲	۰/۱۵۹
۱۳۶۳	۰/۰۴۷۹۸	۱/۵۸۷۹۹	۳/۸۲۰۷۳	۶/۰۶۷۲	۰/۴۶۳۸	۰/۲۹۱	۰/۰۰۷	۰/۳۵۶
۱۳۶۴	۰/۰۴۹۳۲	۱/۱۸۳۸۵	۲/۱۹۴۴۱	۲/۵۹۷۸	۰/۴۷۰۶	۰/۰۱۷	۰/۰۰۹	۰/۰۹۸
۱۳۶۵	۰/۰۴۶۳۸	۱/۰۵۸۹۰	۱/۰۸۸۱۶	۱/۱۵۲۲	۰/۴۵۷۱	۰/۳۶۸	۰/۰۶۵	۰/۲۴۱
۱۳۶۶	۰/۰۴۵۶۵	۱/۸۲۸۲۴	۲/۹۳۱۷۱	۵/۳۵۹۸	۰/۴۳۳۰	۰/۱۵۶	۰/۰۱۶	۰/۱۹۸
۱۳۶۷	۰/۰۴۲۴۸	۲/۴۴۲۵۷	۴/۵۴۵۰۶	۱۱/۱۰۱۱	۰/۴۱۹۵	۰/۰۷۹	۰/۰۰۴	۰/۷۸۷
۱۳۶۸	۰/۰۴۶۰۶	۱/۶۸۹۱۳	۲/۵۱۷۳۷	۴/۲۵۲۱	۰/۴۳۰۹	۰/۱۶۹	۰/۰۰۳	۰/۰۸۶
۱۳۶۹	۰/۰۴۳۵۸	۲/۰۲۸۹۵	۱/۱۰۳۰۹	۲/۲۳۸۱	۰/۴۲۲۷	۰/۱۵۴	۰/۰۳۲	۰/۲۵۶
۱۳۷۰	۰/۰۴۱۵۶	۱/۹۲۸۸۵	۲/۱۷۳۲۸	۴/۱۹۱۹	۰/۴۱۷۱	۰/۲۳۴	۰/۰۶۱	۰/۳۱۶
۱۳۷۱	۰/۰۴۰۱۲	۲/۵۵۰۲۹	۳/۸۳۲۰۹	۹/۷۷۲۹	۰/۴۰۳۴	۰/۰۹۸	۰/۰۰۸	۰/۱۴۷
۱۳۷۲	۰/۰۳۹۸۵	۲/۶۱۲۴۰	۴/۵۶۴۱۷	۱۱/۹۲۳	۰/۴۰۱۲	۰/۰۵۴	۰/۰۰۱	۰/۱۵۸
۱۳۷۳	۰/۰۴۳۸۴	۲/۳۱۴۹۸	۱/۰۲۷۵۴	۲/۳۷۸۷	۰/۴۱۰۸	۰/۰۶۶	۰/۰۱۳	۰/۲۷۹
۱۳۷۴	۰/۰۴۷۵۹	۱/۷۴۶۲۳	۱/۴۸۰۰۶	۲/۵۸۴۵	۰/۴۳۲۶	۰/۱۱۹	۰/۰۷۷	۰/۳۵۶
۱۳۷۵	۰/۰۴۱۰۲	۲/۵۹۶۱۲	۴/۹۱۲۸۸	۱۲/۷۵۴۴	۰/۴۰۹۴	۰/۲۵۶	۰/۰۰۲	۰/۷۸۴
۱۳۷۶	۰/۰۴۵۲۳	۲/۰۰۲۸۱	۱/۵۱۰۹۹	۳/۰۲۶۲	۰/۴۲۴۰	۰/۱۹۸	۰/۰۵۶	۰/۹۸۱
۱۳۷۷	۰/۰۴۲۱۸	۲/۴۹۷۱۷	۱/۴۴۰۲۶	۳/۵۹۶۵	۰/۴۱۸۵	۰/۲۲۴	۰/۰۶۹	۰/۱۱۲
۱۳۷۸	۰/۰۴۸۵۴	۱/۶۸۲۷۹	۱/۲۳۰۹۲	۲/۰۷۱۳	۰/۴۳۱۲	۰/۱۲۵	۰/۰۰۴	۰/۰۷۵
۱۳۷۹	۰/۰۴۱۱۲	۲/۸۶۱۸۰	۴/۲۷۹۰۱	۱۲/۲۴۵	۰/۴۰۵۱	۰/۰۲۳	۰/۰۱۰	۰/۰۶۱
۱۳۸۰	۰/۰۴۳۹۸	۲/۴۵۲۸۳	۱/۴۲۱۴۸	۳/۴۸۶۶	۰/۴۱۴۹	۰/۱۱۴	۰/۰۰۹	۰/۸۷۶
۱۳۸۱	۰/۰۴۶۸۷	۱/۸۸۸۴۷	۶/۱۱۹۸۱	۱۱/۵۵۷۸	۰/۴۲۰۸	۰/۰۷۴	۰/۰۰۵	۰/۴۸۹
۱۳۸۲	۰/۰۳۹۷۹	۲/۷۳۹۸۱	۰/۸۷۱۸۸	۲/۳۸۸۷	۰/۴۰۶۵	۰/۲۲۳	۰/۱۵۹	۰/۳۲۴
۱۳۸۳	۰/۰۴۲۲۶	۲/۶۰۰۲۱	۵/۷۹۰۷۵	۱۵/۰۵۷۷	۰/۴۱۱۳	۰/۱۹۷	۰/۰۴۸	۰/۳۰۷
۱۳۸۴	۰/۰۴۰۰۲	۲/۷۷۶۴۲	۲/۲۷۹۶۶	۶/۳۲۹۲	۰/۴۰۸۷	۰/۳۷۵	۰/۱۲۴	۰/۰۶۹
۱۳۸۵	۰/۰۴۵۱۲	۲/۴۵۶۴۹	۱/۴۴۰۸۹	۳/۵۳۹۵	۰/۴۱۱۹	۰/۱۹۵	۰/۰۱۹	۰/۴۵۱
۱۳۸۶	۰/۰۳۹۷۴	۲/۸۵۴۱۳	۱/۷۲۳۵۴	۴/۹۱۹۲	۰/۴۰۵۷	۰/۱۷۱	۰/۰۳۸	۰/۳۷۸
۱۳۸۷	۰/۰۳۸۹۵	۲/۹۰۲۱۶	۳/۶۰۹۲۳	۱۰/۴۷۴۵	۰/۳۹۹۵	۰/۰۲۷	۰/۰۰۶	۰/۳۲۰
۱۳۸۸	۰/۰۴۱۲۷	۲/۷۵۴۲۵	۳/۰۹۴۹۶	۸/۵۲۴۲	۰/۴۰۷۳	۰/۰۳۰	۰/۰۱۱	۰/۲۲۱
۱۳۸۹	۰/۰۳۸۴۲	۲/۹۷۵۴۴	۲/۶۸۸۹۸	۸/۰۰۰۸	۰/۳۹۴۱	۰/۱۸۸	۰/۰۷۲	۰/۱۷۳
۱۳۹۰	۰/۰۳۷۹۹	۳/۰۹۸۵۴	۵/۱۲۰۷۳	۱۵/۸۶۶۸	۰/۳۹۰۷	۰/۱۶۴	۰/۰۰۱	۰/۸۷۴

منبع: یافته‌های محقق

بعد از برآورد ضریب جینی و پارامترهای فرم تابعی داگوم به معیارهایی برای تشخیص این که آیا مدل ما یک مدل مناسب برای برازش بوده یا نه، نیاز داریم، برای این منظور از دو آزمون مجموع مربعات خطا (SSE) و آزمون کلموگروف – اسمیرونوف (K-S) استفاده کردیم، در این تحقیق چون مقادیر مربوط به آماره K-S کم‌تر از آماره SSE است بنابراین آماره K-S یک انتخاب عالی برای تشخیص خوبی برازش برای مدل داگوم می‌باشد لازم به توضیح است منظور از مجموع مربعات خطا همان رابطه $SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ است که در آن y_i مقادیر مشاهده شده و $\hat{y}_i$ مقادیر برآورد شده هستند. هم‌چنین آزمون کلموگروف– اسمیرونوف برای تطابق توزیع، احتمال‌های تجمعی مقادیر در مجموعه داده‌ها را با احتمال‌های تجمعی همان مقادیر در یک توزیع نظری خاص مقایسه می‌کند. اگر اختلاف آن به قدر کافی بزرگ باشد، این آزمون نشان خواهد داد که داده‌ها با یکی از توزیع‌های نظری مورد نظر تطابق ندارد. در این آزمون اگر معیار تصمیم (P-value) کم‌تر از ۵ درصد باشد فرض صفر رد می‌شود یعنی داده‌ها نمی‌توانند از یک توزیع خاص باشند همان طور که در جدول (۲) مشاهده می‌کنیم مقادیر مربوط به این آماره نشان از این دارد که مدل پیشنهادی داگوم در این تحقیق یک برازش‌گر خوب است.

برای یافتن توزیع مناسب داده‌های درآمد می‌بایست چند توزیع که شکلی مشابه توزیع تجربی داده‌ها دارند به داده ها برازش داده شود و با مقایسه‌ی آن‌ها توزیع مناسب داده‌ها را انتخاب کرد. برای این منظور لازم است با استفاده از روش‌های برآورد پارامترهای توزیع‌ها، مثل روش ماکسیمم درست‌نمایی، پارامترهای این توزیع‌ها را برآورد کرد. در بیش‌تر توزیع‌های متداول یا کلاسیک برآوردگر ماکسیمم درست‌نمایی پارامترها منحصر به فرد بوده و با مشتق‌گیری معمولی از تابع درست‌نمایی یا تابع لگاریتم درست‌نمایی نسبت به پارامترهای مورد نظر به دست می‌آید. لازم به ذکر است که 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 نرم‌افزار R استفاده می‌کنیم. بنابراین در این پژوهش توزیع داگوم را با چهار توزیع وایبل، بتا، گاما و لگ نرمال مقایسه می‌کنیم، با تو جه به این که پارامترهای مدل داگوم در قسمت اول برآورد شده‌اند در اینجا ما به برآورد پارامترهای چهار تابع توزیع دیگر با استفاده از روش حداکثر درست‌نمایی می‌پردازیم، سپس برای مقایسه پنج تابع توزیع با یکدیگر از معیار آکائیک استفاده می‌کنیم.

جدول ۳. برآورد پارامترهای مدل‌های مختلف از روش حداکثر درست‌نمایی

توزیع	توزیع بتا				توزیع گاما			توزیع لگ نرمال			توزیع وایبل	
پارامتر	$\alpha 1$	$\alpha 2$	α	b	α	β	γ	σ	μ	γ	α	β
۱۳٦۱	۱/٤۰	۱/۱۰	۰/۰۱	۱/۲۰	۱/٤۰	۱/۱۰	۰/۰۱	۰/۸۰	۰/۱۲	-۰/۰٦	۱/٤۰	۱/۵۰
۱۳٦۲	۱/۸۱	۱/٤۱	۰/۰۳	۱/۰۱	۱/۷۲	۰/۷۳	۰/۰۳	۰/۷۱	۰/۰۱	-۰/۰۳	۱/۷۱	۱/٤۱
۱۳٦۳	۱/۷۰	۲/۷۰	۰/۰۹	۱/۹۰	۱/۹۱	۰/٦٤	۰/۰۸	۰/٦۳	۰/۰۸	-۰/۰۷	۱/٤۳	۱/۸۰
۱۳٦٤	۱/۸۳	۲/۰۱	۰/۰٤	۱/۳۰	۱/۷۳	۰/۷۲	۰/۰۵	۰/٦۷	۰/۰۸	-۰/۰۸	۱/۷۲	۱/٤۱
۱۳٦۵	۲/۰۱	۲/۱۰	۰/۰۱	۱/۲۰	۲/۰۰	۰/۵٦	۰/۱۲	۰/۵۹	۰/۰۷	-۰/۰۸	۱/۹۰	۱/۳۰
۱۳٦٦	۱/۸٤	۱/۷۰	۰/۰٤	۱/۱۰	۱/۸۰	۰/۷۱	۰/۰٤	۰/٦۸	۰/۰٤	-۰/۰٤	۱/۷۱	۱/٤۳
۱۳٦۷	۱/۸۰	۲/٤۰	۰/۰۳	۱/۷۰	۱/٦۰	۰/۷۹	۰/۰۳	۰/۷۱	۰/۰۵	-۰/۰۵	۱/٦۰	۱/٤۱
۱۳٦۸	۱/۸۷	۲/۳۰	۰/۱۳	۱/٦۰	۱/۵۷	۰/۸۹	۰/۰۱	۰/۷۵	۰/۰۳	۰/۰۲	۱/۷۵	۱/۵۰
۱۳٦۹	۱/۹۲	۲/۲۰	۰/۰۵	۱/٤۰	۱/۹۱	۰/٦٦	۰/۰٦	۰/٦۳	۰/۱۲	۰/۰۹	۱/۸۱	۱/٤۰
۱۳۷۰	۲/۰۱	۱/٦۰	۰/۱۲	۸/۹۰	۲/۱۰	۰/۵۳	۰/۱۱	۰/٦۳	۰/۰۳	۰/۰۲	۲/۰۱	۱/۳۱
۱۳۷۱	۱/۹۰	۱/۹۰	۰/۰۹	۱/۲۰	۲/۰۱	۰/۵۷	۰/۰۸	۰/۵۹	۰/۱۰	-۰/۰۱	۱/۹۰	۱/۳۲
۱۳۷۲	۲/۰۱	۱/۲۰	۰/۱۱	۱/۱۰	۲/۱۳	۰/۸۸	۰/۱۱	۰/٦۳	۰/٤۸	۰/۰۵	۲/۱۰	۱/۹۰
۱۳۷۳	۱/٤٤	۳/٤۰	۰/۰٦	۳/۵۰	۱/۵۲	۰/۹۷	۰/۰٦	۰/۷۷	۰/۱۲	-۰/۰۱	۱/٦۰	۱/٦٤
۱۳۷٤	۲/۲۰	۱/۲۰	۰/۰۳	۷/۲۰	۲/۳۰	۰/۵۵	۰/۰۲	۰/۵۸	۰/۲۰	-۰/۱٤	۱/۸۰	۱/٤۰
۱۳۷۵	۱/۳۱	۱/۱۰	۰/۰۷	۱/٤۰	۱/٤٤	۱/۲۰	۰/۰۷	۰/۸۱	۰/۱۸	-۰/۰۳	۱/۵۰	۱/۷۰
۱۳۷٦	۱/۸۰	۱/۸۰	۰/۰۹	۱/۳۰	۱/۷٦	۰/۷٤	۰/۰۹	۰/۷۱	۰/۰۵	۰/۰۱	۱/۸۲	۱/۵۲
۱۳۷۷	۲/۰۰	۲/۳۰	۰/۰٦	۱/٤۰	۱/۹۲	۰/٦٤	۰/۰٦	۰/٦٤	۰۹.۰	-۰/۰٦	۱/۸۵	۱/٤۰
۱۳۷۸	۱/۹۳	۱/۵۰	۰/۰۷	۱/۱۰	۱/۸۷	۰/۷۵	۰/۰۷	۰/٦۵	۰/۱۸	-۰/۰۸	۱/۷٤	۱/۵۳
۱۳۷۹	۱/۲۰	۷/۹	۰/۰٦	٤/۸۰	۲/۲۰	۰/۵۹	۰/۰٦	۰/۵۸	۰/۲۲	-۰/۱۲	۱/۹	۱/۵۰
۱۳۸۰	۲/۰	۲/۲۰	۰/۱۱	۱/۳۰	۱/۷۰	۰/۷۲	۰/۱۲	۰/٦٦	۰/۱۰	-۰/۰٤	۱/۸۰	۱/۵۰
۱۳۸۱	۲/۰۲	۱/۹۰	۰/۰۸	۵/٦۰	۲/۰۰	۰/۵۹	۰/۰۷	۰/۵۸	۰/۱۸	-۰/۱٤	۱/۸٦	۱/٤۱
۱۳۸۲	۲/٤۰	۱/۲۰	۰/۰٦	٦/٤۰	۲/٤	۰/۵٤	۰/۰٦	۰/۵۹	۰/۱۸۲	-۰/۰۷	۲/۰۰	۱/٤۹
۱۳۸۳	۲/۲۰	۲/٤	۰/۱۳	۱/۳۰	۲/۰۳	۰/٦۱	۰/۱٤	۰/٦۲	۰/۱۱	-۰/۰۱	۲/۰۳	۱/۵۰
۱۳۸٤	۲/۰۵	۱/۷۰	۰/۰٤	۱/۱۰	۲/۳	۰/۵۷	۰/۰۳	۰/۵۸	۰/۱۹	-۰/۱۲	۱/۹۰	۱/٤۰
۱۳۸۵	۲/۰٤	۱/۷۰	۰/۰۷	۱/۲۰	۱/۹	۰/۷۱	۰/۰۷	۰/٦٦	۰/۱۵	-۰/۰۲	۱/۹۳	۱/۵۲
۱۳۸٦	۲/٤۱	۲/٤۰	۰/۰۵	۱/۳۰	۲/۲۰	۰/۵۹	۰/۰٦	۰/۵۹	۰/۲۱	-/۱۰	۱/۹۰	۱/۵۵
۱۳۸۷	۱/۹۰	۱/۰۳	۰/۱۲	۱/۳۳	۲/۰٤	۰/٦۳	۰/۱۲	۰/٦۰	۰/۱۷	-۰/۰٤	۲/۰۱	۱/۵۰
۱۳۸۸	۲/۰۱	۱/۳۰	۰/۰۸	۸/٤۰	۲/۱۰	۰/٦۱	۰/۰۸	۰/٦۱	۰/۱۵	-۰/۰۵	۲/۹٦	۱/۵۰
۱۳۸۹	۱/۸۰	۱/٦۰	۰/۱۷	۱/۱۰	۱/۹۰	۰/٦۲	۰/۱۷	۰/٦۳	۰/۰۷	-۰/۰۱	۲/۰۲	۱/٤۲
۱۳۹۰	۲/۷۰	۱/۹۰	۰/۰۹	۷/۷۰	۲/۸۰	۰/۳۹	۰/۰۹	۰/۵۲	۰/۰۹	-۰/۰۷	۲/۲۰	۱/۳۰

منبع: یافته‌های محقق

جدول ٤. مقایسه برازش مدل‌ها

توزیع	$D = (\alpha, \beta, \delta)$		$B = (\alpha_1, \alpha_2, a, b)$		$G = (\alpha, \beta, \gamma)$		$LN = (\delta, \mu, \gamma)$		$W = (\alpha, \beta)$	
سال	**LLK***	**AIC**	**LLK**	**AIC**	**LLK**	**AIC**	**LLK**	**AIC**	**LLK**	**AIC**
١٣٦١	–٥٢/٦٠	١٠٥/٣٤	–٥٢/٧٧	١٠٥/٦٣	–٥٢/٧٨	١٠٥/٦٥	–٥٢/٧٧	١٠٥/٥٤	–٥٢/٨٠	١٠٥/٨
١٣٦٢	–٦٧/٣٢	١٣٤/٦٤	–٦٧/٨٠	١٣٥/٦	–٦٨/١٤	١٣٦/٢٨	–٦٨/٧٠	١٣٧/٤	–٦٨/٧٩	١٣٧/٥
١٣٦٣	–٧١/٦٠	١٤٣/٣٢	–٧١/٩٨	١٤٣/٩٦	–٧٢/١١	١٤٤/٢٢	–٧٢/٥٤	١٤٥/٠٨	–٧٢/٣٧	١٤٤/٧
١٣٦٤	–٨٨/٤١	١٨٣/٠٤	–٨٨/٦٧	١٧٧/٣٤	–٨٨/٧٣	١٧٧/٤٦	–٨٨/٥٥	١٧٧/١	–٨٨/٩٠	١٧٧/٨
١٣٦٥	–٩٣/٢٠	١٨٦/٤٤	–٩٣/٣٩	١٨٦/٧٨	–٩٤/١٩	١٨٨/٣٨	–٩٣/٨٠	١٨٧/٦	–٩٤/٧٠	١٨٩/٤
١٣٦٦	–٩٨/١٠	١٩٦/٢٢	–٩٨/٥١	١٩٧/٠٢	–٩٩/٠٧	١٩٨/١٤	–٩٨/٦٨	١٩٧/٣٦	–٩٩/٢٨	١٩٨/٥
١٣٦٧	–١٠٢/٢١	٢٠٤/٤٢	–١٠٢/٤٩	٢٠٤/٩٨	–١٠٤/٢١	٢٠٨/٤٢	–١٠٤/٠٨	٢٠٨/١٦	–١٠٣/٤٥	٢٠٦/٠
١٣٦٨	–١٠٩/٦٤	٢١٩/٢٨	–١٠٩/٩٧	٢١٩/٩٤	–١١٠/٤٣	٢٢٠/٨٦	–١١٠/٢٠	٢٢٠/٤	–١١٠/٨٢	٢٢١/٦
١٣٦٩	–١١٨/٧٠	٢٣٧/٤	–١١٩/٠٦	٢٣٨/١٢	–١١٩/٥٥	٢٣٩/١	–١١٩/٨٣	٢٣٩/٦٦	–١٢٠/٢٤	٢٤٠/٤
١٣٧٠	–١٢٧/٣٤	٢٥٤/٦٨	–١٢٧/٩٥	٢٥٥/٩	–١٢٨/٠٤	٢٥٦/٠٨	–١٢٨/٤٥	٢٥٦/٩	–١٢٨/٧٠	٢٥٧/٤
١٣٧١	–١٣٤/٦٠	٢٦٩/٢	–١٣٤/٨٣	٢٦٩/٦٦	–١٣٥/٦٧	٢٧١/٣٤	–١٣٥/٠٩	٢٧٠/١٨	–١٣٥/١٩	٢٧٠/٣
١٣٧٢	١٤٢/٨١	٢٨٥/٦٢	–١٤٢/٩٩	٢٨٥/٩٨	–١٤٣/٤٥	٢٨٦/٩	–١٤٣/٣٨	٢٧٦/٧٦	–١٤٣/٦٦	٢٨٧/٣
١٣٧٣	–١٥٢/٣٢	٣٠٢/٦٤	–١٥١/٨٠	٣٠٣/٦	–١٥٢/٢٣	٣٠٤/٤٦	–١٥٢/٦٥	٣٠٥/٣	–١٥٣/٠٧	٣٠٦/١
١٣٧٤	–١٦٦/٤١	٣٣٢/٨٢	–١٦٦/٧٢	٣٣٣/٤٤	–١٦٧/٢٤	٣٣٤/٤٨	–١٦٧/٤٤	٣٣٤/٨٨	–١٦٧/٨٦	٣٣٥/٧
١٣٧٥	–١٧٨/٦٥	٣٥٧/٣	–١٧٨/٦٩	٣٥٧/٣٨	–١٨٧/٨٠	٣٥٧/٦	–١٧٨/٧٧	٣٥٧/٥٤	–١٧٩/٠٣	٣٥٨/٠
١٣٧٦	–١٨٩/٢٤	٣٧٨/٤٨	–١٨٩/٧٥	٣٧٩/٥	١٩٠/٠٧	٣٨٠/١٤	–١٨٩/٩١	٣٧٩/٨٢	–١٨٩/٥٥	٣٧٩/١
١٣٧٧	–١٩٧/٧٣	٣٩٥/٤٦	–١٩٨/٠٨	٣٩٦/١٦	–١٩٨/٨٤	٣٩٠/٦٨	–١٩٨/٥٥	٣٩٧/١	–١٩٧/٧٦	٣٩٧/٥
١٣٧٨	–٢٠٨/٢٢	٤١٦/٤٤	–٢٠٨/٦١	٤١٧/٢٢	–٢٠٩/٠٨	٤١٨/١٦	–٢٠٨/٧٩	٤١٧/٥٨	–٢٠٩/١١	٤١٨/٢
١٣٧٩	–٢٢١/٣٤	٤٤٢/٦٨	–٢٢١/٨٦	٤٤٣/٧٢	–٢٢٢/٧٤	٤٤٥/٤٨	–٢٢١/٩٧	٤٤٣/٩٤	–٢٢٢/٠٤	٤٤٤/٠
١٣٨٠	–٢٣٧/٦٥	٤٧٥/٣	–٢٣٧/٨٨	٤٧٥/٧٦	–٢٣٨/٤٢	٤٧٦/٨٤	–٢٣٨/٠٩	٤٧٦/١٨	–٢٣٨/٤٧	٤٧٦/٩
١٣٨١	–٢٤٤/٣١	٤٨٨/٦٢	–٢٤٤/٥٦	٤٨٩/١٢	–٢٤٤/٨٧	٤٨٩/٧٤	–٢٤٤/٥١	٤٨٩/٠٢	–٢٤٤/٧٥	٥٨٩/٥
١٣٨٢	–٢٥٨/٩١	٥١٧/٨٢	–٢٥٩/١١	٥١٨/٢٢	–٢٥٩/٧٤	٥١٩/٤٨	–٢٥٩/٤٢	٥١٨/٨٤	–٢٥٩/٦٦	٥١٩/٣
١٣٨٣	–٢٧٣/١٢	٥٤٦/٢٤	–٢٧٣/٢٠	٥٤٦/٤	–٢٧٣/٨٨	٥٧٤/٧٦	–٢٧٣/٥٠	٥٤٧	–٢٧٤/٠١	٥٤٨/٠
١٣٨٤	–٢٨٩/٧٥	٥٧٩/٥	–٢٨٩/٨١	٥٧٩/٦٢	–٢٩٠/٢٧	٥٨٠/٥٤	–٢٩٠/١٢	٥٨٠/٢٤	–٢٩٩/٥٣	٥٨١/٠
١٣٨٥	–٣٠٧/٤٤	٦١٤/٨٨	–٣٠٧/٨٥	٦١٥/٧	–٣٠٧/٩٤	٦١٥/٨٨	–٣٠٧/٨٢	٦١٥/٦٤	–٣٠٨/١٠	٦١٦/٢
١٣٨٦	–٣١٦/٥٢	٦٣٦/٠٤	–٣١٦/٧٩	٦٣٣/٥٨	–٣١٧/٢٣	٦٣٤/٤٦	–٣١٧/٠٨	٦٣٤/١٦	–٣١٧/٦١	٦٣٥/٢
١٣٨٧	–٣٢٨/٢٢	٦٥٦/٤٤	–٣٢٨/٥٧	٦٥٧/١٤	–٣٢٩/١٣	٦٥٨/٢٦	–٣٢٨/٨٤	٦٢٧/٦٨	–٣٢٩/٠٦	٦٥٨/١
١٣٨٨	–٣٤١/٣٧	٦٨٢/٧٤	–٣٤١/٨٢	٦٨٣/٦٤	–٣٤٢/٢٨	٦٨٤/٥٦	–٣٤٢/١٢	٦٨٤/٢٤	–٣٤٢/٥٥	٦٨٥/١
١٣٨٩	–٣٦٧/٤١	٧٣٤/٨٢	–٣٦٧/٧٧	٧٣٥/٥٤	–٣٦٨/٠٩	٧٣٦/١٨	–٣٦٨/٢٧	٧٣٦/٥٤	–٣٦٨/١٩	٧٣٦/٣
١٣٩٠	–٣٨٥/٧٢	٧٧١/٤٤	–٣٨٦/٠٢	٧٧٢/٠٤	–٣٨٦/٧١	٧٧٣/٤٢	–٣٨٦/٥٩	٧٧٣/١٨	–٣٨٦/٧٥	٧٧٣/٥

منبع: یافته‌های محقق. *منظور از LLK همان loglikelihood است.

جدول (٤) شامل مقادیر توابع احتمال لگاریتمی (عنوان loglikelihood در جدول (٤) برای تمام سال‌ها و تمام توزیع‌های مورد بررسی همراه با ارزش آمار معیار آکائیک (AIC) که به صورت زیر تعریف شده است:

(مقدار LLK) * ۲ – (تعداد پارامترها) * ۲ = معیار آکائیک (AIC)

با توجه به جدول برازش مدل‌ها مشاهده می شود که تمامی مقادیر AIC تقریباً مشابه هستند ولی ارزش تابع توزیع داگوم در تمام سال‌های مورد تجزیه و تحلیل، کمتر از بقیه توابع توزیع است. بنابراین می توان نتیجه گرفت که تابع توزیع داگوم به عنوان بهترین برازش‌گر در بین توابع مختلف توزیع برای ایران است.

می‌توان به این نتیجه دست یافت که تابع توزیع داگوم برای همه سال‌های مورد بررسی مقدار بزرگ‌تری از تابع چگالی احتمال را نسبت به توابع توزیع بتا، گاما، لگ نرمال و وایبل نشان می‌دهد بنابراین این موضوع که مدل داگوم مقدار بزرگ‌تری برای ضریب جینی در همه سال‌ها نسبت به سایر مدل‌ها نشان می‌دهد با استفاده از این نتایج مورد تایید قرار می‌گیرد، هم‌چنین همان طور که از شکل‌های بالا قابل مشاهده است هم‌چنین نتایج نشان داد که نابرابری درآمدی در سال ۱۳۶۱ به مراتب بیش‌تر از سال ۱۳۹۰ بوده است. بنابراین تابع چگالی احتمال داگوم برای همه سال‌ها بالاتر از تابع چگالی احتمال بتا، گاما، لگ نرمال و وایبل است.

۵. نتیجه‌گیری

با توجه به این که آمار یکی از مهم‌ترین علوم کاربردی است که با سایر رشته‌های علمی از جمله اقتصاد مرتبط می‌باشد. از این رو، در سال‌های اخیر، علاقه زیادی برای تحقیقات در زمینه مدل‌های پارامتری توزیع درآمد به وجود آمده است. مدل‌های احتمال مربوط به توزیع درآمد، برای ارزیابی استانداردهای سطح زندگی کل مردم یک کشور و هم‌چنین برای مقایسه استاندارد سطح زندگی طبقات اجتماعی و یا مناطق مختلف یک کشور ارایه شده‌اند. بنابراین برای ایجاد یک مدل احتمال ارایه یک تابع توزیع نظری با مشخصه توزیع فراوانی تجربی برای انتخاب روش مناسب تخمین پارامترهای مدل، ضروری است. بنابراین تجزیه و تحلیل آماری توزیع درآمد جمعیت نشان دهنده زمینه لازم برای تصمیم‌گیری در مورد بودجه و سیاست‌های اجتماعی است (باستوروف ۲۰۰۶) با

توجه به این موارد مدل پیشنهاد شده داگوم بسیاری از خواص مربوط به توزیع در آمد همانند خصوصیات رفتاری مدل در چارچوب اقتصادی، هم‌گرایی به قانون پارتو و اهمیت اقتصادی پارامترها، را برآورد می‌کند. (لاتوره ۱۹۸۹) از این مدل هم چنین برای تشریح توزیع اندازه مشارکت در کسب و کار نیز به صورت کاملاً موفق استفاده شده است. مدل داگوم موفقیت‌های زیادی را در مطالعات انجام شده بر روی توزیع درآمد و دستمزد و هم‌چنین توزیع ثروت به دست آورده است که مشخصات و ویژگی‌های این مدل به طور گسترده‌ای توسط نویسندگان مختلف تجزیه و تحلیل شده است. (کلیبر و کوتز ۲۰۰۳) بنابراین در دنیای برابری کامل، هر یک درصد افزایش در جمعیت انباشته، سبب افزایش یک درصد درآمد انباشته می‌شود. تجزیه و تحلیل نظری و تجربی بر روی توزیع درآمد نشان داده‌اند که نرخ رشد درآمد حاصل از درآمد تجمعی سریع‌تر از نرخ رشد جمعیت دریافت کننده درآمد افزایش می‌یابند. و این بدان معناست که کشش درآمد تجمعی کاهش می‌یابد (داگوم ۱۹۹۰).

هدف از این پژوهش تحلیل نابرابری درآمد برای ایران به روش پارامتریک با استفاده از مدل داگوم و مقایسه تابع توزیع داگوم با توابع توزیعی چون بتا، گاما، لگ نرمال، و وایبل بود. نتایج حاصل از این پژوهش بر اساس برآورد ضریب جینی و پارامترهای توابع توزیع با استفاده از روش حداکثر درست‌نمایی در بین سال‌های ۱۳٦۱ تا ۱۳۹۰ نشان دادند که نابرابری درآمدی در ایران در بین سال‌های ۱۳٦۱ و ۱۳۹۰ در حال نوسان بوده است. در یک نتیجه‌گیری کلی از مقادیر ضریب جینی در کشور نتیجه گرفته می‌شود که در دوره‌ی ۱۳۹۰–۱۳٦۱ با وجود فراز و نشیب‌های مقادیر این شاخص در میان خانوارهای کشور، میزان ضریب جینی روندی کاهنده داشته یعنی در واقع شدت نسبی نابرابری درآمد در کشور کاهنده، ولی میزان کاهش آن بسیار محدود بود. هم‌چنین براساس معیار اطلاعات آکائیک و هم‌چنین نمودارهای حاصل از توابع چگالی احتمال مشخص شد که تابع توزیع داگوم یک برازش‌گر خوب است. مقادیر برآورد شده پارامترهای بتا و دلتا در طول این سال‌ها روند صعودی و پارامتر آلفا روند نزولی دارد.

منابع

- ابونوری، اسمعیل، خوشکار، آرش، حیدری، حسین (۱۳۸۵). بررسی تحولات توزیع درآمد در شهرستان بندر لنگه طی برنامه سوم توسعه اقتصادی- اجتماعی. مجموعه مقالات اولین همایش توسعه شهرستان بندر لنگه قابلیت و راهکارها، دانشگاه آزاد اسلامی واحد بندر لنگه.
- بختیاری، صادق، نصراللهی، خدیجه، عمادزاده، مصطفی (۱۳۸۰). تحلیلی از وضعیت توزیع درآمد (هزینه) در استان اصفهان (۷۲ – ۱۳۶۸). *برنامه و بودجه*، ٦ (۹ و ۱۰): ۸۱ – ۵۱.
- خسروی‌نژاد، علی اکبر (۱۳۹۱). برآورد فقر و شاخص‌های فقر در مناطق شهری و روستایی، *فصلنامه مدل‌سازی اقتصادی*، ٦ (۲): ۳۹–٦۰.
- هژبرکیانی، کامبیز، مرادی، علیرضا (۱۳۸۷). نرم افزار R: محیط برنامه‌نویسی برای تحلیل‌های اقتصادسنجی و سری‌های زمانی، *فصلنامه مدل‌سازی اقتصادی*،۲ (۵): ۱٦۳–۱۸٦.

- Abounoori, E. (1987). Mathematico-statistical analysis of distribution of income and effect of oil on economic inequality within OPEC countries.
- Bartosova, J. (2006). Logarithmic-normal model of household income distribution in the Czech Republic after 1990. Forum statisticum slovacum, Slovak Statistical and Demographical Society, Bratislava, 3: 3-10.
- Chotikapanich, D. W. E., & Griffiths, D. S. P., & Valencia V. (2010). Global income distributions and inequality, 1993 and 2000: Incorporating country-level inequality modeled with beta distributions. Forthcoming in the review of economics and statistics.
- Cohen, A. (1951). Estimating parameters of logarithmic normal distributions by maximum likelihood, *Journal of the American Statistical Association*, 46: 206-212..
- Dagum, C. (1983). Income distribution models, in S. Kotz, N. L. Johnson and C. Read (eds.) Encyclopedia of Statistical Sciences, vol. 4, JohnWiley, New York.
- Dagum, C. (1990). Generation and properties of income distribution functions. In C. Dagum and Zenga, Eds. Income and wealth distribution, inequality and poverty, Heildeberg, Springer Verlag.
- Dancelli, L. (1986). Tendenza alla massima ed alla minima concentrazione nel modelo di distribuzione del redito di Dagum. *In Scritti in Honore di Francesco Brambilla*, 1: 249-267.
- Domanski, C., & Jedrzejczak, A. (1998). Maximum likelihood estimation of the dagum model parameters. *International Advances in Economic Research*, 4: 243–252.
- Fisk, P. R. (1961). The graduation of income distributions. *Econometrica*, 29:171–185.

- Gertel, H. R., & Giuliodori, R. F., & Rodríguez, A., & Paula F. A. (2001). Unemployment and income distribution analysis: New evidences using a agum Parametric income distribution model, facultad de ciencias económicas, reunión annual de la aaep, buenos aires.
- Gibrat, R. (1931). Les inegalites économiques, Paris, librairie du recueil sirey.
- Harter, H.L., & Moore, A.L. (1966). Local-maximum-likelihood estimation of the parameter of three-parameter lognormal population from complete and censored samples, *Journal of the American Statistical Society*, 61: 842—851.
- Hill, M.B. (1963). The three-parameter log-normal distribution and bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association*, 58: 112-120.
- Kleiber, C. & Kotz, S. (2003). Statistical size distribution in economics and actuarial sciences, London: Cambridge University Press.
- Kotz, S., & Johnson, N. L., & Read, C. (1983). Encyclopedia of statistical sciences, John Wiley, New York.
- Latorre, G. (1989). Asymptotic distributions of indices of concentration: Empirical erification and application, in: Studies in contemporary economics. income and wealth istribution, inequality and poverty, C. Dagum, M. Zenga (Eds), Springer-Verlag, Berlin.
- Latorre, G. (1988). Propriet`a campionarie del modello di dagum per la distribuzione dei redditi, *Statistica*, 48: 15–27.
- Majumder, A., & Chakravarty, S. R. (1990). Distribution of personal income: Development of a New Model and Its Application to US Income Data. *Journal of Applied Econometrics*, 5: 189–196**.**
- Pareto, V. (1895). La legge Della domanda, giornale degli economisti, 10: 59-68. English translation in rivista di politica economica, 87: 691–700.
- Pareto, V. (1897) Cour's d'Economie Politique, Rouge, Lausanne.
- Shao, Q. (2002). Maximum likelihood estimation for generalised logistic distributions, *Communications in Statistics: Theory and Methods*, 31:1687–1700.
- Stoppa, G. (1995) Explicit Estimators for income distributions, in c. dagum and a. lemmi (eds.) research on economic inequality, 6: Income Distribution, Social Welfare, Inequality and Poverty, Greenwich, CT: JAI Press.
- Thurow, L. C. (1970). Analyzing the American income distribution. *American Economic Review*, 48: 261-269.
- Wingo, D. R. (1984). Fitting Three-parameter log-normal models by numerical global optimization – an improved algorithm, Computation Statistical Data Analysis.

- Yuan, P. (1933). On the logarithmic frequency distribution and semi-logarithmic correlation surface, annals of mathematical statistics.
- Zelterman, D. (1987). Parameter estimation in the generalized logistic distribution, *Computational Statistics & Data Analysis*, 5: 177–184.