رگرسیون بتای آمیخته افزوده و مدل‌بندی نسبت شاغلین در خانوار

زهره فلاح محسن‌خانی و محسن محمدزاده[۱]

گروه آمار، دانشگاه تربیت مدرس

تاریخ دریافت: ۱۳۹۴/۹/۲۵ تاریخ پذیرش: ۱۳۹۶/۴/۳۰

چکیده: مدل رگرسیون بتا معمولاً برای مدل‌بندی داده‌هایی به‌صورت نرخ یا نسبت در بازه (۰,۱) بکار برده می‌شود. در بعضی مطالعات این‌گونه داده‌ها ممکن است مقادیر صفر و یک را نیز شامل شوند. در این مقاله مدل رگرسیون بتای افزوده، که از آمیختن توزیع بتا روی بازه (۰,۱) و دو توزیع تباهیده در صفر و یک ایجاد می‌شود، برای مدل‌بندی داده‌های مشاهده شده در بازه بسته [۰,۱] ارائه شده است. مدل رگرسیون بتای آمیخته افزوده با بازپارامتریدن توزیع بتا، پارامترهای میانگین و دقت را با ساختاری شامل اثرات ثابت و تصادفی مدل‌بندی می‌نماید. لحاظ کردن اثرات تصادفی موجب انعطاف‌پذیری بیشتر مدل‌ها می‌شود و می‌توان وابستگی داده‌ها را نیز در مدل منظور نمود. در اینجا مدل رگرسیون بتای آمیخته افزوده معرفی می‌شود. آنگاه کارایی مدل در یک مطالعه شبیه‌سازی موردبررسی قرار می‌گیرد. سپس نحوه کاربست این مدل برای تحلیل نسبت شاغلین در خانوار نشان داده می‌شود و در انتها بحث و نتیجه‌گیری ارائه خواهد شد.

واژه‌های کلیدی: رگرسیون بتای افزوده، تحلیل بیزی، آمارگیری نیروی کار، توزیع بتا، مدل آمیخته.

رده‌بندی ریاضی (۲۰۱۰): ۶۲J۱۲، ۶۲F۱۵.

۱- مقدمه

در مسائل کاربردی معمولاً برای تحلیل داده‌هایی که مرتبط با متغیرهای تبینی هستند، از مدل‌های رگرسیونی استفاده می‌شود. برازش این مدل با فرض نرمال بودن متغیر پاسخ یا تبدیلی

۱- آدرس الکترونیکی نویسنده مسئول مقاله: mohsen_m@modares.ac.ir

از آن، ثابت بودن واریانس و ناهمبسته بودن مؤلفه‌های خطا انجام می‌شود. در مواردی ممکن است متغیر پاسخ به بازه (۰,۱) محدود شود، برای مثال داده‌های نسبت یا نرخ و با برازش مدل رگرسیونی، پیش‌گویی‌هایی خارج از بازه تعریف شده به‌دست آید. در این حالت استفاده از مدل‌های رگرسیونی معمول مناسب نیستند. راه‌حل متداول، تبدیل متغیر وابسته به‌گونه‌ای است که مقادیر آن در مجموعه اعداد حقیقی قرار گیرند. سپس مدل میانگین پاسخ تبدیل یافته به‌عنوان پیش‌بینی خطی بر اساس مجموعه‌ای از متغیرهای تبینی انجام می‌شود. این رویکرد دارای اشکالاتی است، از جمله این‌که برآورد پارامترهای مدل به‌راحتی قابل تفسیر نیستند. ضعف دیگر این است که توزیع متغیر وابسته که به‌صورت نسبتی است، ممکن است متقارن نباشد، ازاین‌رو استنباط بر اساس فرض نرمال بودن متغیر پاسخ گمراه کننده است. پائولینو [۱] برای اولین بار به‌منظور مدل‌بندی متغیرهای پاسخ از جنس نسبت، فرض کرد متغیر پاسخ از توزیع بتا با تابع چگالی $B(a,b)$

$$\pi(y;a,b)=\frac{\Gamma(a,b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1} \qquad 0<y<1,\ \ a,b>0$$

پیروی می‌کند، که میانگین واریانس آن به ترتیب عبارت‌اند از:

$$E(Y)=\frac{a}{a+b},\quad Var(Y)=\frac{ab}{(a+b)^2(b+a+1)}$$

وی پارامترهای توزیع بتا، را نیز مدل‌بندی نموده و برآورد ماکسیمم درستنمایی آن‌ها را به‌دست آورد. در این مدل رگرسیونی، فرض بر آن است که متغیر پاسخ از توزیع بتا پیروی می‌کند که به ازای مقادیر مختلف پارامترها، توزیعی انعطاف‌پذیر است و می‌تواند شکل‌های متقارن یا نامتقارن داشته باشد. این انعطاف‌پذیری استفاده از این مدل را در کاربردهای زیادی میسر می‌سازد. این روش همچنین برای حالت‌هایی که متغیر پاسخ محدود به فاصله کراندار (a,b) باشد نیز با تبدیل متغیر پاسخ به $y^*=\frac{(y-a)}{(b-a)}$ قابل انجام است، اما از لحاظ محاسباتی با دشواری بیشتری همراه است. مثال‌هایی از داده‌های بالینی با این روش توسط کیسچینیک و همکاران [۲] تحلیل شدند. با توجه به این‌که در مدل‌های رگرسیونی، الگوی رفتار میانگین متغیر پاسخ مشروط بر متغیرهای تبیینی موردبررسی قرار می‌گیرد، فراری و کریباری [۳] توزیع بتای بازپارامتریده را برای مطالعه نسبت‌ها پیشنهاد کردند. آن‌ها پارامترهای توزیع بتا را به‌گونه‌ای بازنویسی کردند که مدل رگرسیونی براساس میانگین متغیر پاسخ بیان شود. برای این منظور با قرار دادن $\mu=\frac{a}{a+b}$ و $\phi=a+b$ توزیع بتا به‌صورت:

$$\pi(y;\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \qquad 0<y<1 \qquad (1)$$

بازپارامتریده می‌شود، که در آن μ میانگین و ϕ پارامتر دقت است ($0<\mu<1,\ \phi>0$). آن‌ها با در نظر گرفتن الگویی خطی برای متغیرهای تبیینی به پردازش مدل پرداختند. در این مدل‌ها متغیرهای تبیینی و میانگین متغیر پاسخ از طریق یک تابع پیوند[1] مناسب به‌صورت $g(\mu_i)=\sum_{j=1}^{k}x_{ij}\beta_j$ به هم مربوط می‌شوند، که در آن β بردار پارامترهای رگرسیونی است. چپیدا و گامرمن [۴] و اسمیتسون ورکولین[۵] نیز مدل رگرسیون بتا را با متغیر در نظر گرفتن پارامتر دقت موردبررسی قرار دادند. آن‌ها به‌طور هم‌زمان لوجیت میانگین و لگاریتم پارامتر دقت را به‌صورت الگوی خطی از متغیرهای تبیینی در مدل لحاظ کردند. برانسکام و همکاران [۶] نیز برازش مدل رگرسیون بتا به داده‌های ژنتیک را با رهیافت بیزی مطالعه کردند.

در برخی از مسائل کاربردی ممکن است با مواردی مواجه شویم که نرخ‌ها هر یک از مقادیر مجموعه {۰و۱} را نیز اختیار کنند، در این شرایط تکیه‌گاه توزیع بتا، یعنی بازه (۰,۱)، نمی‌تواند داده‌ها را به‌طور کامل پوشش دهد، برای حل این مسئله اوسپینا و فراری [۷] آمیختن یک توزیع گسسته به توزیع پیوسته بتا را پیشنهاد کردند که بسته به شرایط می‌تواند جرم احتمال‌های صفر، یک یا هر دو را نیز شامل شود. اگر متغیر پاسخ در بازه [۰,۱] باشد، آمیختن یک توزیع پیوسته بتا در بازه (۰,۱) با توزیع گسسته دوجمله‌ای، که احتمال‌های غیر صفر به نقاط صفر و یک اختصاص می‌دهد، توسط اوسپینا و فراری [۸] تحت عنوان مدل‌های رگرسیون بتای آماسیده[2] ارائه شد.

الویس و همکاران [۹] در مواجهه با قرارگرفتن نسبت‌ها در بازه [۰,۱]، مدل بتای افزوده[3] صفر و یک را پیشنهاد کردند که شامل یک توزیع سه قسمتی است، به‌گونه‌ای که در نقاط صفر و یک تباهیده و در بازه (۰,۱) دارای چگالی بتا است. علت تفاوت دیدگاه آن‌ها نسبت به پیشنهاد اوسپینا و فراری [۷] در مفهوم آماسیده (تورم) است. توزیع احتمال وقتی در یک نقطه آماسیده می‌شود که جرم احتمال آن، بیش از مقدار معمول توزیع در آن نقطه شود. درصورتی‌که اگر متغیر پاسخ در بازه [۰,۱] باشد، لزومی ندارد که همیشه مشاهدات به‌گونه‌ای باشند که در نقاط صفر و یک، جرم‌های احتمال بزرگ‌تر از توزیع بتا را به خود اختصاص دهند (برای مثال ممکن است در

1-Link Function
2-Inflated Beta Regression
3-Zero-and-one Augmented Beta

عمل جرم احتمال در یکی از نقاط صفر یا یک، کمتر از جرم احتمال توزیع بتا در آن نقطه باشد). ازاین‌رو الویس و همکاران مدل افزوده را کلی‌تر از مدل آماسیده دانستند.

در مسائل کاربردی ممکن است علاوه بر محدودیت دامنه پاسخ، بنا به دلایل متعدد، وابستگی نیز بخشی از ماهیت داده‌ها باشد که می‌بایست در مدل‌بندی داده‌ها لحاظ شود. این وابستگی ممکن است به دلیل موقعیت‌های فضایی، زمانی یا عوامل دیگری در داده‌ها ملاحظه شود. در حالتی که متغیر پاسخ به بازه (۰,۱) محدود شده است، مدل‌های آمیخته خطی با لحاظ کردن اثرهای تصادفی توسط فیوگیورا و همکاران [۱۰] پیشنهاد شدند، که در آن با استفاده از یک تابع پیوند مناسب، میانگین شرطی داده‌ها برابر با ترکیبی خطی از اثرات تصادفی و ثابت در نظر گرفته می‌شود. الویس و همکاران [۹] مدل رگرسیون بتای آمیخته افزوده[1] را معرفی و اثر تصادفی را در این مدل در نظر گرفتند، ولی صرفاً برای حالت خاصی در نظر گرفته شد که متغیر پاسخ شرط تابعیت از تکرار قبل را برخوردار باشد، یعنی اگر متغیر پاسخ در تکرار iام مقداری در بازه (۱و۰) اختیار کند در تکرار $(i+1)$ نیز مقدار آن حتماً در بازه (۰,۱) قرار می‌گیرید و اگر متغیر پاسخ در تکرار iام مقدار ۰ یا ۱ را اختیار کند در تکرار $(i+1)$ام نیز حتماً همان مقدار قبلی خود را اختیار می‌کند. ازآنجاکه شرط تابعیت از تکرار قبل در عمل، محدودیت زیادی ایجاد می‌کند، در این مقاله این شرط کنار گذاشته شده، به‌گونه‌ای که متغیر پاسخ در هر تکرار مقادیر خود را در بازه بسته [۰,۱] به‌طور تصادفی اختیار می‌نماید.

در این مقاله ابتدا مدل رگرسیون بتای آمیخته افزوده با لحاظ کردن اثر تصادفی معرفی می‌شود. سپس کاربرد آن برای مدل‌بندی سهم شاغلین در خانوار بر اساس نتایج آمارگیری نیروی کار مرکز آمار بکار می‌رود که در هر تکرار ممکن است مقداری بین صفر و یک، صفر یا یک را به خود اختصاص دهند. به عبارتی دیگر داده‌ها لزوماً تابع تکرار قبلی نیستند. آنگاه کارایی مدل معرفی شده برای این نوع داده‌ها در مطالعات شبیه‌سازی موردبررسی قرار می‌گیرد. نهایتاً به بحث و نتیجه‌گیری پرداخته خواهد شد.

۲-مدل رگرسیون بتای آمیخته افزوده

تابع چگالی احتمال بتای بازپارامتریده شده برای متغیر تصادفی Y برحسب پارامترهای میانگین و دقت به‌صورت (۱) است، که در آن $\Gamma(.)$ تابع گاما، $E(Y)=\mu$ و $Var(Y)=\frac{\mu(1-\mu)}{1+\phi}$

1- Augmented Mixed Beta Regression Model

است. الویس و همکاران [۹] برای داده‌هایی که بازهٔ بسته صفر و یک را اختیار می‌کنند، یک توزیع آمیخته سه قسمتی شامل یک توزیع بتا در بازهٔ (۰,۱) و دو توزیع تباهیده در صفر و یک به‌صورت:

$$f(y \mid p_0, p_1, \mu, \phi) = \begin{cases} p_0 & y = 0 \\ p_1 & y = 1 \\ (1-p_0-p_1)h(y \mid \mu, \phi) & 0 < y < 1 \end{cases}$$

پیشنهاد کردند، که در آن $h(y \mid \mu, \phi)$ تابع چگالی بتا و $p_0, p_1 \geq 0$ و $0 \leq p_0 + p_1 \leq 1$. میانگین واریانس این توزیع به ترتیب عبارت‌اند از:

$$E(Y) = (1-p_0-p_1)\mu + p_1,$$

$$Var(Y) = p_1(1-p_1) + (1-p_0-p_1)\left[\frac{\mu(1-\mu)}{(1+\phi)} + (p_0+p_1)\mu^2 - 2\mu p_1\right]$$

۲-۱- مدل رگرسیون آمیخته بتای افزوده

فرض کنید $\boldsymbol{Y}_i = (y_{i1}, \ldots, y_{in_i})$ بردار به طول n_i برای واحد نمونه iام است. متغیرهای تبیینی و اثرات تصادفی[1] را می‌توان بر اساس یک تبدیل مناسب از μ_i به‌صورت:

$$g(E(\boldsymbol{Y}_i \mid \boldsymbol{b}_i)) = g(\boldsymbol{\mu}_i) = \boldsymbol{X}_i^T \beta + \boldsymbol{Z}_i^T \boldsymbol{b}_i \quad i = 1, \ldots, n$$

مدل‌بندی کرد، که در آن $\boldsymbol{X}_i$ ماتریس طرح $p \times n_i$ بعدی، $\beta = (\beta_1, \ldots, \beta_p)^T$ بردار اثرات ثابت، $\boldsymbol{Z}_i$ ماتریس طرح $q \times n_i$ بعدی و $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iq})^T$ بردار اثرات تصادفی است. برای تابع پیوند $g(.)$ انتخاب‌های متفاوت از جمله لوجیت را می‌توان اختیار نمود. برای پارامترهای p_0 و p_1 نیز می‌توان توابع پیوند لوجیت[2] و برای پارامتر ϕ تابع پیوند لگاریتم انتخاب کرد. ولی عموماً برای سادگی، این پارامترها ثابت در نظر گرفته می‌شوند.

فرض کنید $\theta = (p_0, p_1, \phi, \beta)$ بردار پارامترهای مدل رگرسیون بتای آمیخته افزوده باشد که لازم است برآورد شوند. با فرض استقلال پارامترها، بر اساس نمونه مشاهده شده برای n فرد که با $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ و $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ نشان داده می‌شوند، لگاریتم تابع درستنمایی عبارت است از:

$$\ell(\theta; \boldsymbol{b}, \boldsymbol{X}, \boldsymbol{y}) = \sum_{i,j:y_{ij}=0} \log(p_0) + \sum_{i,j:y_{ij}=1} \log(p_1) + (1-p_0-p_1) \sum_{i,j:y_{ij} \in (0,1)} \ell_1(\theta; \boldsymbol{b}_i, \boldsymbol{X}_i, \boldsymbol{y}_i)$$

1- Random Effects
2- Logit Function

که در آن

$$\ell_{۱}(\theta;\boldsymbol{b}_i,\boldsymbol{X}_i,\boldsymbol{y}_i)=\log\left[\Gamma(\phi)\right]-\log\left(\Gamma(\frac{\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)}{۱+\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)}\phi)\right)$$
$$-\log\left(\Gamma\left((۱-\frac{\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)}{۱+\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)})\phi\right)\right)+\left(\frac{\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)}{۱+\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)}\phi-۱\right)\log(\boldsymbol{y}_i)$$
$$+\left(\left(۱-\frac{\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)}{۱+\exp(\boldsymbol{X}_i\beta+\boldsymbol{b}_i)}\right)\phi-۱\right)\log(۱-\boldsymbol{y}_i)$$

توزیع پسین توأم به‌صورت

$$p(\theta,\boldsymbol{b},\sigma_b^{۲};\boldsymbol{X},\boldsymbol{y})\propto L(\theta;\boldsymbol{b},\boldsymbol{X},\boldsymbol{y})\,\pi_{۰}(\boldsymbol{b};\sigma_b^{۲})\pi_{۰}(\sigma_b^{۲})\pi_{۰}(\theta)$$

است، که در آن $\pi_{۰}(\theta)$ برابر حاصل‌ضرب توزیع‌های پیشین پارامترهای $p_{۰}$، $p_{۱}$، ϕ و β است که عموماً برای این پارامترها، توزیع‌های پیشین ناآگاهی بخش یا کم‌آگاهی بخش در نظر گرفته می‌شوند. ازآنجاکه به‌دست آوردن توزیع‌های پسینی حاشیه‌ای به‌صورت تحلیلی بسیار پیچیده است از الگوریتم مونت‌کارلوی زنجیر مارکوفی[۱] (MCMC) و نمونه‌گیری گیبز[۲] استفاده می‌شود. بدین منظور به توزیع‌های شرطی کامل نیاز است. فرض کنید $\pi_{۰}(.)$ نشان دهنده‌ی توزیع‌های پیشین پارامترهای مدل باشند. در این‌صورت توزیع‌های شرطی کامل عبارت‌اند از:

$$p(\beta\mid p_{۰},p_{۱},\phi,\boldsymbol{b},\sigma_b^{۲},\boldsymbol{y})\propto\prod_{i,j;y_{ij}\in(۰,۱)}f(y_{ij}\mid\beta,b_i)\pi_{۰}(\beta)$$
$$p(\boldsymbol{b}\mid p_{۰},p_{۱},\phi,\beta,\sigma_b^{۲},\boldsymbol{y})\propto\prod_{i,j;y_{ij}\in(۰,۱)}f(y_{ij}\mid\beta,b_i)\pi_{۰}(b_i;\sigma_b^{۲})$$
$$p(\phi\mid p_{۰},p_{۱},\beta,\boldsymbol{b},\sigma_b^{۲},\boldsymbol{y})\propto\prod_{i,j;y_{ij}\in(۰,۱)}f(y_{ij}\mid\beta,b_i,\phi)\pi_{۰}(\phi)$$
$$p(p_{۰}\mid p_{۱},\beta,\phi,\boldsymbol{b},\sigma_b^{۲},\boldsymbol{y})\propto\prod_{i,j;y_{ij}=۰}f(y_{ij}\mid p\;)\pi_{۰}(p_{۰})$$
$$p(p_{۱}\mid p_{۰},\beta,\phi,\boldsymbol{b},\sigma_b^{۲},\boldsymbol{y})\propto\prod_{i,j;y_{ij}=۱}f(y_{ij}\mid p_{۱})\pi_{۰}(p_{۱})$$
$$p(\sigma_b^{۲}\mid p_{۰},p_{۱},\beta,\phi,\boldsymbol{b},\boldsymbol{y})\propto\prod_{i,j;y_{ij}\in(۰,۱)}f(y_{ij}\mid\sigma_b^{۲})\pi_{۰}(\sigma_b^{۲})$$

1- Markov Chain Monte Carlo
2- Gibbs Sampling

با الگوریتم MCMC (ترکیبی از نمونه‌گیری گیبز و متروپولیس در گیبز) و با استفاده از بسته R2WinBUGS که دو نرم‌افزار R و WinBUGS را به هم متصل می‌کند، می‌توان از توزیع‌های شرطی کامل برای پارامترها مقادیر تصادفی تولید نمود [۱۱].

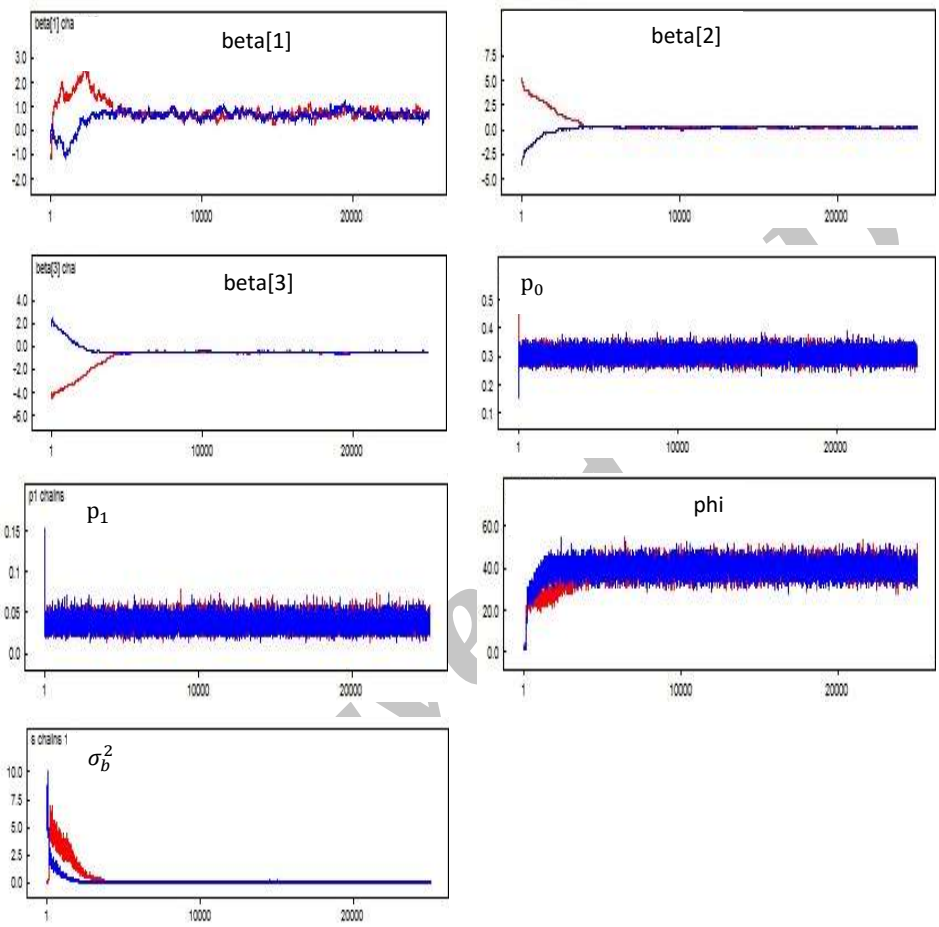
۳- تحلیل داده‌های نیروی کار

آمارگیری نیروی کار یکی از مهم‌ترین نمونه‌گیری‌های خانواری ایران است که جامعه آن همه اعضای خانوارهای معمولی ساکن هستند. در این نمونه‌گیری هر فصل از حدود ۶۰۰۰۰ خانوار آمارگیری می‌شود. برای هر خانوار منتخب نمونه، یک فرم خانواری حاوی مشخصات تمام اعضای خانوار و فرمی فردی حاوی وضعیت فعالیت افراد ۱۰ ساله و بیشتر تکمیل می‌شود. نمونه‌ها بر اساس الگوی چرخش ۲-۲-۲ انتخاب می‌شوند. یعنی هر خانوار ۲ فصل پی‌درپی در آمارگیری حضور دارد، در فصل‌های سوم و چهارم در نمونه‌ها قرار نمی‌گیرد و در دو فصل بعدی پنجم و ششم دوباره به‌عنوان عضو نمونه محسوب می‌شوند. بنابراین برای هر خانوار نمونه ۴ تکرار وجود دارد. برای برآورد پارامترهای جامعه، داده‌هایی که از این آمارگیری برای هر عضو خانوارهای نمونه جمع‌آوری شده است با سه مرحله وزن‌دهی بر اساس وزن پایه، وزن بی‌پاسخی وزن پیش‌بینی‌های جمعیتی تعدیل می‌شوند. در این آمارگیری بر اساس وضع فعالیت تمام افراد ۱۰ ساله و بیشتر خانوارهای نمونه، شاخص‌های مربوط به اشتغال و بیکاری و نسبت شاغلین در هر خانوار تعیین می‌شوند. اگر تعداد اعضای شاغل خانوار کمتر از تعداد اعضای خانوار باشد، این نسبت، عددی بین صفر و یک، اگر تمام اعضای خانوار شاغل باشند، عدد یک و درصورتی‌که هیچ‌یک از اعضای خانوار شاغل نباشند، عدد صفر را اختیار می‌کند. بنابراین متغیر پاسخ یعنی نسبت شاغلین در خانوار، متغیری پیوسته با تحقق‌هایی در بازهٔ بسته $[۰,۱]$ است. در این مطالعه داده‌های مربوط به خانوارهای مشترک در فصل‌های اول و دوم سال‌های ۱۳۹۲ و ۱۳۹۳ آمارگیری نیروی کار شهر تهران در نظر گرفته شده است که شامل اطلاعات ۱۵۹خانوار است [۱۲]. با فرض ثابت بودن پارامترهای $p_۰$، $p_۱$، ϕ و انتخاب تابع پیونگد لوجیت برای میانگین نسبت شاغلین در خانوار از مدل رگرسیون بتای آمیخته افزوده برای مدل‌بندی داده‌ها استفاده می‌شود. اگر $x^{(۱)}$، تعداد افراد ۱۰ ساله و بیشتر و $x^{(۲)}$ جمعیت خانوار به‌عنوان متغیرهای تبینی در این مدل قرار گیرند، لوجیت میانگین نسبت شاغلین به‌صورت

$$g(\mu_{ij}) = \beta_۰ + \beta_۱ x_{ij}^{(۱)} + \beta_۲ x_{ij}^{(۲)} + b_i \qquad i = ۱,\ldots,۱۵۹; j = ۱,\ldots,۴$$

در نظر گرفته می‌شود، که در آن $\beta_۰$ عرض از مبدأ، $\beta_۱$ و $\beta_۲$ پارامترهای رگرسیونی مدل و b_i اثر تصادفی خانوار iام است. ازآنجاکه توزیع t-استیودنت دم کلفت‌تر از توزیع نرمال است، بنا به توصیه فیگورا [۱۰]، برای β توزیع پیشین t-استیودنت چند متغیره $t_۲(\nu_\beta, \mu_\beta, \Sigma_\beta)$ در

نظر گرفته می‌شود، که در آن $\nu_\beta = ۱۰$، μ_β بردار صفر و Σ_β ماتریس قطری ۱۰ منظور شده است. همچنین بنا به پیشنهاد گالویز [۹] برای ϕ توزیع پیشین گاما $G(۰/۱,۰/۱)$، برای $p_۰$ توزیع پیشین یکنواخت $U(۰,۱)$، برای $p_۱$ توزیع پیشین یکنواخت $U(۰,۱-p_۰)$، برای b_i توزیع پیشین نرمال $N(۰,\sigma_b^۲)$ و برای $\sigma_b^۲$ توزیع گامای وارون $IG(۰/۱,۰/۱)$ در نظر گرفته شد.



شکل (۱): نمودارهای ترتیبی نمونه‌های تولید شده از شرطی‌های کامل پارامترها

برآورد بیزی پارامترها بر اساس نمونه‌ای تصادفی به حجم $n=$ ۱۵۹ از توزیع‌های شرطی کامل حاصل شده است، که از اجرای گام‌های الگوریتم مونت‌کارلوی زنجیر مارکوفی برای ۲۵۰۰۰۰

تکرار و مرحله داغیدن ۲۰۰۰۰۰ و از تأخیر دوم به‌دست آمده است. شکل ۱ نمودارهای ترتیبی میانگین نمونه‌های تولید شده از توزیع‌های شرطی کامل پارامترهای مدل هستند که بیان‌گر همگرایی این نمونه‌ها است. میانگین نمونه‌های تولید شده از توزیع‌های شرطی کامل پارامترها، یعنی برآورد بیزی پارامترها به همراه انحراف استاندارد و بازه اطمینان ۹۵٪ در جدول ۱ ارائه شده‌اند. با توجه به بازه‌های اطمینان ۹۵٪ ملاحظه می‌شود که همه پارامترها در مدل معنی‌دار هستند. عرض از مبدأ و متغیر تبیینی تعداد افراد ۱۰ ساله و بیشتر خانوار با اثر مثبت و متغیر تبیینی جمعیت خانوار با اثر منفی وارد مدل شدند. بنابراین انتظار می‌رود با افزایش تعداد افراد ۱۰ ساله و بیشتر خانوار، سهم شاغلین در خانوار افزایش یابد ولی با افزایش بعد خانوار این سهم کاهش یابد. پارامترهای p_0 و p_1 به ترتیب تقریباً ۰/۳۱ و ۰/۰۴ برآورد شدند. این ارقام گویای این است که احتمال غیر شاغل بودن تمام اعضای یک خانوار تقریباً ده برابر احتمال شاغل بودن تمام اعضای یک خانوار است. همچنین پارامتر دقت توزیع بتا، ۳۹/۷۶ واریانس اثر تصادفی نیز ۰/۱۶ برآورد شده است.

جدول (۱): برآورد پارامترها و انحراف استاندارد آن‌ها به همراه میانه چندک‌های ۲/۵، ۵۰ و ۹۷/۵ درصد حاصل از برازش مدل به داده‌های خانوارهای مشترک فصل‌های اول و دوم سال‌های ۱۳۹۲ و ۱۳۹۳

پارامتر	برآورد	انحراف استاندارد	٪۲/۵	٪۵۰	٪۹۷/۵
β_1	۰/۷۲۱۸	۰/۳۱۹۲	۰/۳۶۱۲	۰/۶۷۴۸	۰/۸۵۳۲
β_2	۰/۲۱۲۷	۰/۲۳۷۵	۰/۰۴۱۱	۰/۲۰۲۵	۰/۲۸۰۹
β_3	-۰/۵۶۲۱	۰/۲۹۲۹	-۱/۵۰۵۱	-۰/۵۱۴۱	-۰/۳۶۰۳
p_0	۰/۳۰۸۸	۰/۰۱۸۴	۰/۲۷۳۵	۰/۳۰۸۵	۰/۳۴۵۲
p_1	۰/۰۳۷۵	۰/۰۰۷۵	۰/۰۲۴۳	۰/۰۳۷۱	۰/۰۵۳۵
ϕ	۳۹/۷۶	۳/۷۰۴۱	۳۳/۰۱۲۱	۳۹/۷۶۱۲	۴۶/۳۹۱۱
σ_b^2	۰/۱۵۵۳	۰/۱۴۵۱	۰/۰۹۵۳	۰/۱۳۲۱	۰/۳۹۶۹

یکی از موارد کاربرد این مدل، ارائه آمار کمکی مناسب برای برآورد شاغلین در نواحی کوچک است. ازآنجایی‌که متغیرهای وارد شده در مدل (بعد خانوار و تعداد افراد ۱۰ ساله و بیشتر) بر اساس سرشماری‌ها و فهرست‌برداری‌هایی که در دستور کار مرکز آمار ایران قرار می‌گیرد قابل حصول است، استفاده از این مدل می‌تواند برآوردهایی از سهم شاغلین در خانوار در تمام زیر ناحیه‌های کوچک جغرافیایی ارائه کند. با ضرب سهم شاغلین در خانوار در تعداد خانوار نواحی موردنظر، تعداد شاغلین این نواحی قابل برآورد هستند. با نگاهی محافظه‌کارانه می‌توان از این برآورد به‌طور غیرمستقیم، در برآورد شاغلین نواحی کوچک، استفاده کرد. بدین‌صورت که برآورد

مذکور، خود می‌تواند یک متغیر کمکی مناسب برای برآورد شاغلین در مدل‌های ناحیه‌کوچک[1] برای برآورد تعداد شاغلین ناحیه مذکور باشد. زیرا در برآورد شاغلین نواحی کوچک، بزرگ‌ترین مشکل، عدم وجود آمار کمکی مناسب، برای آن ناحیه، در مدل‌های ناحیه کوچک است. همچنین از این مدل در ارزیابی نتایج آمارگیری نیروی کار استان‌ها نیز می‌توان کمک گرفت. بدین‌صورت که در هر دوره آمارگیری، برآورد شاغلین استان‌ها که مستقیماً از آمارگیری نیروی کار به دست می‌آید با برآورد شاغلین استان‌ها که از مدل حاصل می‌شوند، مقایسه شوند. عموماً انتظار می‌رود، اختلاف قابل‌توجهی در برآوردها مشاهده نشود.

۴-مطالعه شبیه‌سازی

در این بخش برای ارزیابی مدل رگرسیون بتای آمیخته افزوده برای حالتی که متغیر پاسخ در هر تکرار می‌تواند صفر، یک یا عددی بین صفر و یک را اختیار کند، مطالعه شبیه‌سازی انجام شده است. هدف این مطالعه برآورد میانگین، اریبی نسبی، انحراف استاندارد و میانگین توان دوم خطای فراورده‌ای پارامتر رگرسیونی به دست آمده از مدل رگرسیون بتای آمیخته افزوده برای اندازه نمونه‌های مختلف است.

در این شبیه‌سازی برای پارامتر μ_{ij} پیوند لوجیت

$$\log it(\mu_{ij}) = \beta_0 + \beta_1 x_{ij}^{(1)} + \beta_2 x_{ij}^{(2)} + b_i \qquad i = 1,\ldots,N \; ; j = 1,\ldots,5$$

منظور شده است، که در آن b_i دارای توزیع‌های نرمال با میانگین صفر واریانس‌های ۱ و ۵، در نظر گرفته شده است. اندازه‌های نمونه ۵۰، ۱۰۰ و ۲۰۰ در نظر گرفته شده است. پارامترهای p_0 ، p_1 و ϕ ثابت فرض شده و مقادیر آن‌ها به ترتیب برابر ۰/۱، ۰/۱ و ۴۰ در نظر گرفته شده است. متغیرهای $x_{ij}^{(1)}$ و $x_{ij}^{(2)}$ مستقل و دارای توزیع $N(0,1)$ فرض شده‌اند. برای پارامترهای رگرسیونی β_0، β_1 و β_2 به ترتیب مقادیر ۰/۵، ۰/۴ و ۰/۶ قرار داده شده است. برای تولید y_{ij}ها ابتدا یک دنباله مستقل از توزیع برنولی با احتمال ۰/۸ $(1-p_0-p_1)$ تولید شده، اگر مقادیر این دنباله، مقدار صفر را اختیار کرده باشند، y_{ij} برابر صفر و در غیر این‌صورت y_{ij} از توزیع بتای $B(\mu_{ij}\phi,(1-\mu_{ij})\phi)$ به ازای ϕ برابر ۴۰ تولید می‌شوند. سپس دنباله‌ای برنولی با احتمال ۰/۵ از صفر و یک‌ها جایگزین y_{ij}هایی که صفر هستند، در نظر گرفته می‌شود با این الگو، دنباله‌ای از y_{ij}هایی به دست می‌آید که با احتمال ۰/۸ دارای توزیع بتا، با احتمال ۰/۱، مقدار صفر و با احتمال ۰/۱، مقدار یک را اختیار می‌کند.

1- Small Area Models

جدول (۲): فراورده‌ای بیزی پارامترها پس از برازش مدل در اندازه نمونه ۵۰

پارامتر	مقدار واقعی	برآورد	انحراف استاندارد	میانگین توان دوم خطا	اریبی نسبی	٪۲/۵	٪۹۷/۵
β_1	۰/۵	۰/۴۷۶۸	۰/۱۴۶۴	۰/۰۲۱۷	-۰/۰۴۶۴	۰/۴۴۸۱	۰/۵۰۵۵
β_2	۰/۴	۰/۴۰۷۴	۰/۰۳۳۶	۰/۰۰۱۲	۰/۰۱۸۶	۰/۴۰۰۸	۰/۴۱۴۰
β_3	۰/۶	۰/۵۹۹۸	۰/۰۳۱۶	۰/۰۰۰۹	-۰/۰۰۰۲	۰/۵۹۳۶	۰/۶۰۶۰
p_0	۰/۱	۰/۱۰۰۴	۰/۰۲۰۲	۰/۰۰۰۴	۰/۰۰۴۴	۰/۰۹۶۵	۰/۱۰۴۴
p_1	۰/۱	۰/۱۰۱۹	۰/۰۱۸۳۸	۰/۰۰۰۳	۰/۰۱۹۴۱	۰/۰۹۸۳	۰/۱۰۵۵
ϕ	۴۰	۳۹/۱۳۵۸	۳/۹۹۲۴	۱۶/۵۲۶۹	-۰/۰۲۱۶	۳۸/۳۵۳۲	۳۹/۹۱۸۳
σ_b^2	۱	۱/۰۲۹۹	۰/۲۱۸۲	۰/۰۴۸۵	۰/۰۲۹۹	۰/۹۸۷۱	۱/۰۷۲۷
β_1	۰/۵	۰/۴۷۱۶	۰/۳۰۸۳	۰/۰۹۴۹	-۰/۰۵۶۶	۰/۴۱۱۲	۰/۵۳۲۱
β_2	۰/۴	۰/۳۹۲۵	۰/۰۳۹۷	۰/۰۰۱۴	۰/۰۱۸۶	۰/۳۸۴۷	۰/۴۰۰۳
β_3	۰/۶	۰/۶۰۰۹	۰/۰۳۱۴	۰/۰۰۰۹	-۰/۰۰۱۵	۰/۵۹۴۷	۰/۶۰۷۰
p_0	۰/۱	۰/۱۰۲۵	۰/۰۱۸۶	۰/۰۰۰۳	۰/۰۲۵۸	۰/۰۹۸۹	۰/۱۰۶۲
p_1	۰/۱	۰/۱۰۳۱	۰/۰۱۸۸	۰/۰۰۰۴	۰/۰۳۱۴	۰/۰۹۹۴	۰/۱۰۶۸
ϕ	۴۰	۳۹/۱۹۵۰	۴/۲۳۳۵	۱۸/۸۳۵۴	-۰/۰۲۰۱	۳۸/۳۶۵۳	۴۰/۰۲۴۸
σ_b^2	۵	۵/۰۷۹۷	۰/۸۹۷۲	۰/۸۰۳۴	۰/۰۱۵۹	۴/۹۰۳۹	۵/۲۵۵۶

در این شبیه‌سازی برای برآورد بردار پارامترهای $\theta=(p_0,p_1,\phi,\beta_0,\beta_1,\beta_2,\sigma_b^2)\mathfrak{M}$ تعداد ۱۰۰۰۰۰ تکرار مونت‌کارلو در نظر گرفته شده و نتایج با در نظر گرفتن ۵۰۰۰۰ تکرار آخر در جداول ۲، ۳ و ۴ ارائه شده‌اند. این جداول شامل مقادیر برآورد پارامترها، انحراف استاندارد، بازه اطمینان ۹۵ درصدی، اریبی نسبی و میانگین توان دوم خطا است، که به‌صورت

$$MSE(\theta)=\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i-\theta)^2 \quad , \quad Rel.B=\frac{1}{N}\sum_{i=1}^{N}(\frac{\hat{\theta}_i}{\theta}-1)$$

برای ۱۰۰ مجموعه داده (N) محاسبه شده است. فراورده‌ای پارامترها بعد از برازش مدل رگرسیون بتای آمیخته افزوده برای اندازه‌های نمونه ۵۰، ۱۰۰ و ۲۰۰ ارائه شده‌اند.

همان‌طور که در جداول ۲، ۳ و ۴ ملاحظه می‌شود فراورده‌ای بیزی به دست آمده نزدیک مقادیر واقعی پارامترهای مدل هستند و با افزایش اندازه نمونه، انحراف استاندارد و میانگین توان دوم خطاها کاهش یافته است. لازم به ذکر است که آزمون‌های تشخیصی مانند بررسی نمودارهای

ترتیبی، بیانگر مانایی زنجیر هر پارامتر هستند. نظر به این‌که معمولاً چگالی‌های پسین چند مدی، همگرایی زنجیرهای تولیدی را دچار مشکل می‌سازد بررسی چند مدی بودن نمودار این چگالی‌ها از اهمیت زیادی برخوردار است [۱۰]. به همین منظور چگالی‌های پسینی تمام پارامترها در شکل ۳ ارائه شده‌اند که بیانگر تک مدی بودن توزیع آن‌ها است. آزمون‌های همگرایی تشخیصی گلمن روبین [۱۳] و همگرایی تشخیصی هیدل برگ ولچ [۱۴] دلالت بر همگرایی زنجیرهای تولید شده دارند. همچنین معیار همگرایی گلمن و روبین توأم که توسط بروک و گلمن [۱۵]، پیشنهاد شده است، دلالت بر همگرایی زنجیر دارد درصورتی‌که عامل کاهش مقیاس متناسب چند متغیره (mprf) کمتر از ۱/۲ باشد. این معیار همگرایی با استفاده از بسته coda، [۱۶]، در نرم‌افزار R برای داده‌های شبیه‌سازی شده مدل آمیخته رگرسیون بتای افزوده برابر ۱ به دست آمد که مؤید دیگری بر همگرایی زنجیرها است.

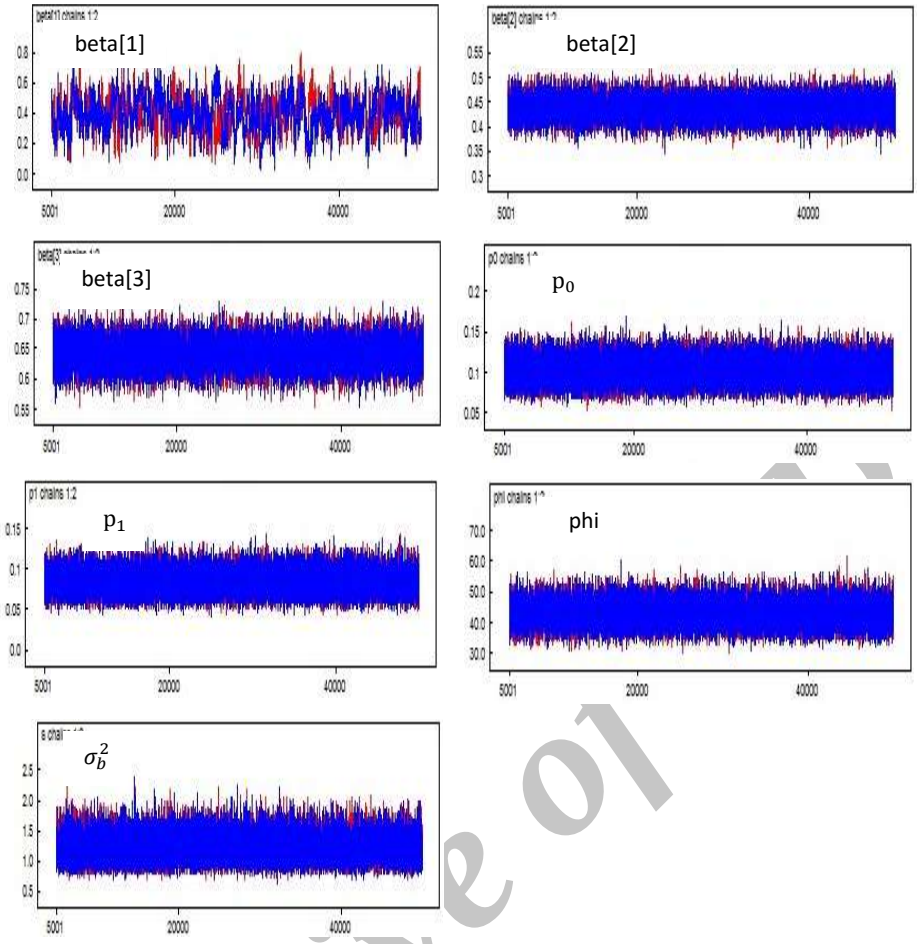
جدول (۳): فراورده‌ای بیزی پارامترها پس از برازش مدل در اندازه نمونه ۱۰۰

پارامتر	مقدار واقعی	برآورد	انحراف استاندارد	میانگین توان دوم خطا	اریبی نسبی	٪۲/۵	٪۹۷/۵
β_1	۰/۵	۰/۵۰۷۵	۰/۱۰۵۹	۰/۰۱۱۲	۰/۰۱۴۸	۰/۴۸۶۷	۰/۵۲۸۳
β_2	۰/۴	۰/۳۹۹۸	۰/۰۲۳۴	۰/۰۰۰۵	-۰/۰۰۰۳	۰/۳۹۵۳	۰/۴۰۴۵
β_3	۰/۶	۰/۶۰۴۷	۰/۰۲۴۸	۰/۰۰۰۶	۰/۰۰۷۹	۰/۵۹۹۹	۰/۶۰۹۶
p_0	۰/۱	۰/۱۰۰۴	۰/۰۱۳۶	۰/۰۰۰۲	۰/۰۰۳۹	۰/۰۹۷۷	۰/۱۰۳۱
p_1	۰/۱	۰/۱۰۱۴	۰/۰۱۳۱	۰/۰۰۰۲	۰/۰۱۴۳	۰/۰۹۸۸	۰/۱۰۴۰
ϕ	۴۰	۳۹/۸۵۰۲	۳/۵۳۲۵	۱۲/۳۷۵۸	-۰/۰۰۳۷	۳۹/۱۵۷۸	۴۰/۵۴۲۵
σ_b^2	۱	۱/۰۱۹۱	۰/۱۴۱۳	۰/۰۲۰۱	۰/۰۱۹۴	۰/۹۹۱۱	۱/۰۴۷۲
β_1	۰/۵	۰/۵۰۲۵	۰/۲۴۱۲	۰/۰۵۷۵	۰/۰۰۵۰	۰/۴۵۵۲	۰/۵۴۹۷
β_2	۰/۴	۰/۴۰۰۷	۰/۰۲۵۱	۰/۰۰۰۶	۰/۰۰۱۸	۰/۳۹۵۸	۰/۴۰۵۶
β_3	۰/۶	۰/۵۹۵۷	۰/۰۲۶۷	۰/۰۰۰۷	-۰/۰۰۷۱	۰/۵۹۰۴	۰/۶۰۰۹
p_0	۰/۱	۰/۱۰۲۴	۰/۰۱۳۰	۰/۰۰۰۳	۰/۰۲۴۱	۰/۰۹۹۸	۰/۱۰۴۹
p_1	۰/۱	۰/۱۰۲۰	۰/۰۱۳۱	۰/۰۰۰۲	۰/۰۲۰۳	۰/۰۹۹۴	۰/۱۰۴۶
ϕ	۴۰	۴۰/۲۱۵۸	۲/۷۵۷۴	۷/۵۷۳۸	-۰/۰۰۵۳	۳۹/۶۷۵۴	۴۰/۷۵۶۳
σ_b^2	۵	۵/۰۵۵۲	۰/۶۷۳۷	۰/۴۵۲۱	۰/۰۱۱۰	۴/۹۲۳۲	۵/۱۸۷۲

جدول (۴): فراورده‌ای بیزی پارامترها پس از برازش مدل در اندازه نمونه ۲۰۰

پارامتر	مقدار واقعی	میانگین	انحراف استاندارد	میانگین توان دوم خطا	اریبی نسبی	٪۲/۵	٪۹۷/۵
β_1	۰/۵	۰/۴۷۴۹	۰/۰۸۵۹	۰/۰۰۷۹	-۰/۰۵۰۱	۰/۴۵۸۱	۰/۴۹۱۷
β_2	۰/۴	۰/۳۹۹۶	۰/۰۱۱۴	۰/۰۰۱۳	-۰/۰۰۰۹	۰/۳۹۷۴	۰/۴۰۱۸
β_3	۰/۶	۰/۵۹۵۴	۰/۰۱۵۷	۰/۰۰۰۲	-۰/۰۰۷۶	۰/۵۹۲۳	۰/۵۹۸۴
p_0	۰/۱	۰/۰۹۹۱	۰/۰۱۰۶	۰/۰۰۰۱	-۰/۰۰۸۷	۰/۰۹۷۰	۰/۱۰۱۲
p_1	۰/۱	۰/۱۰۲۰	۰/۰۰۹۲	۰/۰۰۰۱	۰/۰۲۴۱	۰/۱۰۰۶	۰/۱۰۴۲
ϕ	۴۰	۴۰/۴۳۲۰	۲/۷۳۳۲	۷/۴۶۲۱	۰/۰۱۰۸	۳۹/۸۹۶۳	۴۰/۹۶۷۷
σ_b^2	۱	۱/۰۲۲۹	۰/۰۸۱۲	۰/۰۱۵۶	۰/۰۲۲۹	۱/۰۰۷۰	۱/۰۳۸۸
β_1	۰/۵	۰/۵۰۲۱	۰/۱۰۴۶	۰/۰۱۰۸	۰/۰۰۴۳	۰/۴۸۱۶	۰/۵۲۲۶
β_2	۰/۴	۰/۳۹۹۲	۰/۰۱۵۷	۰/۰۰۰۲	-۰/۰۰۱۸	۰/۳۹۶۱	۰/۴۰۲۳
β_3	۰/۶	۰/۶۰۲۱	۰/۰۳۸۰	۰/۰۰۱۴	۰/۰۰۳۶	۰/۵۹۴۷	۰/۶۰۹۶
p_0	۰/۱	۰/۱۰۲۱	۰/۰۱۰۹	۰/۰۰۰۱	۰/۰۲۱۱	۰/۰۹۹۹	۰/۱۰۴۲
p_1	۰/۱	۰/۱۰۳۴	/۰۱۸۲	۰/۰۰۰۱	۰/۰۳۴۱	۰/۰۹۹۸	۰/۱۰۶۹
ϕ	۴۰	۳۹/۷۸۸۴	۲/۵۴۱۶	۶/۴۳۹۹	-۰/۰۰۵۲	۳۹/۲۹۰۳	۴۰/۲۸۶۶
σ_b^2	۵	۵/۰۱۳۱	/۰۷۹۷۷	۰/۰۰۶۴	۰/۰۰۲۶	۴/۹۹۷۴	۵/۰۲۸۷۶

همچنین تحلیل حساسیت با تغییر ضرایب رگرسیونی، پارامتر دقت توزیع بتا و پارامتر دقت اثر تصادفی σ_b^2، $\sigma_b^2 \sim IG(k,k)$ و $\phi \sim G(k,k)$ برای $k \in \{۰/۱, ۰/۰۱\}$ انجام شد که در هر مورد تغییر اساسی در نتایج حاصله نسبت به نتایج ارائه شده در جدول ۲ ملاحظه نشد. به عبارتی با توجه به نتایج مطالعات شبیه‌سازی می‌توان گفت فراورده‌ای مدل رگرسیون آمیخته بتای افزوده برای حالتی که داده‌ها در هر تکرار ممکن است مقداری بین صفر و یک، صفر یا یک را به خود اختصاص دهند یا به عبارتی تابع تکرار قبلی نیستند، قابل اعتماد است.

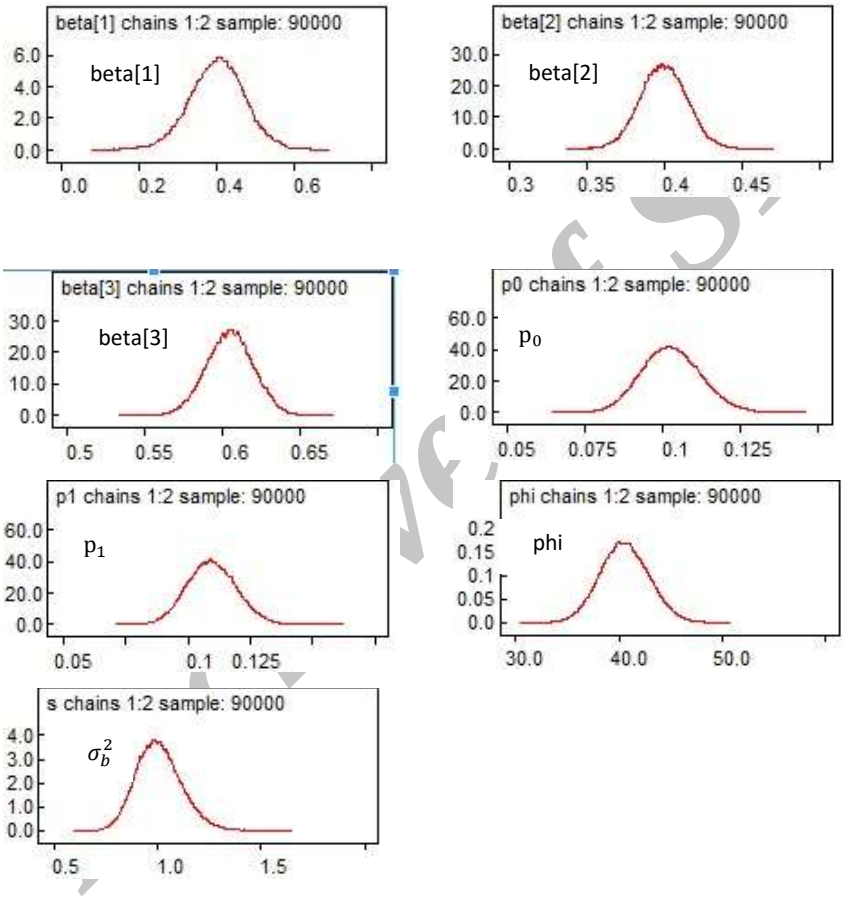


شکل (۲): نمودارهای اثری برای داده‌های شبیه‌سازی شده با اندازه نمونه ۲۰۰ تایی

۵- بحث و نتیجه‌گیری

در این مقاله مدل رگرسیون آمیخته بتای افزوده به‌عنوان تعمیمی از مدل رگرسیون بتای افزوده مطالعه و برای حالتی که مقادیر متغیر وابسته در تکرارهای مختلف لزوماً تابع تکرار قبلی نیستند، معرفی شد. در مطالعه شبیه‌سازی با بررسی اریبی نسبی، میانگین توان دوم خطا و آزمون‌های همگرایی تشخیصی، کارایی این مدل موردبررسی قرار گرفت. برای تحلیل داده‌های آمارگیری نیروی کار مرکز آمار ایران، داده‌های مربوط به خانوارهای مشترک در فصل‌های اول و دوم سال‌های ۱۳۹۲ و ۱۳۹۳ شهر تهران، در نظر گرفته و نسبت شاغلین در خانوار مدل‌بندی شد.

نتایج حاصل از این مدل‌بندی حاکی از آن است که تعداد افراد ۱۰ ساله و بیشتر در هر خانوار با ضریب مثبت و جمعیت خانوار با ضریب منفی به‌عنوان متغیرهای تبیینی معنی‌دار در این مدل حضور دارند. ازآنجاکه مهم‌ترین مشکل برآورد وضعیت نیروی کار در سطح کمتر از استان، عدم وجود متغیرهای کمکی مناسب است یکی از موارد کاربرد این مدل ساختن متغیر تبیینی مناسب برای برآورد وضعیت نیروی کار نواحی کوچک است زیرا با ضرب سهم شاغلین در خانوار در تعداد خانوار این نواحی، تعداد شاغلین این نواحی قابل برآورد است. بنابراین حتی اگر این برآوردها را با نگاهی محافظه‌کارانه نتوان مستقیماً به‌عنوان فراورده‌ای شاغلین نواحی موردنظر در نظر گرفت، می‌توان به‌عنوان متغیرهای تبیینی مطلوب برای استفاده از روش‌های نواحی کوچک به‌منظور برآورد وضعیت نیروی کار نواحی کوچک، در نظر گرفت.



شکل (۳): نمودارهای چگالی پسین برای داده‌های شبیه‌سازی شده با اندازه نمونه ۲۰۰ تایی

همچنین از این مدل‌بندی در بحث کنترل کیفیت نتایج آمارگیری نیروی کار استانی نیز می‌توان کمک گرفت. بررسی اثر تصادفی با توزیع t–استیودنت در مدل رگرسیون آمیخته بتای افزوده، مدل‌بندی هم‌زمان پارامترهای میانگین، دقت، p_0 و p_1 در این مدل رگرسیونی برای حالت‌های مختلف و همچنین بررسی رفتار مدل با تابع‌های پیوند مختلف، مسائل مهمی هستند که لازم است در مطالعات آتی به آن‌ها پرداخته شوند.

تقدیر و تشکر

نویسندگان از داوران محترم مجله که نظرات ارزنده آن‌ها موجب بهبود مقاله شد و از حمایت قطب علمی تحلیل داده‌های فضایی-زمانی دانشگاه تربیت مدرس قدردانی می‌نمایند.

منابع

[1] Paolino, P. (2001). Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables, *Political Analysis*, **9**, 325-346.

[2] Kieschnik, R. and McCullough, B.D. (2003). Regression Analysis of Variates Observed on (0,1): Percentage, Proportions and Fractions. *Statistical Modelling*, **3**, 193-213**.**

[3] Ferrari, S. and Cribari, F. (2004). Beta Regression for Modelling R ates and Proportions, *Journal of Applied Statistics*, **31**, 799-815.

[4] Cepeda, E.D. and Gamerman, D. (2005). Bayesian Methodology for Modeling Parameters in the Two Parameter Exponential Family, *Revista Estadística*, **57,** 168-169.

[5] Smithson, M. and Verkuilen, J. (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables, *Psychological Methods*, **11**, 54-71.

[6] Branscum, A.J., Johnson, W.O. and Thurmond, M. (2007). Baysian Beta Regression Applications to Houshold Expenditure Data and Genetic Distance Between Food and Mouth Dieseas Viruses, *Australian & New Zealand Journal of Statistics*, **49**, 287-301.

[7] Ospina, R. and Ferrari, S. (2010). Inflated Beta Distributions. *Statistical Paper*, **23,** 193-213**.**

[8] Ospina, R. and Ferrari, S. (2012). A General Class of Zeror-One Inflated Beta Regression Models. *Computational Statistics & Data Analysis*, **56**, 1609-1623.

[9] Galvis, M.D., Dipankar, B. and Victor, H.L. (2014). Augmented Mixed Beta Regression Models for Periodontal Proportion Data, Preprinted, *Statistics in Medicine,* **33,** 3759-3771.

[0] Figueroa-Zúñiga, J.I., Arellano-Valle, R.B. and Ferrari, S.L. (2013). Mixed Beta Regression: A Bayesian Perspective, *Computational Statistics & Data Analysis*, **61**, 137– 147.

[11] Sturtz, S., Ligges, U. and Gelman, A. (2005). R2winbugs: A Package for Running Winbugs from R, *Journal of Statistical Software*, **12**, 1–16.

[۱۲] نتایج آمارگیری نیروی کار (۱۳۹۲)، مرکز آمار ایران، تهران.

[13] Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science* **7**, 457–511.

[14] Heidelberger, P. and Welch, P.D. (1981). A Spectral Method for Confidence Interval Generation and Run Length Contral in Simulations, *Communications of the ACM*, **24**, 233-245.

[15] Brooks, S.P. and Gelman, A., (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.

[16] Plummer, M., Best, N., Cowles, K. and Vines, K., (2006). The coda Package. R Project. http://cran.r-project.org/doc/packages/coda.pdf.

Augmented Mixed Beta Regression and Modeling of Employed Proportions in Households

Zohre Fallah Mohsenkhani and Mohsen Mohammadzadeh

Department of Statistics, Tarbiat Modares University, Tehran, Iran

Abstract

The Beta regression model is usually used for modeling the rates or proportions confined in an open interval (0,1). In some studies, the data may also include zero and one. In this paper, an augmented Beta regression model that is a mixture of Beta distribution with two degenerated distributions at 0 and 1 is presented for rates or proportions confined in [0,1]. For the augmented mixed Beta model with reparametrization of Beta distribution, the mean and precision parameters were modeled including fixed and random effects. This is while taking into account that the random effects make these models applicable to correlated data. Here, the augmented mixed Beta model is presented. Then this model is evaluated in a simulation study. Next, the application of this model is shown for analyzing the proportions of employed persons in every household. Finally, conclusion and results are presented.

Keywords: Augmented Beta Regression; Bayesian Analysis; Labour Force Survey; Beta Distribution; Mixed Model.

Mathematics Subject Classification (2010):12J15, 62F62.