Universitatea din Craiova

Facultatea de Economie și Administrarea Afacerilor

Școala Doctorală de Științe Economice

Selecția Bayesiană a variabilelor și estimarea coeficienților în modelul de regresie cuantilică Tobit

O teză prezentată pentru obţinerea titlului de Doctor in Stiinţe in domeniul: Cibernetică și statistică

Coordonator științific,
Prof. Univ. dr. Vasile Georgescu

Doctorand,
Fadel Hamid Hadi Alhusseini

Craiova

2017

Rezumat

Scopul acestei teze este de a efectua selecţia Bayesiană a variabilelor și estimarea coeficienților în modelul de regresie cuantilică Tobit prin intermediul unor noi metode propuse.

În capitolul 2, a fost propusă o nouă formulare pentru regresia cuantilică (Qreg) Bayesiană prin utilizarea unei formulări bazat pe o mixtură scalată de distribuţii uniforme. Apoi, a fost dezvoltată o abordare Bayesiană completă care conduce la un algoritm simplu și eficient de eșantionare Gibbs cu distribuții aposteriori condiționate uşor de manevrat. Superioritatea abordării propuse (Lasso U) a fost demonstrată prin simulare și pe date reale. Unele extensii la această abordare sunt discutate în regresia cuantilică Tobit.

În capitolul 3, a fost propusă o nouă metodă de regresie cuantilică Tobit Lasso Bayesiană pentru selecția variabilelor și estimarea coeficienților care asignează o mixtură scalată independentă de distribuții uniforme (SMU) pentru coeficienții de regresie. Apoi, a fost prezentat un algoritm MCMC (Markov Chain Monte Carlo) simplu și eficient pentru eşantionatorul Bayesian. Atât studii bazate pe simulare cât și un set real de date au fost utilizate pentru a investiga performanța metodei propuse. Ambele arată că metoda propusă se comportă destul de bine în comparație cu celelalte metode într-o varietate de scenarii.

În capitolul 4, a fost elaborat un MCMC simplu și eficient pentru modelul de regresie cuantilică Tobit compozit, bazat pe o mixtură între o distribuție exponențială şi una normală scalată a distribuției Laplace asimetrice. Studiile de simulare arată că metoda propusă este eficientă în estimarea coeficienților cu distribuții diferite. Pe baza studiilor de simulare și a analizei datelor reale, susținem că este necesar să combinăm informațiile bazate pe estimatori la diferitele cuantilice pentru a obține un câștig de eficiență.

În capitolul 5, modelul Bayresian Tobit de regresie cuantilică și modelul Bayesian Tobit cuantilic compozit au fost folosite pentru a analiza datele privind investițiile băncilor irakiene, în două moduri. În primul rând, estimarea coeficienților prin treizeci de nivele cuantile Tobit. În al doilea rând, selectarea variabilelor prin determinarea importanței relative a variabilelor independente din modelul nostru, prin intermediul a treizeci de nivele cuantile Tobit. Pe de altă parte, abordarea bazată pe modelul Bayesian compozit cuantilic Tobit este utilizată prin intermediul a șase grupuri de nivele cuantilice compozite Tobit, de asemenea în două moduri. În primul rând: pentru modelarea relației dintre investițiile băncilor irakiene și nouă variabile independente. În al doilea rând: pentru selecția variabilelor prin un grup de șase nivele cuantilice compozite Tobit. Un set de concluzii a fost dedus din punct de vedere teoretic și aplicativ.

Cuvinte cheie

- regresie Tobit
- regresie cuantilă
- regresie Tobit cuantilică
- Estimarea coeficienților
- Selecția variabilelor
- Abordare Bayesiană
- Eșantionare Gibbs
- Algoritmul MCMC
- Studii de simulare
- Date reale
- Distributii a priori
- Distributii **SMN**
- Distributii **SMU**
- Bănci irakiene
- Investiții bancare

Cuprins

Direcţii de cercetare şi principale contribuţii

Modelarea relației dintre mediile unei variabile răspuns (variabila dependentă) Y cu setul de covariate X nu este întotdeauna convenabilă. În multe studii aplicative, regresia mediilor poate fi inadecvată pentru a descrie comportamentul variabilei răspuns (variabila rezultat) Y cu covariatele X. De exemplu, efectul proprietăților demografice și al comportamentului maternal asupra greutății copiilor născuți a fost studiat de către (Abrevaya, J., & Dahl, CM (2008)) [6] în Statele Unite. Acest studiu a fost axat pe greutatea scăzută la naștere pentru sugari, care provoacă multe probleme de sănătate. Aceste date au fost analizate prin regresia standard a mediilor; media condiționată nu a fost o abordare atractivă pentru distribuţia cozii inferioare (low tail). Regresia cuantilică (Q Reg) a fost propusă de Koenker și Bassett (1978) [38] ca o extensie pentru regresia standard a mediei în diferite cuantile condiționate ale variabilelor dependente.

Modelul de regresie cuantilică este capabil sa furnizeze informatii complete despre diferitele cuantile ale relatiilor stochastice intre variabila dependentă si predictori. Recent, modelul Q Reg a primit o atenție majoră în studiile teoretice și aplicative. Modelul Q Reg este aplicat în diferite domenii cum ar fi: genetică (Microarray) (Wang and He, (2007)) [69], economia agricolă (Kostov și Davidova, 2013) [41], studii ecologice (Cade and Noon, (2003)) [14], indicele de masă corporală (Yu et al., 2013), și așa mai departe "Fadel Hamid Hadi Alhusseini, 2017" [24]. Modelele de regresie cuantilică au proprietăți bune în comparație cu alte modele de regresie. Modelul Q Reg aparține unei familii de modele robuste (Koenker și Geling, (2001)) [43]. Modelul de regresie cuantilică nu necesită nicio ipoteză cu privire la distribuția reziduală, oferind o eficiență statistică mai mare decât alte modele de regresie atunci când eroarea este non-normală. "Fadel Hamid Hadi Alhusseini, 2017" [24]. De asemenea, modelul Q Reg este robust în raport cu problemele economice. Toate aceste caracteristici au făcut modelul Q Reg un model important în domeniile de aplicare. Următoarea formulă matematică descrie modelul Q Reg.

$$y_i = x_i^T \beta_\theta + \varepsilon_i, \qquad \theta \in (0,1), \qquad [1]$$

Pentru fiecare a θ-a cuantilă $(0 < \theta < 1)$, a θ-a regresie cuantilică poate fi notată prin $Q_{y_i|x_i}(\theta) = x'_i \beta_\theta$, unde y_i este variabila de răspuns, x_i^T este un vector de dimensiune k al covariatelor, β_θ este vectorul coeficienţilor modelului Q Reg.

Există un număr infinit de puncte cuprinse în intervalul $(0 < \theta < 1)$. Deci există un număr infinit de nivele cuantilice. Modelul Q Reg este estimat în fiecare nivel cuantilic. Prin urmare, modelul Q Reg are o flexibilitate foarte mare și capacitatea de a furniza informații cu privire relațiilor dintre varibila răspuns și predictori la diferite nivele cuantilice, spre deosebire de modelul classic de regresia în care numai o linie de regresie este estimată prin media condiționată a variabilei răspuns (y) atunci când se dă x, adică $E\,(y|x)$. (Koenker și Hillock(2001))[40]. A se vedea figura următoare.

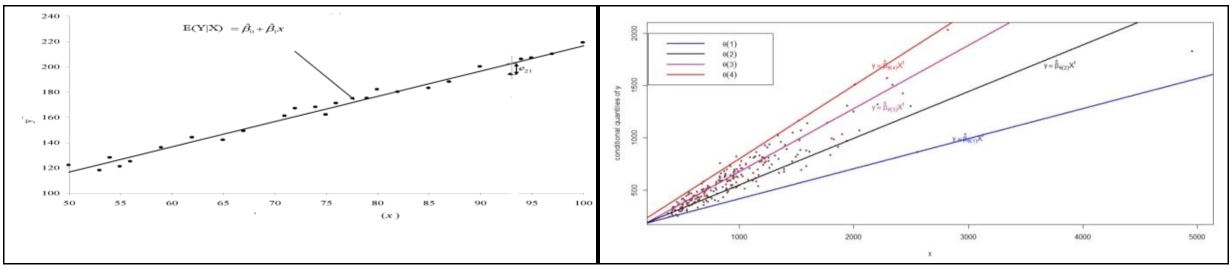

Figura 1: Linii de regresie estimate prin modelul de regresie classic şi modelul Q Reg.

Figura din stânga prezintă linia de regresie estimată de modelul de regresie clasică. Uneori, modelul de regresie clasică nu ne poate oferi informații complete despre relația dintre variabila dependentă și covariate (variabilele independente). Figura din dreapta prezintă patru linii de regresie cuantilică estimate în patru niveluri cuantilice. Întotdeauna, modelul Q Reg ne dă o imagine clară despre relația dintre o variabilă de răspuns și covariate (variabilele explicative), deoarece sunt estimate numeroase linii de regresie cuantilice. Fiecare linie aparține unui nivel cuantilic specific. Problema discutată se concentrează asupra proprietăților modelului Q Reg. O altă problemă importantă în construirea modelului Q Reg este selecția covariatelor active. În ultimii ani, selectarea subseturilor importante de covariate a primit o mare atenție în literatura de specialitate. Multe metode de selecție a variabilelor au fost propus, cum ar fi: metoda Lasso (Tibshirani, (1996)) [64], metoda SCAD (Fan and Li, (2001)) [20], metoda reţelei elastice (elastic net) (Zou and Hastie, 2005). [75], metoda Lasso adaptivă (Zou, (2006)). Tehnicile de selecție a variabilelor sunt utilizate și în modelul Q Reg, cum ar fi penalizarea Lasso, care a fost aplicată modelului Q Reg cu efecte mixte sau date longitudinale de către (Koenker (2004)) [39]; o cale de soluție a fost introdusă de către (Li și Zhu (2008)) [53] la modelul de regresie cuantilică cu normă L_1 penalizată. Pentru a estima coeficienții modelului Q Reg Lasso, (Li și Zhu (2008)) [53] au propus următoarea ecuație:

$$\min_{\beta_\theta} \sum_{i=1}^{n} \rho_\theta(y_i - x_i^T \beta_\theta) + \lambda \parallel \beta_\theta \parallel, \qquad [2]$$

unde $\rho_\theta(s)$ este funcția de verificare (pierdere) definită de $\rho_\theta(s) = s\{\theta - \mathrm{I}(s \leq 0)\}$, şi unde $I(s < 0)$ este funcția indicator iar $\lambda((\lambda \geq 0)$ este parametrul de contracție. Din păcate, ecuația (2) nu este diferenţiabilă în zero, deci nu există o soluție exactă pentru ecuația (2). (Koenker, (2005)) [42] arată că minimizarea lui (2) poate fi obținută printr-un algoritm de programare lineară (Koenker și D'Orey, (1987)) [46].

De asemenea, pentru a estima coeficienții modelului Q Reg Lasso, se poate folosi abordarea Bayesiană. (Park și Casella (2008)) [60] au propus modelul Lasso Bayesian în modele clasice de regresie liniară prin utilizarea unei mixturi scalate de distribuții a priori normale (SMN) pentru coeficienții de regresie și distribuții a priori exponențiale pentru dispersiile lor "Fadel Hamid Hadi Alhusseini 2017 " [24]. (Li et al, (2010)) [52] au sugerat un model Bayesian Q Reg Lasso, de asemenea prin utilizarea distribuțiilor a priori (SMN) pentru coeficienții de regresie și a distribuțiilor a priori exponențiale pentru dispersiile lor. (Li et al, (2010))[52] a propus un algoritm Markov Chain Monte Carlo (MCMC) simplu și efficient pentru updatarea tuturor coeficienților modelelor din distribuția aposteriori. Formularea Bayesiană este o procedură flexibilă de estimare a parametrului de penalizare împreună cu coeficienții de regresie. Recent, pentru regresia liniară, (Mallick and Yi (2014))[55] a furnizat un proces diferit al modelului bazat pe Lasso prin utilizarea mixturilor scalate de distribuţii uniforme (SMU) ca formulare a funcției de densitate Laplace. (Mallick și Yi (2014)) [55] au oferit o nouă metodă de estimare a coeficienților și de selecție a variabilelor în modelul de regresie clasică. Această metodă a performat destul de bine în comparație cu alte metode existente în același domeniu.

Contribuția noastră constă într-o nouă formulare pentru modelul Bayesian Q Reg Lasso, prin utilizarea formulării SMU ca o nouă distribuţie a priori. Am dezvoltat, de asemenea, un tratare Bayesiană completă, care a condus la un algoritm Gibbs de eșantionare eficient și simplu cu distribuții aposteriori condiționate complete, uşor de manevrat. Distribuțiile aposteriori condiționate complete ale algoritmului nostru de eșantionare Gibbs au fost colectate în doi paşi. În primul rând, funcția de probabilitate care aparţine familiei de distribuţii Laplace asimetrice (ALD), a se vedea (Yu and Moyeed (2001)) []. Aici, nu putem folosi direct ALD deoarece ar duce la dovezi eronate și la un algoritm Gibbs de eșantionare ineficient. Prin urmare, am folosit o formulă alternativă a ALD, propusă de Kozumi și Kobayashi, (2011) [49] care este o mixtură scalată de distribuții normale (SMN) "Fadel Hamid Hadi Alhusseini" [24]. Variabila răspuns (y_i) urmează o distribuție normală cu media ($x_i^T \beta_\theta + (1 - 2\theta) m_i$) și varianța ($2\sigma^{-1} m_i$) as $\sim N(x_i^T \beta_\theta + (1 - 2\theta) m_i, 2\sigma^{-1} m_i)$:

$$f(y|y_i, x_i\beta, \sigma) = \frac{1}{\sqrt{4\pi\sigma^{-1} m_i}} Exp \frac{-(y_i - x_i^t \beta - (1 - 2\theta) m_i)^2}{4\sigma^{-1} m_i} \qquad [3]$$

unde m_i este distribuția exponențială cu parametrul de scară $\theta(1 - \theta)\sigma$. Majoritatea cercetătorilor din domeniul regresiei cuantilice bayesiene și regresiei cuantilice normalizate Bayesiene utilizează formularea Kozumi și Kobayashi. Aceasta oferă un algoritm MCMC simplu și eficient.

Este contribuția noastră de a utiliza distribuții a priori SMU pentru β_j și distribuții a priori exponențiale pentru σ^2 , presupunând că acestea sunt independente. Am dezvoltat un model alternativ Q Reg Bayesian Lasso ierarhic. Potrivit lucrării lui (Mallick și Yi (2014)) [55], distribuția a priori Laplace a lui β_j poate fi dată de:

$$\frac{\lambda}{2} e^{\{-\lambda|\beta_j|\}} = \frac{\lambda}{2} \int_{u_j > |\beta_j|}^{\cdot} \lambda \; exp\{-\lambda u_j\} \, du_j \quad ,$$

$$= \int_{-u_j < \beta_j < u_j}^{\prime} \frac{1}{2u_j} \frac{\lambda^2}{\Gamma(2)} {u_j}^{2-1} \, exp\{-\lambda u_j\} \, du_j \quad , \qquad [4]$$

$$= \int_{-u_j < \beta_j < u_j}^{\prime} \frac{\lambda^2}{2} \, exp\{-\lambda u_j\} \, du_j$$

unde $\Gamma(2) = (2 - 1)! = 1$.

Ecuația (4) consideră reformularea distribuției a priori Laplace în cadrul SMU. Această procedură ne oferă dovezi confortabile pentru eșantionarea noastră Gibbs și atrăgătorul algoritm MCMC. β_j urmează o distribuție uniformă cu parametrii $(-u_j, u_j)$, iar u_j o distribuția Gamma cu parametrul formei (2) și parametrul ratei (λ); deasemenea parametrul λ urmează distribuția Gamma cu parametrul formei (c) și parametrul ratei (d) "Fadel Hamid Hadi Alhusseini" [24]. Din informațiile de mai sus, modelul nostru Bayesian ierarhic poate fi formulat după cum urmează:

$$\begin{aligned}
& m_i|\sigma \sim exp\{\theta(1 - \theta)\sigma\}, \\
& \beta_j|u_j \sim Uniform\,(-u_j, u_j), \\
& u_j|\lambda \sim Gamma(2, \lambda)\,, \qquad [5] \\
& \sigma \sim \sigma^{a-1} exp(-b\lambda) \\
& \lambda \sim \lambda^{c-1} exp(-d\lambda)\,.
\end{aligned}$$

Aici, $exp(\theta(1 - \theta)\sigma)$ se referă la distribuția exponențială cu parametru ratei $\theta(1 - \theta)\sigma$.

De asemenea σ urmează o distribuție Gamma cu parametrul formei (a) și parametrul ratei (b), unde a, b, c și d sunt patru hiperparametri fixați. Din funcția de verosimilitate prezentată în ecuația (3) şi modelul ierarhic descris în ecuația (5) vom obține o distribuție aposteriori condiționată a lui $\beta, m, (m_1, \dots m_n)^T, u = (u_1, \dots u_n)^T$ iar λ poate fi actualizat utilizând o tehnică eficientă de calcul bazată pe MCMC.

Distribuția aposteriori condiționată a lui β este normală trunchiată, cu media $\left(\sigma_{\beta_j}^2 \sum_{i=1}^{n} \frac{\sigma x_{ij} \left(y_i - (1-2\theta)m_i - \sum_{,j\neq l}^{p} x_{ij}\beta_j\right)}{2m_i}\right) I\{|\beta_j| < u_j\}$ și varianța $\left(\sum_{i=1}^{n} \frac{\sigma x_{ij}^2}{2m_i}\right)^{-1}$. Distribuția aposteriori condiționată a lui m este distribuția Gaussiană inversă cu media $1/\sqrt{(y_i - x_i^t\beta)^2}$ și parametrul formei $(\frac{\sigma}{2})$. Distribuția aposteriori condiționată a lui u_j este o distribuție exponențială trunchiată la stânga, dată de $u_j|\beta, \lambda \sim Exp(\lambda) I\{u_j > |\beta_j|\}$. Distribuția aposteriori condiționată a parametrului de penalizare λ este distribuția Gamma cu parametru formei $(c + 2p)$ și parametrul ratei $(d + \sum_{j=1}^{p} |\beta_j|)$. De asemenea, distribuția aposteriori condiționată a lui σ este distribuția Gamma cu parametrul formei $(a + \frac{3n}{2})$ și parametrul ratei $(\sum_{i=1}^{n} \left(\frac{\left(y_i - x_i^t\beta + (1-2\theta)m_i\right)^2}{4m_i} + \theta(1-\theta)m_i\right) + b)$. Din distribuțiile aposteriori condiționate complete procedăm la eșantionarea fiecărui parametru necunoscut $(\beta, m, (m_1, \dots m_n)^T, u = (u_1, \dots u_n)^T$, λ și σ. Vom obtine un bun eșantionator Gibbs pentru estimări ale coeficienților și selecția variabilelor în noul model Bayesian de regresie cuantilică Lasso.

Pentru a evalua performanța metodei propuse de noi, New Bayesian Lasso Quantile Regression (LassoU), aceasta va fi comparată cu alte trei metode (LassoN, rq și MCMCquantreg) prin studii de simulare și date reale, unde studiile de simulare sunt descrise după cum urmează:

În prima simulare, datele noastre de simulare sunt generate de un model rar (sparse):

$$y_i = 3x_{1i} + 1x_{2i} + 2x_{5i} + \varepsilon_i, \qquad [i = 1,2, \dots ,100]$$

unde y_i este variabila de răspuns, iar parametrii noștri adevărați sunt $\beta = (0,3,1,0,0,2,0,0,0)$. Cele 8 covariate $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i})$ sunt simulate dintr-o distribuție normală multivariată cu media 0 și $cov(x_h, x_g) = 0.5^{|h-g|}$. Distribuțiile erorii aleatoare (ε_i) sunt generate dintr-o distribuție $\chi_{(3)}^2$ cu trei grade de libertate, a distribuție $t_{(3)}$ cu trei grade de libertate și o distribuție normală cu media (μ) și varianța (9), $\varepsilon_i \sim N(\mu, 9)$. Metodele supuse testării sunt evaluate prin mediana abaterilor absolute medii (MMAD), unde $\text{MMAD} = median\left(mean\left(|x^T\hat{\beta} - x^T\beta^{true}|\right)\right)$, și deviațiile standard (SDs). În primul studiu de simulare și sub trei alegeri ale distribuției erorilor, algoritmul nostru a fost rulat pentru 11000 iterații și primele 1000 au fost eliminate. Rezultatele au fost după cum urmează: performanța metodei propuse New Bayesian Lasso Q Reg (LassoU) pare foarte bună în comparație cu metodele Bayesiane și non-Bayesiane (LassoN, rq și MCMCquantreg). In general, MMAD generată de New Bayesian Lasso Q Reg este mult mai mică decât MMAD generată de celelalte trei metode (LassoN, rq și MCMCquantreg), la toate nivelele cuantilice și toate distribuțiile luate în considerare. De asemenea, în primul studiu de simulare cu diferite distribuții ale erorii, SD obținut prin metoda propusă este mult mai mic decât SD obținut prin celelalte abordări (LassoN, rq și MCMCquantreg). Algoritmul nostru MCMC este foarte stabil și acest lucru este clar din coeficienții factorului de reducere a scării potențiale multivariate (MPSRF), unde MPSRF este stabil și se apropie de 1 după aproximativ 2000 de iterații. Aceasta arată că convergența distribuției aposteriori condiționate complete pentru metoda propusă a fost rapidă, iar mixtura a fost bună.

În a doua simulare, datele noastre de simulare sunt generate de modelul rar (sparse):

$$y_i = 3x_{1i} + \varepsilon_i, \qquad [i = 1,2, \dots ,100]$$

unde parametrii adevărați sunt $\beta = (0,3,1,0,0,2,0,0,0)^T$. Ca și înainte, 8 covariate $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i})$ sunt simulate dintr-o distribuţie normală multivariată cu media 0 şi $cov(x_h, x_g) = 0.5^{|h-g|}$. MMAD şi SD sunt calculate prin LassoU, care este mult mai mic decât MMAD calculat prin alte trei metode (LassoN, rq şi MCMCquantreg), prin intermediul a trei niveluri cuantilice și trei distribuții diferite de erori. Din rezultatele celei de-a doua simulări concluzionăm că metoda propusă (LassoU) are o performanță bună comparativ cu metodele Bayesiene și non-Bayesiene (LassoN, rq și MCMCquantreg).

În cea de-a treia simulare, datele noastre de simulare sunt generate de modelul dens

$$y_i = x_i^T \beta_\theta + \varepsilon_i, \qquad [i = 1,2, \dots ,100]$$

unde $\beta = (0.00,0.85,,0.85,,0.85,,0.85,,0.85,,0.85,,0.85,,0.85)^T$. La fel ca în prima și a doua simulare, opt covariate $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i})$ sunt simulate dintr-o distribuţie multivariată normală cu media 0 și $cov(x_h, x_g) = 0.5^{|h-g|}$. MMAD și SD sunt calculate de LassoU și sunt mult mai mici decât MMAD calculat prin metodele Bayesiene și non-Bayesiene (LassoN, rq și MCMCquantreg), prin intermediul a trei niveluri cuantilice și trei distribuții diferite de erori. Din rezultatele celei de-a treia simulări concluzionăm, de asemenea, că metoda propusă (LassoU) are o performanță bună comparativ cu metodele Bayesiene și non-Bayesiene (LassoN, rq și MCMCquantreg).

De asemenea, din cele trei rezultate ale simulării, metoda propusă (LassoU) are cele mai mici MMAD și SD, comparativ cu celelalte trei metode. Prin urmare, metoda propusă are o precizie mai mare decât celelalte metode în estimarea coeficienților și selectarea variabilelor.

De asemenea, pentru evaluarea metodei propuse, am putea folosi estimarea lui β în mod direct. Alegem doar cazul erorii aleatoare $\varepsilon_i \sim N(\mu, 9)$, la trei nivele cuantilice diferite ($\theta_1 = 0.50$, $\theta_2 = 0.75$, and $\theta_3 = 0.95$). Estimările privind β prin metoda propusă, new Bayesian Lasso Q Reg (LassoU), este foarte apropiat de valorile adevărate ale parametrilor, în comparație cu alte metode cum ar fi LassoN, rq și MCMCquantreg. În general, LassoU se comportă bine în estimarea coeficienților de regresie în comparație cu metodele Bayesiene și non-Bayesiene (LassoN, rq și MCMCquantreg). Scenariile noastre de simulare arată că LassoU este eficientă în selectarea variabilelor și estimarea coeficienților în modelul de regresie cuantilică. Studiile de simulare arată, de asemenea, că metoda propusă (LassoU) este robustă atunci când distribuția termenului de eroare nu este ALD " Fadel Hamid Hadi Alhusseini, 2017" [24].

Metoda propusă este evaluată cu date reale. În acest scop, am folosit date privind poluarea aerului care conține variabila de răspuns exprimată în logaritm(concentrația de NO2 pe oră) și șapte covariante - x_1 (log (nr. de mașini pe oră)), x_2 (temperatură), x_3 (viteza vântului în metri pe secundă), x_4 (diferența de temperatură), x_5 (direcția vântului), x_6 (durata zilei în ore) și x_7 (numărul zilei). Pentru evaluarea performanței metodei LassoU, am comparat-o cu alte trei metode (LassoN, rq și MCMCquantreg). Criteriul erorii medii pătratice (MSE) a fost utilizat prin intermediul a trei niveluri cuantilice $\theta \in \{0.50, 0.75, 0.95\}$. În general, (MSE) și SD calculate de către noul nostru model Bayesian Lasso QReg au fost mai mici decât (MSE) calculate prin celelalte trei metode (LassoN, rq și MCMCquantreg) în cazul majorității nivelelor cuantilice. Prin urmare, metoda noastră modelul Bayesian Lasso QReg are o performanță mai bună comparativ cu metodele Bayesiene și non-Bayesiene (LassoN, rq și MCMCquantreg). Atât din scenariile de simulare cât și datele reale rezultă că metoda propusă de noi, adică modelul Bayesian Lasso Q Reg, poate fi considerată o nouă extensie la estimarea coeficienților și selecția variabilelor în modelul QReg.

Am propus, de asemenea, o nouă contribuție la estimarea coeficienților și selecția variabilelor în modelul de regresie Tobit Qreg (Tobit Q Reg). De când a fost propus de Tobin (1958) [66], modelul de regresie Tobit a devenit cunoscut în variate domenii ştiinţifice (Greene, W (2010)) [31], cum ar fi științele economice, științele medicale, științele sociale și științele inginerești. Modelele de regresie Tobit oferă informații limitate privind variabila de răspuns. Prin urmare, în punctul cenzurat egal cu zero ($c = 0$), modelul de regresie Tobit este definit ca:

$$y_i = \begin{cases} T_i^* = \alpha + \beta x_i^T + \varepsilon_i & if\ T_i^* > 0 \\ 0 & if\ T_i^* \le 0 \end{cases} \qquad [6]$$

unde $\varepsilon_i \sim N(0, \sigma^2)$, $[i = 1,2, \ldots\ldots, n]$

y_i este variabila de răspuns cenzurată, limitată în punctul cenzurat egal cu 0. T_i^* este variabila latentă, x_i^T este este un vector $1\times k$ de variabile explicative (covariate), α este termenul intercepţie, β este vectorul parametrilor necunoscuți în modelul de regresie Tobit şi ε_i este un termen de eroare aleatoare, ce urmează o distribuție normală cu medie zero și varianța (σ^2). Prin urmare, variabila latentă T_i^* este distribuită normal cu media ($\alpha + \beta x_i$) și varianța (σ^2), unde $T_i^* \sim N(\alpha + \beta x_i, \sigma^2)$. Din ecuația (6) , modelul de regresie Tobit conține două părți. În primul rând, atunci când variabila latentă T_i^* este observată, adică $T_i^* > 0$, funcția densitate de probabilitate (pdf) a variabilei latente T_i^* aparţine variabilei latente observate (observație non-limită) unde $y_i = T_i^*\ \ if\ T_i^* > 0$ şi ea este:

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{\frac{-\left(y_i - \alpha + x_i^T\beta\right)^2}{2\sigma^2}} \qquad [7]$$

Ecuația (1.7) poate fi reformulată ca în ecuația (8)

$$f(y_i) = \frac{1}{\sigma}\phi\left(\frac{y_i - (\alpha + x_i^T\beta)}{\sigma}\right) \qquad [8]$$

unde $\phi(.)$ este o funcție de densitate a probabilitate (pdf) care aparține variabilei latente observate. A doua parte este dedicată variabilei latente neobservate (T_i^*) când $T_i^* \le 0$, Deci, va urma funcția de distribuție cumulativă a distribuției normale. $pro(y_i = 0)\ \ if\ \ \ pro(T_i^* \le 0) \longrightarrow \Phi\left(\frac{y_i - (\alpha + x_i^T\beta)}{\sigma}\right) = \Phi\left(\frac{0 - (\alpha + x_i^T\beta)}{\sigma}\right)$

$$= \Phi\left(\frac{-(\alpha + x_i^T\beta)}{\sigma}\right) = 1 - \Phi\left(\frac{(\alpha + x_i^T\beta)}{\sigma}\right) \qquad [9]$$

$\Phi(.)$ este o funcție de distribuție cumulativă (cdf) (Greene, W (1999)) [30]. Modelul de regresie Tobit este o funcție mixtă între funcția densității de probabilitate și funcția de distribuție cumulativă a distribuției normale. Modelul de regresie Tobit este un bun instrument statistic pentru modelarea relației dintre variabilele de răspuns cenzurate și un set de variabile explicative (covariate). Dar modelul de regresie Tobit este foarte sensibil la problemele de regresie. Modelul de regresie Tobit nu este suficient atunci când datele conțin valori aberante. De asemenea, modelul de regresie Tobit nu poate furniza estimatori buni atunci când ipotezele normale nu sunt realizate. Pentru a evita aceste probleme dure, modelul Tobit de regresie cuantilica (Tobit Q Reg) propus de Powell (1986) [59] are mai multe proprietăți. Tobit Q Reg analizează toate caracteristicile distributive condiționate ale variabilei dependente (răspuns). Modelul Tobit de regresie cuantilică este axat pe un set de funcții cuantilice la nivele diferite de cuantile. Prin urmare, are flexibilitate și capacitatea de a ne da o imagine completă a distribuției complete a relației dintre variabila de răspuns și variabilele explicative. Modelul Tobit Q Reg este o extensie normală a modelului clasic de regresie Tobit. Modelul Tobit Q Reg are proprietăți atractive, ceea ce îl face un model important pentru descrierea relației dintre variabila de răspuns cenzurată și un set de variabile explicative. Formularea matematică a modelului Tobit Q Reg poate fi scrisă ca:

$$y_i = \begin{cases} T_i^* = \alpha_\theta + x_i^T \beta_\theta + \varepsilon_i & if\ T_i^* > 0 \\ 0 & if\ T_i^* \le 0 \end{cases} \quad [10]$$

unde θ is $(0 < \theta < 1)$. Modelul Tobit Q Reg poate avea o altă formulare matematică: $y = \max(C, T_i^*)$, $\quad unde\ T_i^* = \alpha_\theta + x_i^T \beta_\theta + \varepsilon_i \quad și\ C\ est\ egal\ cu\ zero$.

unde T_i^* este o variabilă latentă de răspuns, $(\alpha_\theta, \beta_\theta)$ sunt intercepţia, respectiv parametrii necunoscuți ai regresiei cuantilice Tobit, $\theta \in (0,1)$. Pentru estimarea coeficienților în Tobit Q Reg se minimizează următoarea funcție de pierdere:

$$\min_{\alpha_\theta, \beta_\theta} = \sum_{i=1}^{n} \rho_\theta \left(y_i - max\{0, T_i^*\}\right) \quad [11]$$

unde $\rho_\theta(\varepsilon)$ se numește funcția de verificare a lui (Koenker și Bassett (1978)) [38] la cuantila θ, ecuația (11), de asemenea, nu este diferenţiabilă în 0. Dar minimizarea ecuației [11] poate fi rezolvată printr-un algoritm de programare liniară (Koenker și D'Orey, (1987)) [46] pentru a ne da o estimare a coeficienților modelului Tobit Q Reg. Deși sunt studiate proprietățile asimptotice pentru Tobit Q Reg și sunt propuși numeroși algoritmi, majoritatea acestor algoritmi sunt ineficienți atunci când variabila de răspuns are prea multe date cenzurate. În prezent, există o posibilă estimare a coeficienților modelului Tobit Q Reg de către funcția (crq) în pachetul (quanTobit Req) (Koenker, (2011)) [45]. Recent, (Yu și Stander, (2007)) [] au propus o abordare Bayesiană a estimării coeficienților în Tobit Q Reg, chiar și atunci când există multe date cenzurate. În ultimii ani, selectarea unui subgrup important al variabilelor explicative a captat o atenție deosebită în estimarea coeficienților din literatura de specialitate pentru estimarea coeficientilor in modelul Tobit Q Reg Lasso cu penalizare, după cum urmează:

$$\min_{\alpha_\theta, \beta_\theta} = \sum_{i=1}^{n} \rho_\theta \left(y_i - max\{0, \alpha_\theta + x_i^T \beta_\theta + \varepsilon_i\}\right) + \lambda \parallel \beta_\theta \parallel \quad [12]$$

unde (α_θ, β_θ) sunt parametrul intercepţie, respectiv vectorul parametrilor necunoscuţi și $\lambda \parallel \beta_\theta \parallel$ este penalizarea pentru estimarea și selectarea coeficienților cuantilici Tobit. De asemenea, minimizarea ecuației [12] poate fi rezolvată printr-un algoritm de programare liniară sau prin modelul Bayesian penalizat în Tobit Q Reg pentru a obține estimarea coeficienților și selectarea variabilelor în modelele Tobit Q Reg. Modelul Bayesian adoptiv Lasso în regresia cuantilică Tobit a fost dezvoltat de (Alhamzawi (2013)) [4], care în (2014) a propus și rețeaua elastică Bayesian adaptivă Tobit Q Reg.

Contribuția noastră constă în faptul că am propus o mouă metodă Bayesiană Lasso Tobit Q Reg pentru a realiza o selecție a variabilelor și o estimare a coeficienților în modelul de regresie Tobit cuantilică folosind abordarea Bayesiană. Cele mai multe metode din domeniul Bayesian Tobit Q Reg penalizat au asignat o mixtură de distribuții normale (SMN) înaintea de a realiza Lasso Bayesian în modele de regresie. (Mallick and Yi, (2014)) [55] a furnizat o tehnică nouă pentru realizarea Lasso-ului Bayesian în modelul de regresie tradițională prin distribuția SMU a priori, în locul funcției de densitate Laplace. Am dezvoltat un nou Lasso Bayesian în modelul Tobit Q Reg prin utilizarea distribuției (SMU) a priori pentru coeficienții modelului Tobit Q Reg. Metoda propusă a generat noi distribuții aposteriori condiționate, care sunt foarte importante pentru construirea algoritmului eficient MCMC. Pentru această metodă propusă, am fost inspirați de sugestia lui (Konker și Machado (1999)) [48] și (Yu and Moyeed (2001)) []. Acești cercetători au observat convergența dintre funcția de pierdere (10) și distribuția asimetrică Laplace (SLD). Prin urmare, termenul eroare aleatoare ε_i este distribuit ca distribuția asimetrică Laplace cu funcția densității de probabilitate (pdf) luând următoarea formă:

$$f(\varepsilon_i|\mu,\sigma,\theta) = \frac{\theta(1-\theta)}{\sigma}\ \exp{-\rho_\theta}\left\{\left(\frac{\varepsilon_i-\mu}{\sigma}\right)\right\} \qquad [13]$$

Când media este egală cu 0 și varianța este egală cu 1, atunci funcția densitate de probabilitate care se referă la eroarea aleatoare ε_i este după cum urmează:

$$f(\varepsilon_i|\sigma,\theta) = \theta(1-\theta)\ \exp{-\rho_\theta}\{(\varepsilon_i)\} \qquad [14]$$

$\rho_\theta(.)$ este funcția de verificare (pierdere). Distribuția comună (joint) a variabilei de răspuns $y = (y_1, \dots, y_n)^T$, dându-se $X = (x_1, \dots, x_n)^T$, este:

$$f(y|X,\alpha,\beta,\sigma,\theta) = \theta^n(1-\theta)^n exp\left\{-\sum_{i=1}^{n}\rho\theta(y_i - max\{0,\alpha_\theta + x_i^T\beta_\theta + \varepsilon_i\}\right\} \qquad [15]$$

Maximizarea funcției de probabilitate a ecuației [15] este echivalentă cu minimizarea ecuației [11]. Prin folosim directă a distribuției asimetrice Laplace (ALD), se ajunge la calcule dificile. Prin urmare, (Kozumi și Kobayashi, (2011)) [49] presupun că ALD poate fi reformulată într-o mixtură scalată de distribuții normale. Funcția de probabilitate a ecuației [14] devine după cum urmează:

$$f(T_i^*|\alpha_\theta, x_i^T, \theta, \beta_\theta, m_i) = \left[\frac{1}{\sqrt{4\pi m_i}}\right]^n e^{-\sum_1^n \frac{\left(T_i^*-\alpha_\theta-x_i^T\beta_\theta-(1-2\theta)m_i\right)^2}{4m_i}} \qquad [16]$$

unde m_i este funcția densitate a distribuţiei experimentale cu parametrul θ (1-θ). Ecuația [16] este o parte importantă pentru construirea eşantionelor noastre Gibbs. (Tibshirani, (2011)) [65] introduce distribuția a priori Laplace asignată abordării de tip Lasso Bayesian. Utilizarea directă a distribuției a priori Laplace conduce la calculul dificil al distribuțiilor aposteriori condiționate complete. Prin urmare, folosirea formulelor simplificate pentru distribuţia a priori Laplace conduce la un algoritm MCMC simplu și eficient. Există o formulă SMN alternativă propusă de (Andrews and Mallows, (1974)) [5], după cum urmează:

$$\frac{\lambda_j}{2}\ e^{-\lambda_j|\beta_j|} = \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} e^{\left(-\frac{\beta_j^2}{2s_j}\right)}\ \frac{\lambda_j^2}{2} e^{\left(-\frac{s_j\lambda_j^2}{2}\right)} ds_j \qquad [17]$$

Ecuația (17) este o formulă simplă a distribuției a priori Laplace cu două funcții. Prima funcție poate fi atribuită distribuţiei a priori a parametrilor (β_j), care au o distribuție normală cu medie zero și varianță (s_j) după cum urmează:

$$p(\beta_j|s_j) = \frac{1}{\sqrt{2\pi s_j}} exp\left\{-\frac{\beta_j^2}{2s_j}\right\} \qquad [18]$$

unde varianța necunoscută a lui β_j este s_j . Distribuția a priori exponențială pentru s_j ia forma următoare:

$$p(s_j|\lambda_j) \propto \frac{\lambda_j}{2}\ exp\left\{-\frac{s_j\lambda_j}{2}\right\} \qquad [19]$$

În final, mixtura scalată de distribuții a priori normale este considerată o alternativă bună pentru distribuția a priori Laplace. Acest lucru ne oferă un calcul simplu pentru distribuțiile aposteriori condiționate complete. Acesta este motivul pentru care majoritatea cercetătorilor au utilizat distribuția SMN în mod prioritar în modele de regresie Bayesiană penalizată. De exemplu, (Park și Casella, (2008)) [60] au prezentat modelul Bayesian Lasso în modelul tradițional de regresie. Aceste metode au fost extinse la regresia cuantilică Tobit. De exemplu, (Alhamzawi, (2013)) [4] a propus un Lasso adaptiv în regresia cuantilică Tobit folosind tehnica Bayesiană. De asemenea, (Alhamzawi și Yu, (2014)) [2], a sugerat o tehnică bayesiană pentru estimarea coeficienților în modelul Tobit Q Reg, folosind distribuția g-prior cu parametru ridge. De asemenea, (Alhamzawi, (2014)) [2] a propus o penalizare elastică de reţea Bayesiană în Tobit Q Reg. În metoda pe care am propus-o am folosit distribuția a priori SMU ca formulă alternativă despre distribuția a priori Laplace pentru coeficienții din modelul nostru:

$$\frac{\lambda}{2} e^{\{-\lambda|\beta_j|\}} = \int_{s_j>|\beta_j|}^{\infty} \frac{1}{2u_j} \frac{\lambda^2}{\Gamma(2)} \; u_j^{\,2-1} \, exp\{-\lambda u_j\} \, du_j \quad ,$$

$$= \int_{|\beta_j|}^{\infty} \frac{1}{2u_j} \frac{\lambda^2}{\Gamma(2)} \; u_j^{\,2-1} \, exp\{-\lambda u_j\} \, du_j \qquad [20]$$

Ecuația (20) are două funcții. Prima funcție este atribuită distribuției uniforme a priori pentru u_j, iar a doua funcție este atribuită distribuției a priori Gamma cu parametrul de formă (2) și parametrul de scalare (λ). Parametrul scalei (λ) are distribuție a priori Gamma cu parametrii (a, b). Acest parametru este necesar pentru reducerea coeficienților aproape de zero. Distribuția a priori a lui α_θ este atribuită distribuției uniforme a priori standard. Parametrii a, b, sunt hiper-parametrii fixaţi care iau valori inițiale.

Prin urmare, abordarea noastră Bayesiană ierarhică cu SMU pentru parametrii modelului nostru va fi după cum urmează:

$$y_{i=max\{0,T_i^*\}}, \qquad i=1,\dots,n,,$$

$$T_i^*|\alpha_\theta, \beta_\theta, z_i \sim N(\alpha_\theta + x_i^T\beta_\theta + (1-2\theta)m_i, 2m_i),$$

$$p(\alpha_\theta) \propto 1,$$

$$m_i \sim Exp\left(\theta(1-\theta)\right), \qquad [21]$$

$$\beta_j|u_j \sim Uniform\left(-u_j, u_j\right),$$

$$u_j|\lambda \sim u_j^{\,2-1}\exp(-\lambda u_j),$$

$$\lambda \sim Gamma(a, b),$$

Ierarhia Bayesiană din ecuația (21) este o parte importantă pentru generarea eșantionului Gibbs. Prima parte este dedicată funcției de probabilitate ca în ecuația (21) iar a doua parte va fi prezentată în ierarhia Bayesiană din ecuația (21), unde vom obține distribuțiile noastre aposteriori condiționate care sunt extrase din formula matematică după cum urmează:

$$posterior \;\; distribution \propto likelihood\, function \times prior\, distribution$$

Sub ierarhia modelelor Bayesiene (21) și (16), algoritmul de eșantionare Gibbs este utilizat pentru a estima și a actualiza parametrii. Distribuțiile aposteriori condiționate complete pentru metoda propusă vor fi următoarele:

Distribuția condiționată aposteriori a variabilei (T_i^*) urmează distribuția normală trunchiată care este dată de:

$$T_i^*|y_i, x_i, m_i, \alpha_\theta, \beta_\theta \sim \begin{cases} \gamma(y_i) & if\ y_i > o \\ N(\alpha_\theta + x_i^T\beta_\theta + (1-2\theta)m_i, 2m_i)I(T_i^* \leq 0) & \text{in caz contrar} \end{cases} \quad [22]$$

unde $\gamma(y_i)$ este distribuția degenerată. Distribuția completă condiționată aposteriori a lui α_θ este distribuția normală cu medie egală cu $(\sum_{i=1}^{n}\frac{(T_i^*-x_i^T\beta_\theta-(1-2\theta)m_i)}{2m_i})$ și varianța egală cu $(\sum_{i=1}^{n}\left(\frac{1}{2m_i}\right))$. De asemenea, distribuția aposteriori condiționată completă a lui m_i este Gaussiană inversă cu media $\sqrt{\frac{1}{(T_i^*-\alpha_\theta-x_i^T\beta_\theta)^2}}$ și parametrul formei $(\frac{1}{2})$, iar distribuția aposteriori condiționată completă a β_j este normală trunchiată cu media $\acute{\sigma}^2\sum_1^n\frac{x_{ij}\left(T_i^*-(1-2\theta)m_i-\sum_{j=1,j\neq k}^{p}x_{ij}\beta_j\right)}{2m_i}$ I$\left[|\beta_j|<u_j\right]$ și varianța $\acute{\sigma}^2=\left(\sum_1^n\frac{x_{ij}{}^2}{2z_i}\right)^{-1}$. Distribuția aposteriori condiționată completă a lui u_j este exponențială trunchiată la stânga cu parametrul ratei (λ). Distribuția completă condiționată aposteriori a lui λ este o distribuție Gamma cu parametru ratei $(a+2p)$ și parametrul de scalare $\left(b+\sum_{j=1}^{p}|\beta_j|\right)$, unde a, b sunt hiperparametri care iau valori inițiale. Distribuțiile noastre aposteriori ierarhice bayesiene vor genera un algoritm MCMC atrăgător pentru metoda pe care am propus-o, new Bayesian Lasso Tobit quantile regression (New B L Tobit Q Reg) "Fadel Hamid Hadi Alhusseini", 2017) [25].

Am testat performanța metodei propuse prin compararea cu alte două metode. Acestea sunt metoda clasică Tobit Q Reg (crq) prin utilizarea funcției crq () care folosește abordarea lui Powell în (Koenker (2013)) [50] și rețeaua elastică Bayesiană adaptivă Tobit Q Reg (BAnet), raportată de (Alhamzawi (2014))[2]. Pentru testarea comparativă a acestor metode, vom folosi abordarea prin simulării precum și cea bazată pe date reale. În abordările prin simulare, vom folosi două criterii. Primul este rădăcina reziduului mediu pătratic (RMSR: Root Mean Square Residual), RMSR$(\beta)=\sqrt{\frac{1}{S}\sum_{i=1}^{S}(\hat{\beta}_{kj}-\beta_j{}^{True})^2}$, $j=1,2,\ldots\ldots p$. unde S este numărul de simulări, $\hat{\beta}_{kj}$ sunt parametrii estimați pentru cel de-al $j-lea$ parametru al modelului în a $k-a$ iterație, iar $\beta_j{}^{True}$ sunt parametrii adevărați (Lawrence și Arthur (1990)) [54]. Al doilea criteriu folosit este abaterea standard a MADS. Performanța metodei propuse este investigată prin abordări bazate pe simulare. Modelul folosit în aceste simulări este definit după cum urmează:

$$y = \max(C, T_i^*)\ , \qquad i = 1,2,\ldots\ldots\ldots n\ , n = 100,$$

$$unde\ \ T_i^* = \alpha_\theta + x_i^T\beta_\theta + \varepsilon_i,\ \ și\ C = zero,\ .$$

θ este nivelul cuantilic Tobit. În simulările noastre, am folosit trei niveluri cuantilice Tobit: nivel cuantilic scăzut $\theta_1 = 0.30$, nivel cuantilic mediu $\theta_2 = 0.60$ și nivel cuantilicic mare $\theta_3 = 0.90$. Termenul eroare ε_i , $i = 1,2\ldots\ldots n$, este generat din trei distribuții diferite: o distribuție normală standard, o distribuție $\chi^2_{(4)}$ cu patru grade de libertate și o distribuție Laplace standard $u_i \sim Laplace$ (0,1). Numărul de simulări a fost de 100 pentru fiecare caz. Algoritmul nostru a fost rulat în 13000 de iterații. Primele 3000 au fost eliminate. Pentru evaluarea performanței metodei propuse, aceasta a fost comparată cu alte două metode prin intermediul a patru abordări prin simulare. In prima simulare, datele noastre de simulare au fost generate din cazuri foarte rare și cu adăugarea termenului de intercepţie la parametrii adevărați $\beta = (0{,}5{,}0{,}0{,}0{,}0{,}0{,}0{,}1)^T$. Modelul real va fi următorul:

$y = \max(0, T_i^*)$, $T_i^* = 0 + x_{1i} + x_{8i} + \varepsilon_i,$ $i = 1,2, \dots \dots .100$

Am simulat variabilele explicative $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i})$ dintr-o distribuție Gaussiană multivariată $X \sim N_8(\mu, \Sigma)$, unde μ este vectorul mediilor $\mu \epsilon R^n$ și Σ este matricea de covariație $(\Sigma_x)_{ij} = (2^{-1})^{|i-j|}$. Metoda pe care am propus-o (New B L Tobit Q Reg) are cel mai mic RMSR în comparație cu BAnet și crq la diferite distribuții ale erorii și la niveluri cuantilice Tobit diferite. Metoda propusă are, de asemenea, o abatere standard (SD) mai mică pentru diferite distribuții ale erorii și niveluri diferite cuantilice, comparativ cu celelalte două metode. Acest lucru înseamnă că propunerea noastră (New B L Tobit Q Reg) are o performanță mai bună și o precizie mai ridicată în estimarea coeficienților și selecția variabilelor în modelul Tobit Q Reg comparativ cu celelalte metode.

In a doua abordare prin simulare, datele noastre de simulare sunt generate din cazul dens și cu adăugarea termenului de intercepţie la parametrii adevărați $\beta = (0{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85)^T$. Prin urmare, modelul real este dat după cum urmează:

$y = \max(0, T_i^*)$, $T_i^* = 0 + 0.85x_{1i} + 0.85x_{1i} + 0.85x_{2i} + 0.85x_{3i} + 0.85x_{4i} + 0.85x_{5i} + 0.85x_{6i} + 0.85x_{7i} + 0.85x_{7i} + \varepsilon_i,$ $i = 1,2, \dots \dots .100$

Variabilele explicative $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i})$ sunt simulate în funcție de distribuția Gaussiană multivariată $X \sim N_8(\mu, \Sigma)$, unde μ este vectorul mediilor $\mu \epsilon R^n$ și Σ este matricea de covarianţă $(\Sigma_x)_{ij} = (2^{-1})^{|i-j|}$. În a doua simulare putem vedea că metoda propusă (New B L Tobit Q Reg) are performanțe mai bune decât metodele Bayesiene și non-Bayesiene (BAnet și crq) întrucât RMSR generată de metoda propusă (New B L Tobit Q Reg) este foarte mică comparativ cu celelalte două metode la diferite distribuții ale erorii. În simulaea a treia datele noastre sunt simulate din structurile de grup, inclusiv termenul de intercepție și $\beta = (0, (0{,}0{,}0), (2{,}2{,}2), (0{,}0{,}0), (2{,}2{,}2), (0{,}0{,}0))^T$. Următorul model adevărat a fost utilizat.

$y = \max(0, T_i^*)$, $T_i^* = 0 + 2x_{4i} + 2x_{5i} + 2x_{6i} + 2x_{10i} + 2x_{11i} + 2x_{12i} + \varepsilon_i,$ $i = 1,2, \dots \dots .100$

Variabilele explicative $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i,} x_{9i,} x_{10i,} x_{11i,} x_{12i,} x_{13i,} x_{14i,} x_{15i})$ sunt simulate dintr-o distribuție Gaussiană multivariată $X \sim N_{15}(\mu, \Sigma)$,unde μ este vectorul mediilor $\mu \epsilon R^n$ și Σ este matricea de covarianţă $(\Sigma_x)_{ij} = (2^{-1})^{|i-j|}$.

Din rezultatele prezentate în a treia simulare putem vedea că performanța metodei propuse (New B L Tobit Q Reg) este mai bună decât metodele Bayesiene și non-Bayesiene (BAnet respectiv crq). Acest lucru este clar din RMSR rezultat, unde RMSR calculat prin metoda propusă este mai mică decât RMSR calculată prin celelalte două metode (BAnet și crq) pentru toate nivelele cuantilice și toate distribuțiile diferite ale erorii.

A patra simulare:

Datele noastre de simulare sunt generate din structurile multi-grup, inclusiv termenul de intercepţie:

$\beta = (0, (0{,}0{,}0), (2{,}2{,}2), (0{,}0{,}0), (2{,}2{,}2), (0{,}0{,}0), (0{,}0{,}0), (2{,}2{,}2), (0{,}0{,}0), (2{,}2{,}2), (0{,}0{,}0))^T.$

Prin urmare, adevăratul model va fi după cum urmează:

$y = \max(0, T_i^*)$, $T_i^* = 0 + 2x_{4i} + 2x_{5i} + 2x_{6i} + 2x_{10i} + 2x_{11i} + 2x_{12i} + 2x_{19} + 2x_{20i} + 2x_{21i} + 2x_{25} + 2x_{26i} + 2x_{27i} + \varepsilon_i,$ $i = 1,2, \dots \dots .100$

unde variabilele explicative:

$(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i}, x_{9i}, x_{10i}, x_{11i}, x_{12i}, x_{13i}, x_{14i}, x_{15i}, x_{16i}, x_{17i}, x_{18i}, x_{19i}, x_{20i}$

$, x_{21i}, x_{22i}, x_{23i}, x_{24i}, x_{25i}, x_{26i}, x_{27i}, x_{28i}, x_{29i}, x_{30i})$ sunt simulate dintr-o distribuție Gaussiană multivariată $X \sim N_{30}(\mu, \Sigma)$, unde μ este vectorul mediilor $\mu \epsilon R^n$ și Σ este matricea de covarianţă cu $(\Sigma_x)_{ij} = (2^{-1})^{|i-j|}$. Din rezultatele obţinute din a patra simulare putem vedea că metoda noastră propusă este mai bună decât metodele Bayesiane și non-Bayesiane (BAnet și crq, respectiv) în ceea ce privește criteriile RMSR și SD. RMSR și SD generate de (New B L Tobit Q Reg) sunt mult mai mici comparativ cu RMSR și SD generate de celelalte două metode. De asemenea, săa evaluăm propunerea noastră (New B L Tobit Q Reg) comparativ cu celelalte două metode prin estimarea coeficienților prin criteriul căii directe. Din rezultatele care aparțin estimării criteriului căii directe, coeficienții estimaţi prin metoda propusă (New B L Tobit Q Reg) a fost foarte apropiată de parametrii adevăraţi comparativ cu metodele Bayesiene și non-Bayesiene (BAnet, respectiv crq). Acest lucru indică faptul că metoda (New B L Tobit Q Reg) are o performanță destul de bună în comparație cu celelalte două metode. De asemenea, pentru a demonstra performanța metodei propuse (New B L Tobit Q Reg), a fost utilizată abordarea bazată pe date reale. Datele aventurilor extramaritale sunt folosite pentru testarea metodei propuse (New B L Tobit Q Reg). Aceste date sunt disponibile în pachetul R "AER" și au fost prezentate de către (Fair in (1978)) [21]. Aceste date sunt utilizate de către Ji et al., (2012) [37], (Alhamzawi, (2014)) [2] și alții și conțin variabila de răspuns cenzurată reprezentând numărul de întâlniri sexuale extramaritale care au avut loc în trecutul an (aventuri). Cele opt variabile explicative sunt sexul (1 pentru femeie și 2 pentru bărbați), vârsta, numărul de ani de căsătorie, copiii (2 pentru existența copiilor în căsătorie și 1 fără copii), religiozitatea (scara de la 1 la 5), nivelul de educație, cât de mult își prestează ocupația (scară de la 1 la 7) și o evaluarea fericirii în căsătoria lor (scară de la 1 la 5). Dimensiunea eșantionului de date privind aventurile extraconjugale este de 601 de observații. Variabila de răspuns (affairs) are valorile mari cenzurate, unde sunt cenzurate 451 observații și restul observațiilor necenzurate (Fadel Hamid Hadi Alhusseini, 2017) [25]. (MSE) este calculată prin metoda propusă și este mai mică decât cel al metodei Banet, pentru toate nivelele cuantilice Tobit. Aceste rezultate demonstrează că metoda noastră propusă poate fi considerată mai bună decât metoda Banet.

Din toate rezultatele care sunt enumerate în abordările de simulare și datele reale, vom vedea că metoda propusă (New B L Tobit Q Reg) are o performanță destul de bună în estimarea coeficienților și selecția variabilelor în modelul Tobit Q Reg comparativ cu alte metode. De asemenea, este considerată o nouă extensie la modelul Bayesian penalizat Tobit Q Reg.

Pentru implementarea selecției variabilelor prin metoda propusă (NewBL Tobit Q Reg) vom construi un nou algoritm MCMC pentru determinarea valorii de probabilitate pentru fiecare variabilă independentă din modelul nostru. Coeficienții estimați prin abordarea Bayesian sunt implementați prin mii de iterații. La fiecare iterație se va genera un nou estimator în conformitate cu algoritmul propus (Gramacy, R. B. și Lee, H. K. H. (2008)) [32]. Toate aceste estimări sunt comparate cu intervalul (-0,05, 0,05). Dacă coeficienții estimați care aparțin variabilelor independente în afara intervalului deschis (-0,05, 0,05), la valoarea de probabilitate este mai mare de 0,5. Aceasta înseamnă că variabila independentă are o mare importanță relativă în model. Dimpotrivă, dacă coeficienții estimați care aparțin variabilelor independente din afara intervalului deschis (-0,05, 0,05), la valoarea de probabilitate sunt mai mici de 0,5. Aceasta înseamnă că variabila independentă are o importanță relativă slabă în model (Reed, C, (2011)) [62] și (Fadel Hamid Hadi Alhusseini 2017) [27] [28]. Metoda noastră este pusă în aplicare cu privire la datele privind aaventurile Extramaritale la patru nivele cuantilice Tobit. În modelul Tobit Q Reg la nivelul cuantilic Tobit ($\theta_1 = 0.15, \theta_2 = 0.35, \theta_3 = 0.75$ *și* $\theta_4 = 0.95$). În modelul Tobit Q Reg la nivelul cuantilic Tobit $\theta_1 = 0.15$.

Există patru variabile independente (x_1, x_2, x_5, x_8) care au valori de probabilitate mai mari decât 0 și există patru variabile independente (x_3, x_4, x_6, x_7) care au valori de probabilitate mai mici de 0,5. Aceasta înseamnă variabilele independente (x_1, x_2, x_5, x_8) au o mare importanță relativă și variabilele independente (x_3, x_4, x_6, x_7) au o importanță relativă scăzută în Tobit Q Reg la nivelul cuantilic Tobit $\theta_1 = 0.15$.

La nivelul cuantilic $\theta_2 = 0.35$, vedem că variabilele independente $(x_1, x_2, x_5, x_6, x_7, x_8)$ au o mare importanță relativă în modelul nostru din cauza valorii lor de probabilitate depășind 0,5. Dar variabilele independente (x_3, x_4) au o importanță relativă scăzută în modelul regresiei cuantice Tobit la nivel $\theta_2 = 0.35$. Din cauza valorilor lor de probabilitate, nu trebuie să depășească 0,5. La nivelul cuantilic Tobit θ_3=0.75 and nivelul cuantilic Tobit θ_4=0.95 all independent variables have a high relative importance in our models (Tobit Q Reg) . Din cauza valorii lor de probabilitate mai mare de 0,5. Mai multe detalii despre metoda propusă (noul BL Tobit Q Reg) sunt prezentate în capitolul trei,
Avem, de asemenea, o contribuție nouă la estimarea coeficienților și selecția variabilelor în modelul compozit Tobit Q Reg prin intermediul metodei propuse Bayesian composite Tobit regression (regresia compozită Bayesiană Tobit Q Reg). Abordările menționate mai sus pentru modelarea lui Tobit Q Reg se concentrează pe un singur nivel cuantilic. Cu toate acestea, eficiența estimatorilor Tobit Q Reg depinde de nivelul cuantilic Tobit. Deoarece distribuția este necunoscută, este dificil să selectăm cea mai informativă cuantilă Tobit, care poate oferi un estimator eficient. (Zou și Yuan (2008)) [76] au fost propuse o nouă metodă de estimare a coeficienților în modelul de regresie numită regresie cuanelă compusă (Composite Q Reg) și arată eficiența relativă a acestor estimatori este mai mare de 70% estimator, indiferent de distribuția de erori. Estimatoarele regresie cuantifică compozită (compozit Q Reg) sunt robuste pentru variabilele dependente de greutăți sau valori extreme și sunt mai eficiente decât o singură regresie cuanle. Pentru aceste caracteristici folosim această abordare în acest studiu. Modelul compozit Tobit Q Reg va fi după cum urmează:

$$T_i^* = \alpha_h + x_i^T \beta_h + \varepsilon_i, \qquad , \qquad y = \max(C, T_i^*) \qquad [22]$$

Unde $h = 1, \ldots\ldots, H,.$, H și diferite cuantile, $0 < \theta_1 < \theta_2 < \cdots < \theta_H < 1$ and $i = 1, \ldots\ldots, n$.
În compozitul Tobit Q Reg parametrii sunt estimați prin rezolvarea următoarei ecuații

$$\left(\hat{\alpha}1, \hat{\alpha}2, \ldots, \hat{\alpha}\mathrm{H}, \hat{\beta}\right) = \underset{\alpha_1 \ldots\ldots \alpha_{2,\beta}}{Min} \quad \sum_{h=1}^{H} \left\{ \sum_{i=1}^{n} \rho_{\theta_h} \left(y_i - \max(c, \alpha_h + x_i^T \beta \right) \right\}, \qquad [23]$$

Ecuația (23) nu poate fi diferențiată la 0 puncte. Prin urmare, minimizarea poate fi realizată prin unele modificări ale algoritmului propus de (Koenker și D'Orey (1987)) [46]. Pentru a estima coeficienții Composite Tobit Q Reg, o metodă Bayesiană este considerată o nouă abordare în acest scop. Prin intermediul unui nou eșantionator Gibbs este propus pentru inferența de distribuție aposteriori. Din cele cunoscute, repartizarea aleatorie a erorilor de Tobit Q Reg aparține (ALD). Prin urmare, distribuția în comun a variabilei de răspuns $[y^T_{(i=1,2,\ldots\ldots,n)}]$ dat fiind $[x^T_{(i=1,2,\ldots\ldots,n)}]$, $\left[\alpha^T_{(j=1,2,\ldots.h)}\right]$$și$ $[\beta^T_{(j=1,2,\ldots.h)}]$ pentru compozitul Tobit Q Reg este

$$f(y|X, \alpha, \beta) = \prod_{h=1}^{H} \theta_h^n (1 - \theta_h)^n exp \left\{ -\sum_{i=1}^{n} \rho_{\theta_h} (y_i - max\{c, \alpha_h + x_i^T \beta\} \right\} \qquad [24]$$

Ecuația (24) este dificil de rezolvat direct din cauza amestecului de H components. În urma (Huang și Chen (2015)) [34], folosim o matrice de atribuire a clusterului C al cărui $(i, h)^{lea}$element C_{ih} este egal cu 1 dacă i_{it} aparține clusterului h_{th}, altfel $C_{ih} = 0$. Elementul

C_{ih} este tratat ca valoare lipsă. Astfel, probabilitatea noastră are forma "Fadel Hamid Hadi Alhusseini și Vasile Georgescu, 2017" [23].

$$f(y|X,\alpha,\beta) = \prod_{h=1}^{H}\prod_{i=1}^{n}\left[\theta_h(1-\theta_h)exp\left\{\rho_{\theta_h}(y_i - max\{c,\alpha_h + x_i^T\beta\})\right\}\right]^{C_{ih}} \qquad [25]$$

De asemenea, atunci când se utilizează ecuația (25) conduce direct la algoritmul MCMC dificil. Acum vom folosi propunerile Kozumi și Kobayashi prin reformularea ALD la un amestec de distribuții normale. Pentru compozitul nostru ierarhic Bayesian Tobit Q Reg, formularea (Kozumi și Kobayashi (2011)) [49] poate fi scrisă ca:

$y_i|\alpha_h,\beta,z_i \sim Normal\ (max\{c,\alpha_h + x_i^T\beta\} + (1-2\theta_h)m_i, 2m_i)$, $unde\ m_i \sim Exp\big(\theta_h(1-\theta_h)\big)$. Sub această formulă, funcția de probabilitate a variabilei de răspuns este dată după cum urmează

$$f(T_i^*|\alpha_h,\beta,z_i) = \prod_{i=1}^{n}\frac{1}{\sqrt{4\pi m_i}}\, e^{\frac{1}{2}\sum_{i=1}^{n}\sum_{h=1}^{H}\frac{c_{ih}\,(T_i^*-\alpha_h - x_i^T\beta-(1-2\theta_h)m_i)^2}{2z_i}} \qquad [26]$$

Parametrii Compozit Tobit Q Reg au caracteristici de conjugare condiționate de construcție pentru un algoritm simplu și atractiv de sortare Gibbs pentru montarea modelului nostru pe date. Ecuația (26) ia în considerare funcția de probabilitate pentru estimarea coeficientului și selecția variabilă a metodei propuse (model compozit Tobit Q Reg). Prin urmare, avem nevoie de distribuții ierarhice a priori pentru a obține distribuții aposteriori complete condiționate ale metodei propuse. Distribuțiile ierarhice a priori ale regresiei compuse Tobit cuantilice vor fi după cum urmează:

Distribuțiile a priori la parametrii modelului (compozit Tobit Q Reg) sunt rezumate după cum urmează: Distribuția uniformă este atribuită parametruluiα_h , unde $p(\alpha_h) \propto 1$ și distribuția a priori normală cu media 0 și varianța s_j aparține parametrului β_j . Parametrul s_j are distribuția a priori exponențială cu parametrul ratei (λ_j) . Parametrul λ_j este repartizat ca distribuție a priori gamma cu parametrul de formă (a) și parametrul de scală (b). A și b sunt doi parametri hiper care iau valorile inițiale, unde a=0.1 și b= 0.1. Prin urmare, modelul nostru ierarhic Bayesian va fi următorul:

$y_{i=max\{c,T_i^*\}},\qquad i=1,\ldots,n,$

$T_i^*|\alpha_h,\beta,z_i \sim [N(\alpha_h + x_i^T\beta + (1-2\theta_h)m_i, 2m_i)]^{c_{ih}},$

$p(\alpha_h) \propto 1$ [27]

$m_i \sim Exp\ (\theta_h(1-\theta_h))$,

$\beta_{j\sim N(s_j)}$

$s_{j\sim}Exp\left(\frac{\lambda_j}{2}\right),$

$\lambda_j \propto \lambda_j^{a-1}exp\{-b\lambda_j\}.$

Din ecuația (26) și din ecuația (27), vom obține distribuțiile aposteriori condiționate. Prin urmare, modelul ierarhic al modelului compozit Tobit Q Reg cu penalizare Lasso sunt:

Distribuția condiționată completă aposteriori a variabilei latente (T_i^*) dată de

$$T_i^*|y_i, m_i, \alpha_h, \beta \sim \begin{cases} \{\Upsilon(y_i), & if\ y_i > c; \\ \left\{\prod_{h=1}^{H} [N(\alpha_h + x_i^T\beta + (1-2\theta_h)m_i, 2m_i)]^{c_{ih}}\right\} I(T_i^* \leq c), otherwise \end{cases} \quad [28]$$

unde c este egal cu zero și $\Upsilon(y_i)$ denotată la o distribuție degenerată. Distribuția completă condiționată aposteriori a m_i for $i = 1, \ldots, n,$ este Gaussian invers cu medie$\sqrt{\sum_{h=1}^{H} C_{ih}/(T_i - \alpha_h - x_i^T\beta)^2)}$ și parametrul de formă $\sum_{h=1}^{H} C_{ih}/2$. Distribuția completă condiționată aposteriori a α_h este distribuția normală cu media $(\tilde{\sigma}_h^2 \sum_{i=1}^{n} C_{ih}\,(T_i^* - x_i^T\beta - (1-2\theta_h)m_i)/2m_i)$ și varianța $(\sum_{i=1}^{n}(C_{ih}/2m_i))^{-1}$.

Distribuția completă condiționată aposteriori a β_j[j=1, 2,….k] este distribuția normală cu media $(\tilde{\sigma}_j^2 \sum_{h=1}^{H}\sum_{i=1}^{n} C_{ih}x_{ij}(y_i - \alpha_h - \sum_{l\neq j} x_{il}\beta_l - (1-2\theta_h)m_i)/(2m_i))$ și varianța $(\tilde{\sigma}_j^2 = (\sum_{i=1}^{n}\sum_{h=1}^{H} C_{ih}\, x_{ij}^2/2m_i) + s_j^{-1})^{-1})$. Distribuția completă condiționată aposteriori a $s_i^{-1}[j = 1,2, \ldots \ldots k]$este Gaussian invers cu media$\sqrt{\lambda_j/\beta_j^2}$ și parametrul de formă λ_j. Tdistribuția aposteriori completă condiționată a λ_j este distribuția gamma cu parametru de formă $(a+1)$ și parametrul de scală $(b + \frac{s_j}{2})$. Distribuția completă condiționată aposteriori a $C_{ih} = (C_{i1}, C_{i2}, \ldots \ldots, C_{iH})^T$ este o distribuție multinomială; $p(C_i|y, X, \alpha, \beta, z) \propto$ $\text{multinomial}(1, \hat{p}_1, \ldots, \hat{p}_H)$, unde $\hat{p}_H = \frac{\exp[-(y_i - \alpha_h - x_i^T\beta - (1-2\theta_h)m_i/2m_i]}{\sum_{h=1}^{H}\exp[-(y_i - \alpha_h - x_i^T\beta - (1-2\theta_h)m_i/2m_i]}$ Din distribuțiile aposteriori complete condiționate sunt arătate în formulele de mai sus, vom obține un algoritm simplu și eficient Gibbs sampler. Algoritmul nostru a fost rulat pentru 16.000 de iterații, iar primele 1000 au fost eliminate ca fiind arse. Apoi, credem că iterațiile ulterioare prin păstrarea fiecarei simulări a 5-a remiză și aruncând restul. Pentru evaluarea metodei propuse (Bayesian Composite Tobit Q Reg) sa folosit abordarea de simulare și date reale "Fadel Hamid Hadi Alhusseini și Vasile Georgescu, 2017" [23].

În studiul de simulare, vom utiliza trei abordări de simulare pentru evaluarea performanței metodei noastre (Bayesian Composite Tobit Q Reg) prin compararea acestei metode cu metoda bayesiană și non-Bayesiană "crq, BAnet". Metodele comparative sunt evaluate prin (MMAD) unde MMAD= $median(media(|x^T\hat{\beta} - x^T\beta|))$. De asemenea, sunt prezentate deviațiile standard (SD) ale MAD.

Simularea 1 aparține unui caz foarte restrâns. Astfel, parametrii adevărați ai modelului sunt $\beta = (1,0,0,0,0,0,0,0)^T$. Datele sunt generate de modelul real după cum urmează:

$$y_i = \max(0, T_i^*), \quad i = 1,2, \ldots \ldots .., 100 \ , \qquad T_i^* = x^T\beta + \varepsilon_i$$

Unde X este distribuția distribuției normale multivariate a $N_k(0, \Sigma_x)$ cu $(\Sigma_x)_{hl} = 0.5^{|h-l|}$, and(k=8) .$T_i^*$ este variabila latent cu media 0. Reziduurile sunt generate din 5 distribuții: $\varepsilon_i \sim N(0,1)$ (ε_i e standard distribuit normal cu zero și varianță 1) , a $\varepsilon_i \sim t_{(3)}$($\varepsilon_i$ este distribuită o distribuție t la 3 grade de libertate),a $\varepsilon_i \sim 0.5N(1,1) + 0.5N(-1,1)$($\varepsilon_i$ distribuția distribuției normale a amestecului) a $\varepsilon_i \sim$ Laplace (0,1),(ε_i este repartizată distribuția standard Laplace cu parametrul de localizare 0 și parametrul de scară 1) și $\varepsilon_i \sim 0.5\text{Laplace}(1,1) + 0.5\text{Laplace}(-1,1)$ (ε_i distribuția amestecului distribuție Laplace)

Am stability un H=3 astfel încât cele trei niveluri cuantilice Tobit sunt: $\theta_1 = 0.25, \theta_2 = 0.50, și\ \ \theta_3 = 0.75$ unde $\theta_h = h/(H+1)$, "Fadel Hamid Hadi Alhusseini and vasile Georgescu, 2017"[23].

Din rezultatele obținute am observat că MMAD și deviațiile standard (SD) calculate prin metoda propusă (Bayesian Composite Tobit Q Reg) sunt mult mai mici decât deviațiile standard (SD) și MMAD calculate de crq și BAnet pentru toate distribuțiile de erori. Aceasta înseamnă că metoda propusă (Bayesian Composite Tobit Q Reg) are o performanță mai bună decât celelalte două metode.

Cea de-a doua simulare aparține cazului dens. Prin urmare, parametrii adevăratți sunt $\beta = \underbrace{(0.85, \ldots \ldots \ldots, 0.85)}_{15}^{T}$. Datele sunt generate de modelul real după cum urmează:

$$y_i = \max(0, T_i^*), \quad i = 1,2, \ldots \ldots \ldots, 100 \ , \qquad T_i^* = x^T \beta + \varepsilon_i$$

De asemenea, covariatele (X), au distribuit distribuția normală multivariată a $N_k(0, \Sigma_x)$ cu $(\Sigma_x)_{hl} = 0.5^{|h-l|}$, ;i (k=15) The T_i^* este o variabilă latentă cu media 0. Reziduurile sunt generate pentru toate distribuțiile luate în considerare. În această simulare, rezultatele arată că metoda noastră propusă (Bayesian Composite Tobit Q Reg) are o performanță mai bună decât alte două metode la trei niveluri cuantilice Tobit și toate distribuțiile reziduale. Acest lucru se limitează la rezultatele obținute de SDS și MMAD generate de metodele în care (SD) și MMAD sunt generate de metoda propusă mai mică decât (SD) și MMAD sunt generate de alte două metode.Simularea 3 aparține Casei structurale de grup. Prin urmare, parametrii adevărați ai modelului sunt $\beta = ((1.5,1,0), (0,0,0), (1.8,1,0), (0,0,0), (1,1,0)^T$. Datele sunt generate de modelul real după cum urmează::

$$y_i = \max(0, T_i^*), \quad i = 1,2, \ldots \ldots \ldots, 100 \ , \qquad T_i^* = x^T \beta + \varepsilon_i$$

De asemenea covariatele (X), au distribuit multivariat distribuția normală $N_k(0, \Sigma_x)$ cu $(\Sigma_x)_{hl} = 0.5^{|h-l|}$, și (k=15). T_i^* este variabila latentă cu media 0. Reziduurile sunt generate pentru toate distribuțiile luate în considerare. Din rezultatele MMAD și SD sunt generate de metoda propusă (Bayesian Composite Tobit Q Reg) sunt cele mai mici decât SD și MMAD sunt generate de BAnet și crq pentru toate distribuțiile reziduale. Aceste rezultate indică faptul că performanța metodei propuse a fost bine comparată cu alte două metode (BAnet și crq). Din rezultatele înregistrate de trei simulări, concluzionăm că metoda propusă (Bayesian Composite Tobit Q Reg) are o mare importanță pentru selecția variabilă și estimarea coeficienților în modelul de regresie Tobit cuantilic în comparație cu alte metode din același domeniu. Pentru a evalua performanța metodei noastre (Bayesian Composite Tobit Q Reg) cu date reale, vom folosi datele de participare la forța de muncă disponibile în pachetul AER în R. Aceste date au fost introduse de către (Mroz (1987)) [56] folosind Tobit Q Reg în (Yu și Stander (2007)) []. Descrierea datelor privind participarea la forța de muncă este cuprinsă în variabila de răspuns y: (ore) este considerată orele de lucru ale soției în 1975 și șase variabile explicative sunt educația x_1: (*education*) (educația) soției în ani,, x_2: (experiență) ani reali din experiența anterioară a pieței forței de muncă a soției, x_3: (vârstă), x_4: (*taxe*) nivelul impozitării marginale față de soție, x_5: (copii mari) numărul de copii cu vârste cuprinse între șase și optsprezece ani în gospodării și familii x_6: (*fincome*) venitul familiei. dimensiunea eșantionului acestor date este *n* = 753 observații , din acesta 325 sunt observații cenzurate, iar celelalte (423) observații sunt observații necenzurate. Vom folosi patru nivele cuantilice Tobit $(H = 4, h = 1,2,3, H = 4))$. Se determină după cum urmează: $\theta_h = h/(H+1)$.

unde$\theta_1 = \frac{1}{5} = 0.20, \theta_2 = \frac{2}{5} = 0.40\ , \theta_3 = \frac{3}{5} = 0.60\ \ și\ \theta_4 = \frac{4}{5} = 0.80$.

Apoi $\theta \in \{0.20, 0.40, 0.60, 0.80\}$.

Pentru evaluarea metodei propuse (Bayesian Composite Tobit Q Reg) am comparat-o cu alte două metode (Banet, crq) s-a calculat eroarea medie pătrată (MSE) pentru toate metodele studiate. Prin toate nivelele cuantilice Tobit (MSE) calculată prin metoda propusă (Bayesian Composite Tobit Q Reg) mult mai mică decât (MSE), se calculează prin alte două metode. Prin urmare, metoda propusă (Bayesian Composite Tobit Q Reg) are o performanță mai bună comparativ cu metodele (Banet, crq). Din rezultatele prezentate în simulări și abordări de date reale, concluzionăm că metoda propusă (Bayesian Composite Tobit Q Reg) este destul de bună în estimarea coeficienților și selecția variabilelor în modelul Tobit Q Reg "Fadel Hamid Hadi Alhusseini și Vasile Georgescu, 2017" [23].

În metoda Bayesian Composite Tobit Q Reg, vom propune un nou algoritm MCMC pentru selecția variabilă prin determinarea importanței relative pentru variabilele explicative din modelul nostru. În cazul în care toate metodele anterioare din domeniul selecției variabile conțin algoritmii lor în următoarea condiție, dacă estimarea coeficienților este închisă de la zero, acestea sunt exact zero. În algoritmul MCMC vom propune o nouă procedură în selectarea variabilei prin determinarea relevanței relative pentru fiecare variabilă explicativă prin estimarea Bayesiană comparată pentru fiecare variabilă explicativă cu interval deschis (-0,05, 0,05), pentru calculul valorii de probabilitate pentru această variabilă explicativă. Dacă estimarea Bayesiană pentru această variabilă explicativă în afara intervalului (-0,05, 0,05) cu valoarea probabilității mai mare de 0,5, atunci această variabilă explicativă este activă în modelul nostru. Dimpotrivă, dacă estimarea Bayesiană pentru această variabilă explicativă în afara intervalului (-0,05, 0,05) cu o valoare de probabilitate mai mică de 0,5, atunci această variabilă explicativă este inactivă în modelul nostru. Prin urmare, îl putem șterge din modelul nostru. A se vedea (Reed, C (2011)) [62] și (Alhamzawi, R., (2016)) [5], (Fadel Hamid Hadi Alhusseini, 2017). Am folosit metoda propusă de Bayesian Composite Tobit Q Reg pentru a determina importanța relativă a variabilelor explicative ale datelor de participare la forța de muncă la două grupe de niveluri cuantilice Tobit (H = 5 și H = 10). Din rezultatele obținute la cinci niveluri quantle Tobit (H = 5) există două variabile explicative (x_3: vârsta, x_5: copii mari) au o mare importanță relativă în modelul nostru. Deoarece aceste variabile explicative au valori de probabilitate mai mari de 0,5, există patru variabile explicative (x_1: Educația, x_2: Experiența, x_4:Taxele și x_6:Venitul) cu o importanță relativă scăzută în modelul nostru (compozit Tobit Q Reg la cinci nivele compuse de cuantile (H=5)). Am folosit și Bayesian Composite Tobit Q Reg la 10 nivele cuantilice Tobit (H=10). Rezultatele arată 4 variabile (x_1:Educație,x_3:Vârstă, x_4:Taxe, x_5:Copii mari) cu importanță majoră la 10 nivele cuantilice Tobit datorită probabilității cu valoare mai mare de 0.5. Dar restul variabilelelor explanatorii (, x_2:Experiență, x_6:Venit) au o importanță relativ mică Tobit Q Reg la 10 nivele cuantilice Tobit datorită probabilităâii mai mici de 0.5. Prin urmare putem elimina următoarele variabile(x_2: Experiență, x_6: Venit) din modelul compus Tobit Q Reg la nivelul de 10 cuantile. Mai multe detalii sunt prezentate în capitolul patru, rezultatele inițiale sunt publicate în [23] Fadel Hamid Hadi Alhusseini și Vasile Georgescu 2017 "Bayesian composite Tobit cuantilice regression." Journal of Applied Statistics (2017): pp 1-13.Conform rezultatelor, considerăm că metodele propuse (new Bayesian Lasso Tobit Q Reg și Bayesian Composite Tobit Q Reg) au o bună posibilitate în estimarea coeficienților și selecția variabilelor în modelul Tobit Q Reg.

Prin urmare, vom folosi aceste două metode propuse pentru modelarea relației dintre variabila de răspuns (investițiile băncilor irakiene) și nouă variabile explicative x_1:Depozite bancare , x_2:Ptofitul bancar, x_3:Cpaitalul bancar, x_1:Rezervele bancare, x_5: Împrumuturile bancare, x_6:Cheltuielile de promovare, x_7:Vârsta băncii, x_8:Numărul filialelor băncii, x_9:Datorii negative. The response variabilă (investițiile băncilor irakiene) este cenzurată din partea stângă la punctul zero. Modelul de regresie Tobit consideră modelul de regresie eficient cu variabila de răspuns cenzurat. Dar modelul de regresie Tobit este sensibil din multe probleme. De asemenea, modelul de regresie Tobit nu este capabil să furnizeze o informație completă despre relațiile stocastice dintre variabilele dependente și variabilele explicative.

Pentru a depăşi aceste probleme, a fost utilizat modelul Tobit cuantilice regression (Tobit Q Reg). Pentru a analiza datele noastre (date privind investițiile bancare), vom folosi modelul de regresie cuantilică Tobit după cum urmează:

$y_i = \max(0, T_i^*)$ $\quad ,T_i^* = \alpha + \beta_{1\theta}x_{1j} + \beta_{2\theta}x_{2j} + \beta_{3\theta}x_{3j} + \beta_{4\theta}x_{4j} + \beta_{5\theta}x_{5j} + \beta_{6\theta}x_{6j} + \beta_{7\theta}x_{7j} + \beta_{8\theta}x_{8j} + \beta_{9\theta}x_{9j} + u_{i\theta}$ $\quad j$=1,2,.......47

unde: y este variabila de răspuns cenzurat (investițiile băncilor irakiene). T_i^* este variabila latentă $x_i, [i = 1,2, \ldots 9]$ sunt variabile explicative ca descrierea de mai sus. $(\beta_{1\theta}, \ldots, \beta_{9\theta})$ iar parametrii evaluați prin metodele propuse, după cum urmează:

Noul Bayesian Lasso Tobit Q Reg este utilizat pentru estimarea coeficienților și selectarea variabilelor în modelul Tobit Q Reg, care este prezentat mai sus prin intermediul a treizeci de nivele cuantilice Tobit. Aceasta înseamnă că vom obține 30 de modele de regresie quantificală Tobit. Am angajat metoda propusă (New Bayesian Lasso Tobit Q Reg) în două părți:În primul rând: estimarea coeficienților modelului de regresie Tobit cuantilic prin treizeci de nivele cuantilice Tobit.

În al doilea rând: vom determina importanța relativă a variabilelor explicative în modelul regresiei cuantice Tobit prin abordarea probabilistică. Dacă variabila explicativă are o probabilitate mai mare de 0,5, aceasta înseamnă că are o mare importanță relativă în construirea modelului nostru. Dar atunci când, valoarea sa de probabilitate mai mică de 0,5. Aceasta înseamnă că are o importanță relativă neglijabilă în construirea modelului nostru. Prin urmare, îl putem elimina din acest model. Pentru a identifica variabilele independente active în băncile din Irak, investiții care depind de importanța relativă a acestor variabile, au fost utilizate treizeci de nivele cuantilice diferite Tobit.

Metoda propusă (compozitul Bayesian Tobit Q Reg) are o înaltă calitate pentru estimarea coeficienților și selecția variabilelor în modelul compozit Tobit Q Reg Aici vom folosi șase grupuri de nivele Tobit cuantilice (cinci nivele cuantilice Tobit, zece nivele cuantilice Tobit, douăzeci de nivele cuantilice Tobit, douăzeci și cinci de nivele cuantilice Tobit și treizeci de nivele cuantilice Tobit). Fiecare grup are un model compozit specific Tobit Q Reg, bazat pe numărul de nivele cuantilice Tobit. Vom folosi metoda noastră (compozitul Bayesian Tobit Q Reg) în două părți după cum urmează:În primul rând: estimarea coeficienților în modelul compozit Tobit Q Reg la șase grupuri de nivele cuantilice Tobit: În al doilea rând: selecția variabilă în modelul de regresie compozit Tobitla șase grupuri de nivele cuantilice Tobit.

CONCLUZII ȘI CERCETĂRILE VIITOARE

Obiectivul principal al acestei lucrări este implementarea estimării Bayesiene a variabilelor și a estimării coeficienților în modelul de regresie cuantilică Tobit printr-un set de metode noi propuse. Metodele propuse în modelul Tobit Q Reg și modelul compozit Tobit Q Reg sunt considerate o nouă adăugare în abordarea Bayesiană regularizată. Și aceste metode propuse sunt foarte eficiente în comparație cu alte metode din același domeniu, acest lucru fiind clar din rezultatele exemplelor de simulare și ale datelor reale care au fost utilizate. Metodele propuse (noul Bayesian Lasso Tobit Q Reg și compozitul Bayesian Tobit Q Reg) sunt folosite pentru a analiza datele privind investițiile băncilor irakiene. În cazul în care metoda noastră (noul Bayesian Lasso Tobit Q Reg) este folosită în două părți: în primul rând, ea este folosită pentru a modela relația dintre investițiile bancare irakiene și nouă variabile independente la treizeci de niveluri cuantilice Tobit. În al doilea rând, acesta este utilizat pentru determinarea importanței relative a variabilelor independente în modelul Tobit Q Reg, de asemenea, la treizeci de niveluri cuantilice Tobit pentru a realiza selecția variabilă. Metoda noastră compusă Bayesian Tobit Q Reg este utilizată în două părți, în primul rând, este utilizată pentru a modela relația dintre investițiile bancare irakiene și nouă variabile independente la șase grupuri de niveluri cuantilice Tobit. În al doilea rând, este utilizat pentru determinarea importanței relative a variabilelor independente în modelul compozit Tobit Q Reg, de asemenea, la șase grupuri de nivele compuse Tobit cuantilice pentru a realiza selecția variabilă. Din studiul teoretic și Aplicat la metodele propuse, vom obține următoarele concluzii.

CONCLUZII TEORETICE:

- Din rezultatele studiilor de simulare și date reale la metodele propuse, vom ajunge la următoarele concluzii:
- - În metoda noastră propusă, noua regresie Bayesian Lasso cuantilice (noul Bayesian Lasso Q Reg) este atribuită amestecului independent al scalei de distribuții uniforme pentru coeficienții de regresie. Apoi, a fost prezentat un algoritm MCMC simplu și eficient pentru prelevatorul Bayesian. Studiile de simulare și un set real de date sunt utilizate pentru a investiga performanța metodei propuse în comparație cu alte metode existente. Ambele exemple de date simulate și reale arată că metoda propusă este destul de bună în comparație cu celelalte metode în cadrul unei serii de scenarii.
- - Studiile de simulare arată că eșantionul nostru Gibbs este eficient în contracția și estimarea coeficienților de regresie în cadrul unei serii variate de scenarii. De asemenea, scenariile noastre de simulare arată că metoda propusă funcționează bine chiar și atunci când distribuția adevărată pentru termenul de eroare nu este distribuția asimetrică Laplace (ALD). Acest caz a fost observat și de Yuan și Yin (2010)) [71], (Li et al., 2010) [52], (Alhamzawi et al., 2012) [1] și (Ji et al.)) [37], printre altele.
- Metoda propusă (noua regresie cuanelă Bayesiană) ne oferă un algoritm eficient și simplu de eșantionare Gibbs cu distribuții aposteriori condiționale bune. Algoritmul de eșantionare propus de Gibbs are multe avantaje și nu este complicat de utilizat, deoarece realizarea selecției variabilelor este evidentă și metoda noastră este mai echilibrată în comparație cu alte metode studiate.
- algoritmul MCMC propus pentru noua metodă Bayesian Lasso Q Reg a fost foarte stabil, așa cum se arată în rezultatele care aparțin factorului de reducere a scării potențiale multivariate (MPSRF). Metoda noastră a devenit stabilă și aproape de una după 2000 de

iterații, prin fiecare nivel de cuantil (0,50, 0,75 și 0,95), arătând că convergența distribuției aposteriori pentru metoda propusă a fost rapidă și amestecarea a fost bună.

- Metoda propusă (noul Bayesian Lasso Q Reg) consideră o nouă adăugare în domeniul modelului Q Reg regularizat Bayesian. Prin reformulare, distribuția a priori Laplace devine o nouă formulă de distribuție mixtă între distribuția uniformă și distribuția gama, uniformă a amestecului.
- - Rezultatele metodei de simulare și date reale indică faptul că metoda propusă (noul Bayesian Lasso Q Reg) are o performanță bună comparativ cu alte metode non-Bayesian și Bayesian. În plus, metoda noastră propusă este luarea în considerare a uneia dintre metodele competitive în selectarea variabilelor de estimare a coeficienților și precizia previziunilor.
- - una dintre caracteristicile bune și valoroase ale modelului Q Reg are o robustitate și funcționează bine chiar dacă încalcă presupunerea normală a erorii etc. Dar, în cadrul parametrului Bayesian Q Reg, am presupus că eroarea aparține distribuției asimetrice Laplace (ALD). În timp ce această ipoteză provoacă unele îngrijorări privind pierderea (verificarea) naturii modelului nonparametric Q Reg. Deci, rezultatele metodei propuse (noul Bayesian Lasso Q Reg) sunt destul de insensibile la această presupunere și se comportă bine pentru datele generate de alte distribuții de erori.
- - Pentru a evalua metoda propusă (noul Bayesian Lasso QReg), am testat-o cu alte trei metode, prima metodă aparține metodelor non-Bayseiene, iar a doua și a treia metodă aparțin metodelor Bayesiene. Din rezultatele care aparțin criteriului de direcție directă, estimarea coeficienților aparținând metodei propuse este mult închisă cu parametri adevărați comparativ cu metodele non-Bayesian și Bayesian. Acest lucru indică faptul că metoda propusă a fost mai bună decât alte metode.

- întreaga literatură din domeniul modelelor Bayesian Q Rag regularizate a utilizat distribuția de transformare Laplace la un amestec de distribuție normală, care a fost propus de Andrews și Mallows (1974) [5], având următoarea formulă:

$$\frac{\lambda_j}{2}\, e^{-\lambda_j|\beta_j|} = \int_0^{\infty} \frac{1}{\sqrt{2\pi s_j}} e^{\left(-\frac{\beta_j^2}{2s_j}\right)} \frac{\lambda_j^2}{2} e^{\left(-\frac{s_j\lambda_j^2}{2}\right)} ds_j$$

în scopul obținerii algoritmilor eficienți și simpli pentru estimarea coeficienților și selecția variabilelor în modelul Q Reg, vezi (Li et al. (2010)) [2010], (Alhamzawi et al., 2012) [1]. În metoda noastră propusă (noul Bayesian Lasso Q Reg), a fost utilizată o altă formă de transformare pentru distribuția Laplace, care este o uniformă a amestecului de scară ca următoarea formulă:

$$\frac{\lambda}{2} e^{\{-\lambda|\beta_j|\}} = \int_{u_j>|\beta_j|}^{\infty} \frac{1}{2u_j} \frac{\lambda^2}{\Gamma(2)}\; u_j{}^{2-1}\, exp\{-\lambda u_j\}\, du_j$$

- Această formulă a fost propusă de Mallick și Yi (2014) [55]. Din noua structură a distribuției Laplace am obținut un algoritm simplu și tractabil al metodei propuse pentru selectarea variabilelor și estimarea coeficienților în modelul Q Reg.
- Caracteristica metodei propuse (New Bayesian Lasso Q Reg) este că distribuția adecvată înainte este flexibilă, cu nivele diferite de cuantilă. Comportamentul amestecului de distribuție uniformă înainte este, evident, perfect robust, cu câteva nivele de cuantil. Amestecarea distribuției uniforme a priori este foarte importantă în modelul Bayesian Q Reg regularizat. Prin urmare, metoda noastră de New Bayesian Lasso QReg pare a fi foarte importantă în numeroase aplicații, de exemplu, selecția variabilă și studiile longitudinale.

- Metoda propusă are o performanță bună cu datele reale. În acest scop au fost utilizate date privind poluarea aerului. Acest set de date a fost măsurat de Administrația Drumurilor Publice din Norvegia și constă din 500 de observații, 7 covariate și o variabilă de răspuns. Eroarea medie pătrată este generată de metoda propusă este mai mică decât eroarea medie pătrată care este generată de alte metode. Într-un caz, vedem în studiul real de date că metoda non-Bayesiană (rq) a fost cea mai bună din metodele Bayesiene, până la metoda propusă (noul laeso QReg Bayesian) la nivelul cuantil mediu (θ_2 = 0,75). Dar metoda noastră propusă a înregistrat o bună performanță în comparație cu alte metode prin majoritatea nivelelor cuantilice.
- În această teză am folosit extensii din noua metodă Bayesian Lasso Q Reg pentru noua metodă, deoarece noul Bayesian Lasso Tobit Q Reg este considerat o nouă metodă în modelul regizat Bayesian Tobit Q Reg, realizând împreună selectarea variabilelor și estimarea coeficienților. Ierarhia Bayesiană aplicată pentru generarea distribuțiilor aposteriori complete condiționate se calculează din funcția de densitate condițională în comun și din distribuția precedentă uniformă a amestecului de scară pentru a crea algoritmul de eșantionare Gibbs.
- - Metoda propusă are scopul de a genera o nouă ierarhie bayesiană prin utilizarea uniformă a amestecului uniform (SMU) înainte de distribuţia coeficienților modelului Tobit Q Reg, pentru a obține estimarea coeficienților și selectarea variabilelor, unde SMU este considerat un înlocuitor bun pentru scară amestec normal (SMN) la modelul Bayesian Lasso Tobit Q Reg reglementat. Algoritmul MCMC derivat din distribuțiile aposteriori complete condiționate, simplu și eficient. Performanța metodei propuse este evaluată comparativ cu alte metode prin exemple de simulare și date reale. Rezultatele atât în studiul de simulare, cât și în datele reale au înregistrat pentru metoda propusă noul Bayesian Lasso Tobit Q Reg rezultate mai bune decât în comparație cu alte metode prin diferite nivele de dozare. Prin urmare, aceasta poate fi considerată o metodă bună pentru estimarea coeficienților și selecția variabilelor în modelul Tobit Q Reg.
- Atunci când modelul găzduiește un număr mare de variabile independente, nu există nici o garanție că parametrul de pedeapsă pentru complexitatea modelului este adecvat pentru realizarea selecției variabilelor cu multe dimensiuni. Prin urmare, metoda propusă (noul Bayesian Lasso Tobit Q Reg) consideră o abordare bună pentru obținerea selecției variabile în modelul Tobit Q Reg cu multe dimensiuni. De asemenea, metoda propusă nu are nevoie de mult timp pentru realizarea selecției variabilelor, care se face automat.

  Metoda propusă (noul Lasso Bayesian în Tobit Q Reg) folosește un amestec de distribuție uniformă (SMU) pentru o nouă distribuție a priori a ierarhiei la parametrii modelului care conduc la producerea de distribuții aposteriori condiționale complete pentru a construi tractabile și eficiente (MCMC) algoritm pentru implementarea selecției variabilelor și estimarea coeficienților în modelul Tobit Q Reg.

De asemenea, toți autori care lucrează în domeniul reglementat Bayesian Tobit Q Reg folosesc distribuția de transformare Laplace la un amestec de distribuție normală, propus de Andrews și Mallows: (1974)) [5] ca formula următoare.

$$\frac{\lambda_j}{2}\, e^{-\lambda_j|\beta_j|} = \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} e^{\left(-\frac{\beta_j^2}{2s_j}\right)} \frac{\lambda_j^2}{2} e^{\left(-\frac{s_j\lambda_j^2}{2}\right)} ds_j$$

.

Pentru obținerea algoritmului eficient și simplu de estimare a coeficienților și de selecție a variabilelor în modelul Tobit Q Reg (a se vedea (Ji și alții (2012)) [37], Alhamzawi (2014) [3] printre altele.

În metoda noastră propusă (noul Bayesian Lasso Tobit Q Reg) se folosește o altă formă de transformare pentru distribuția Laplace este uniformitatea amestecului de scală, după cum urmează

$$\frac{\lambda}{2}e^{\{-\lambda|\beta_j|\}} = \int_{s_j>|\beta_j|}^{\infty} \frac{1}{2u_j}\frac{\lambda^2}{\Gamma(2)}\ {u_j}^{2-1}\, exp\{-\lambda u_j\}\, du_j$$

Formula a fost propusă de (Mallick și Yi (2014)) [55]. Din noua structură a distribuției Laplace am obținut un algoritm simplu și tractabil pentru selectarea variabilelor și estimarea coeficienților în modelul Tobit Q Reg.

- algoritmul nostru de eșantionare Gibbs generează o distribuție aposteriori completă și informativă în selecția variabilelor în modelul Tobit Q Reg. Aceste avantaje în eșantionarea noastră Gibbs aparțin unei distribuții a priori corespunzătoare, care este compus din două părți. Prima parte este alocată distributiei uniforme a priori, iar a doua parte este atribuită distribuției Gamma cu parametrul de formă doi și parametrul de scală. Această distribuție a priori a ierarhiei asigură algoritmului Gibbs o eșantionare a algoritmului de eșantionare cu modele de dimensiuni mari în comparație cu alte metode prin care se pierd aceste avantaje.

- Estimarea coeficienților în modelul Tobit Q Reg este implementată prin minimizarea funcției de verificare (pierdere):

$$\underset{\alpha_\theta,\beta_\theta}{min} = \sum_{i=1}^{n} \rho_\theta\,(y_i - max\{0, T_i^*\})$$

Cu toate acestea, nu este diferențiat la punctul de origine, deci nu există o formă exactă a soluției pentru acești parametri (Koenker, (2005)) [42]. Minimizarea acestei funcții de verificare poate fi rezolvată printr-un algoritm de programare liniară (Koenker și D'Orey, (1987)) [46]. De aceea, metoda noastră propusă consideră o nouă metodă care va contribui la estimarea coeficienților în modelul Tobit Q Reg și consideră metode eficiente atunci când variabila de răspuns are date cenzurate ridicate.

- Pentru evaluarea metodei propuse (noua regresie Bayesian Lasso Tobit cuantilice), am comparat-o cu alte două metode, primul aparține metodelor non-Bayseiene, iar a doua metodă aparține metodelor Bayesiene. Rezultatele sunt generate prin intermediul a patru studii de simulare, care se încheie cu comportamentul metodei noastre propuse în depășirea metodei Bayesian și non-Bayesian până la datele cenzurate. De asemenea, estimarea coeficienților care aparține metodei propuse este închisă cu parametri adevărați comparativ cu metodele non-Bayesian și Bayesian. Aceasta indică performanțele metodei propuse față de alte metode.

- Metoda propusă noastră consideră o abordare bună cu datele reale. Pentru a evalua metoda propusă, au fost utilizate datele extramaritalilor. Este introdus de Fair în (1978). Aceste date au fost găsite în pachetul AER de la R. Eroarea medie pătrată generată de metoda propusă este mai mică decât eroarea medie pătratică care este generată de celelalte metode. Prin urmare, metoda propusă are o performanță mai bună decât alte metode cu date reale.

- algoritmul nostru MCMC aparține metodei noastre (noua lassă Bayesian Tobit Q Reg), care este puternică și atractivă pentru a realiza selecția variabilă prin calcularea importanței relative pentru fiecare covarianți din modelul nostru l Tobit Q Reg.

- S-a demonstrat că modelele compuse Q Reg sunt tehnici influente în dezvoltarea preciziei de predicție (Zou and Yuan, (2008)) [76]; (Bradic și colab., (2011)) [12]; (Zhao și Xiao, 2014) [77]. Metoda propusă ne propune să studiem compozitul Tobit Q Reg dintr-o abordare Bayesiană. O metodă de calcul eficientă și simplă bazată pe MCMC este derivată pentru inferențele de distribuție aposteriori utilizând un amestec de distribuție normală exponențială și scalată a distribuției asimetrice Laplace (ALD). Abordarea este studiată prin exemple de simulare și date reale. Aceste rezultate arată că informațiile colectate în diferite nivele cuantilice pot oferi o metodă bună în estimarea statistică eficientă. Aceasta este considerată prima lucrare de studiu a regresiei compuse Tobit cuantilice printr-o perspectivă Bayesiană.

- În modelul de regresie cuantică Tobit, există un număr de infinit de linii Tobit Q Reg la diferite nivele de cuante Tobit. Prin urmare, procesul de alegere a celei mai bune linii Tobit Q Reg este o problemă dificilă. Pentru a depăși această problemă, este necesar să se folosească regresia compozitului Tobit cuantilice pentru a obține estimatori la nivele cuantilice diferite pentru a obține un câștig de eficiență. Metoda propusă (Bayesian Composite Tobit Cuantilice Regression) consideră o nouă adăugare în estimarea coeficienților modelului compozit Tobit Q Reg.

- În noua metodă propusă (Bayesian Composite Tobit Q Reg), noua distribuție ierarhică a priori și funcția Likelihood (ALD) pentru eroarea propusă de Kozumi și Kobayashi (2011) [49] vor produce distribuții aposteriori condiționate . Aceste distribuții aposteriori complete condiționate sunt informative pentru construirea unui algoritm MCMC puternic și eficient pentru metoda propusă pentru a realiza estimarea coeficienților și selectarea variabilelor în modelul compozit Tobit Q Reg cu o precizie ridicată.

- Algoritmul nostru MCMC a fost prezentat pentru un compozit Bayesian Tobit Q Reg; a fost foarte stabil, acest lucru este clar din rezultatele factorului de reducere a scării potențiale multivariate (MPSRF); se calculează prin simularea 1 la cinci tipuri de distribuții de erori, unde devine stabilă și aproape de una după 3000 de iterații. Aceasta arată că convergența distribuțiilor posterioare condiționale totale pentru algoritm a fost foarte rapidă, iar amestecarea lanțului a fost bună.

În această teză am dezvoltat un algoritm MCMC simplu și eficient bazat pe tehnica de calcul pentru compozitul Tobit Q Reg bazat pe un amestec de distribuție normală exponențială și scalată a distribuției asimetrice Laplace. Studiile de simulare arată că metoda noastră propusă este eficientă în estimarea coeficienților și distribuțiilor de erori diferite. Metoda propusă (Bayesian Composite Tobit Q Reg) are o performanță superioară față de alte metode. Vedem performanțele metodei propuse cu distribuții de eroare ale amestecului care sunt mai bune decât distribuțiile de erori non-amestec.

După evaluarea metodei propuse folosind metoda de simulare, vom evalua metoda noastră propusă (compozitul Bayesian Tobit Q Reg) cu un set de date real. Am utilizat datele de participare la forța de muncă disponibile în pachetul AER în R, introduse de Mroz (1987) [6]. Setul de date constă din n = 753 observații din care 325 sunt observații cenzurate. Aceste date conțin variabila de răspuns (răspunsul soției ore de lucru în 1975 (ore)) și șase covariate. Metoda propusă are o bună performanță în comparație cu o altă metodă cu nivel ridicat de cenzurare în variabila de răspuns. Prin urmare, metoda propusă (compozitul Bayesian Tobit Q Reg) are o performanță bună cu setul de date real. Algoritmul nostru nou aparținând metodei noastre (compoziția bayesiană Tobit Q Reg) a fost eficient și a reprezentat o modalitate simplă de a realiza selecția variabilă prin calcularea importanței relative pentru fiecare covariante din modelul nostru compozit model Tobit Q Reg.

CONCLUZII DE ORDIN APLICATIV

- Metodele propuse (New Bayesian Lasso in Tobit quantile regression și Bayesian Composite Tobit Cuantile Regression) sunt aplicate pe datele financiare pentru modelarea relației dintre variabila răspuns (investiții bancare irakiene) și un set de variabile independente. Prin diferite nivele cuantilice Tobit și grupuri de nivele cuantilice Tobit din rezultatele noastre, am ajuns la un set de concluzii generale:
- Din rezultatele pătratului Pseudo-R, cea de-a treizecea linie de regresie cuantilică Tobit, care aparține intervalului (($\theta_{30} = 0.99$) a fost cea mai bună din toate liniile de regresie cuantilică Tobit utilizate pentru a reprezenta datele studiate. În cazul în care pătratul Pseudo-R aparține modelului Tobit Q Reg la $\theta_{30} = 0.99$ este egal cu 0.573283, ceea ce înseamnă că 57.32 % din variația variabilei de răspuns cenzurate poate fi explicată printr-un set de variabile independente (x_1 :Depozit bancar ,x_2 Profit bancar ,x_3 Capital bancar ,x_4 :Rezerve bancare ,x_5 :Împrumuturi bancare , x_6 : cheltuieli de publicitate ,x_7 :Vârsta băncii ,x_8 Numărul filialelor bancare,x_9: Datorii negative). Deși este cea mai puternică linie pentru interpretarea datelor studiate comparativ cu restul liniilor de regresie cuantice Tobit, nu are o mare rezistență în interpretarea datelor respective. Prin urmare, am folosit linia de regresie cuantică Tobit, care poate interpreta puternic datele studiate, dar procesul de identificare este unul foarte dificil.Rezultatele sunt prezentate ca toate modele compozite Tobit cuantilice care apartin sase grupe de nivele cuantice Tobit au o abilitate ridicata in explicarea datelor studiate. Rezultatul este clar din pseudo-pătrat.
- Variabilele independente (x_1 :depozit bancar, x_2: Profit bancar, x_4:Rezerve bancare) au un efect statistic semnificativ asupra variabilei de răspuns (investițiile băncilor irakiene) prin toate grupurile de nivele cuantilice Tobit. De asemenea, variabilele independente (x_3: Capital bancar, x_7:Vârsta băncii, x_8:Number of Banks Branches, x_9: Datorii negative) au un efect statistic semnificativ asupra variabilei de răspuns (investițiile băncilor irakiene) prin majoritatea grupurilor de nivele cuantilice Tobit. Dar există două variabile independente (x_5:Imprumuturi bancare, x_6:Cheltuieli de promovare) sunt nesemnificative în variabila de răspuns (investiții bancare irakiene) prin toate grupele de niveluri de cuantilice Tobit.
- la șase grupuri de niveluri Cuantilice Tobit, există un set de variabile independente care sunt active în construirea modelului compozit Tobit Q Reg prin diferite niveluri CuantiliceTobit după cum urmează: La primul grup (H = 5), există șase variabile independente (x_1: Depozite bancares, x_2: Profit bancar, x_3: Capital bancar, x_6 : Cheltuieli de promovare, x_8 :Numărul filialelor băncii, x_9: Datorii negative). În al doilea grup (H = 10), există șase variabile independente (x_1: Depozite bancare, x_2:Profit bancar, x_3: Capital bancar, x_6:Cheltuieli de promovare, x_8 :Numărul filialelor băncii, x_9: Datorii negative) eficiente în construirea acestui model. În al treilea grup (H=15), există șapte variabile independente (x_1: Depozite bancare, x_2: Profit bancar, x_3: Capital bancar, x_4: Rezerve bancare,x_6 : Cheltuieli de promovare,x_8 :Numărul filialelor băncii,x_9: Datorii negative, x_9: Datorii negative) puternic în structura acestui model. În al patrulea grup (H = 20), există opt variabile independente (x_1: Depozite bancare, x_2: Profit bancar, x_3: Capital bancar, x_4: Rezerve bancare, x_6 : Cheltuieli de promovare, x_7: Vârsta băncii, x_8 :Numărul filialelor băncii, x_9: Datorii negative) puternice în structura acestui model. În al cincilea grup (H = 25), există nouă variabile independente (x_1 :Depozit bancar ,x_2 Profit bancar ,x_3 Capital bancar ,x_4 :Rezerve bancare ,x_5 :Imprumuturi bancare , x_6 : cheltuieli de promovare ,x_7 :Vârsta băncii ,x_8 Number of Banks Branches ,x_9: Datorii negative) active

în construcția modelului nostru. La al șaselea grup (H=30), sunt nouă variabile independente (x_1 :Depozit bancar ,x_2 Profit bancar ,x_3 Capital bancar ,x_4 :Rezerve bancare ,x_5 :Imprumuturi bancare , x_6 : cheltuieli de promovare ,x_7 :Vârsta băncii ,x_8 Number of Banks Branches ,x_9: Datorii negative) wcu importanță relativ înaltă în construcția modelului nostru.

- Variabilele independente (x_1: Depozite bancare, x_2: Profit bancar, x_3: Capital bancar, x_6 : Cheltuieli de promovare, x_8 :Numărul filialelor băncii, x_9: Datorii negative) sunt foarte importante în modelarea relației cu variabilele de răpsuns (Iraqi banks investments) prin intermediul grupurilor de nivele cuantilice Tobit. Restul variabilelor independente (x_4 :Rezerve bancare, x_5:Imprumuturi bancare, x_7:Vârsta băncii) sunt importante în construirea relației cu variabilele de răspuns (Investițiile băncilor irakiene) prin intermediul majorității grupurilor de nivele cuantilice Tobit. Tabelul de mai jos demonstrează cele spuse:

Table 6.1: Rezumatul statutului variabilelor independente prin intermediul a 6 grupuri de niveluri Cuantilice Tobit.

Variabile Independente / Grupe	x_1: Depozite banzare	x_2: Profit bancar	x_3 Capital bancar	x_4 :Rezerve bancare	x_5 :Imprumuturi bancare	x_6 : cheltuieli de promovare	x_7 :Vârsta băncii	x_8 Numărul filialelor băncii	x_9: Datorii negative
H=5	Activ	Activ	Activ	InActiv	InActiv	Activ	InActiv	Activ	Activ
H=10	Activ	Activ	Activ	InActiv	InActiv	Activ	InActiv	Activ	Activ
H=15	Activ	Activ	Activ	Activ	InActiv	Activ	InActiv	Activ	Activ
H=20	Activ	Activ	Activ	InActiv	Activ	Activ	Activ	Activ	Activ
H=25	Activ	Activ	Activ	Activ	Activ	Activ	Activ	Activ	Activ
H=30	Activ	Activ	Activ	Activ	Activ	Activ	Activ	Activ	Active

- • Pentru a determina puterea și slăbiciunea variabilelor independente din modelul Tobit Q Reg prin diferite nivele cuantilice Tobit prin metoda propusă (New Bayesian Tobit Q Reg) după cum urmează:

$\boldsymbol{x_1}$**: Depozite bancare** : Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_{28} = 0.92$. Valoarea sa de probabilitate mai mare este (0.839), mai mare de 0.5 . Prin urmare este importantă construcția modelului Tobit Q Reg la $\theta_{28} = 0.92$. Valoarea de probabilitate mai mică a acestei variabile în modelul Tobit Q Reg este la $\theta_{23} = 0.79$. Valoarea probabilității sale mai mici este (0.599), mai mare de 0.5 Este importantă în construcția modelului Tobit Q Reg la $\theta_{28} = 0.92$ În general x_1: Depozite bancare sunt active în modelul Tobit Q Reg la toate nivelele cuantilice Tobit. Nu putem șterge această variabilă din model.

$\boldsymbol{x_2}$**: Profit bancar :**Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_{28} = 0.92$. Unde valoarea probabilității este (0.853) mai mare de 0.5 . Este importantă în construcția modelului Tobit Q Reg la $\theta_{28} = 0.92$. Valoarea de probabilitate mai mică a acestei variabile în modelul Tobit Q Reg este la $\theta_1 = 0.01$. Valoarea probabilității sale mai mici este (0.413), mai mică de 0.5. Prin urmare nu este importantă în construcția modelului Tobit Q Reg la $\theta_1 = 0.01$. În general**,** x_2: Profit bancar este activă în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Putem depinde de aceată variabilă în cosntrucția modelului nostru.

x_3 **Capital bancar**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_{30} = 0.99$. Valoarea probabilității sale este (0.797), mai mare de 0.5 . Prin urmare este importantă în construcția modelului Tobit Q Reg la $\theta_{30} = 0.99$. Valoarea de probabilitate mai mică a acestei variabile în modelul Tobit Q Reg este la $\theta_{19} = 0.66$. Valoarea probabilității sale este (0.298) mai mică de 0.5. Prin urmare este slabă în construcția modelului $\theta_{19} = 0.66$. În general, x_3 Capital bancar nu este important în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Putem ignora această variabilă în structura modelului nostru.

x_4: **Rezerve bancare:** Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_3 = 0.08$. Valoarea probabilității sale este (0.778) mai mare de 0.5. Este importantă în construcția modelului Tobit Q Reg la $\theta_3 = 0.08$. Valoarea probabilității sale mai mici la $\theta_{21} = 0.74$ este (0.309) mai mică de 0.5. Prin urmare este slabă în construcția modelului **la** $\theta_{21} = 0.74$. În general, x_4: Rezerve bancare este important în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Nu putem ignora această variabilă în structura modelului nostru.

x_5 **:Imprumuturi bancare**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la θ_3 $\theta_{30} = 0.99$. Valoarea probabilității sale este (0.802) este mai mare de 0.5 . Este importantă în construcția modelului Tobit Q Reg la $\theta_{30} = 0.99$. Valoarea probabilității sale mai mici la $\theta_{12} = 0.42$. Valoarea probabilității sale mai mici la (0.332) este mai mică de 0.5. rin urmare este slabă în construcția modelului **la** $\theta_{12} = 0.42$. În general, x_5 :Imprumuturi bancare este important în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Nu putem ignora această variabilă în structura modelului nostru.

x_6 **: Cheltuieli de promovare**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_{30} = 0.99$. Valoarea probabilității sale este (0.811) mai mare de 0.5 . So, it is strong in building a Modelul Tobit Q Reg at $\theta_{30} = 0.99$. Valoarea probabilității sale mai mici la $\theta_{19} = 0.66$. Valoarea probabilității sale este (0.261) mai mică de 0.5. Este medie la construcția modelului la $\theta_{19} = 0.66$. În general, x_6 : Cheltuieli de promovare este slab la majoritatea nivelelor cuantilice Tobit Q. O putem șterge din modelul nostru.

x_7 **:Vârsta băncii**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_{30} = 0.99$. Valoarea probabilității sale este (0.822) mai mare de 0.5 . . Este importantă în construcția modelului Tobit Q Reg la $\theta_{30} = 0.99$. Valoarea probabilității sale mai mici la $\theta_{18} = 0.63$. Valoarea probabilității sale mai mici este (0.339) mai mică de 0.5. Prin urmare nu este eficientă în constrcția modelului Modelul Tobit Q Reg la $\theta_{18} = 0.63$. In general, x_7 :Vârsta băncii este activă în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Putem depinde de aceată variabilă în cosntrucția modelului nostru.

x_8: **Numărul de filiale ale băncii**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_{30} = 0.99$. Valoarea probabilității sale este (0.828) mai mare de 0.5 . Este eficientă în structura modelului Tobit Q Reg la $\theta_{30} = 0.99$. Valoarea probabilității sale mai mici la $\theta_{17} = 0.60$ este (0.325), mai mică 0.5. Nu este atât de important a construcția modelului Tobit Q Reg lat $\theta_{17} = 0.60$. In general, x_8 Numărului de filiale ale băncii nu este important în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Putem ignora această variabilă în structura modelului nostru.

x_9**: Datorii negative:** Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg este la $\theta_{30} = 0.99$. Valoarea probabilității sale este (0.875) mai mare de 0.5 . So, it is very strong in building a Modelul Tobit Q Reg at $\theta_{30} = 0.99$. Valoarea probabilității sale mai mici la $\theta_{12} = 0.42$. Valoarea probabilității sale mai mici este (0.591) mai mare de 0.5 . și este eficientă în construcția modelului Tobit Q Reg la $\theta_{12} = 0.42$. In general, x_9: Datorii negative este activă în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Putem depinde de aceată variabilă în cosntrucția modelului nostru.

- Variabilele independente (x_1: Depozite bancare,x_9: Datorii negative) au valoarea probabilității mai mare de 0.5 in Tobit Q Reg la toate nivelele cuantilice. Prin urmare aceste variabile independente sunt foarte active în modelul nostru. Și variabilele independente (x_2: Profit bancar,x_5 :Imprumuturi bancare, x_7 :Vârsta băncii) au valoarea probabilității mai mare de 0.5 in Modelul Tobit Q Reg la majoritatea nivelelor cuantilice Tobit. Prin urmare aceste variabile independente sunt foarte puternice în modelul nostru. Dare variabilele independente (x_3: Capital bancar, x_6 : Cheltuieli de promovare, x_8 :Numărul filialelor băncii) au valoarea probabilității mai mică de 0.5 in Modelul Tobit Q Reg la majoritatea nivelelor cuantilice Tobit. Prin urmare aceste variabile sunt foarte slabe în cosntrucția modelului nostru, așa că le putem șterge.

Pentru a determina puterea și slăbiciunea variabilelor independente din compozitul Modelul Tobit Q Reg am folosit șase grupuri de nivele cuantilice în compozitul Tobit prin metoda propusă (compozitul Bayesian Tobit Q Reg) după cum urmează:

$\boldsymbol{x_1}$**: Depozite bancare** : Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la zece nivele cuantilice în compozitul Tobit [H=10] . Valoarea probabilității sale este (0.917) mai mare de 0.5 . Prin urmare este foarte activă în construcția compozitului Modelului Tobit Q Reg la [H=10] . Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la cinci nivele cuantilice în compozitul Tobit [H=5] . Valoarea probabilității sale este (0.893) mai mare de 0.5 . Este foarte activ în Modelul Tobit Q Reg la [5] . In general, x_1: Depozite bancare este activă în modelul Tobit Q Reg la majoritatea nivelelor cuantilice. Putem depinde de aceată variabilă în cosntrucția modelului nostru.

$\boldsymbol{x_2}$**: Profit bancar :**Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la twenty five nivele cuantilice în compozitul Tobit [H=25] . Valoarea probabilității sale este (0.907) mai mare de 0.5 . Prin urmare este foarte eficientă în Modelul Tobit Q Reg la [25] . Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la cinci nivele cuantilice în compozitul Tobit [H=5] . Valoarea probabilității sale este (0.880) mai mare de 0.5 . Este de asemenea importantă în construcția Modelului Tobit Q Reg la [5] . In general, x_2: Profit bancar are o importanță relativ mare în Modelul Tobit Q Reg la cele șase grupuri de nivele cuantilice Tobit. Putem depinde de această variabilă în construcția modelului nostru.

$\boldsymbol{x_3}$ **Capital bancar**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la treizeci de nivele cuantilice în compozitul Tobit [H=30] . Valoarea probabilității sale este (0.703) mai mare de 0.5 . Prin urmare este foarte eficientă în Modelul Tobit Q Reg la $[\mathrm{H} = 30]$. Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la cinci nivele cuantilice în compozitul Tobit [H=5] . Valoarea probabilității sale este (0.642) mai mică de 0.5. Este de asemenea importantă în construcția Modelului Tobit Q Reg la [5]. In general, x_3 Capital bancar are o importanță relativ mare în Modelul Tobit Q Reg la cele șase grupuri de nivele cuantilice Tobit. Putem depinde de aceată variabilă în cosntrucția modelului nostru.

$\boldsymbol{x_4}$: **Rezerve bancare:** Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la twenty composite Tobit cuantilice levels [H=20]. Valoarea probabilității sale este (0.617) mai mare de 0.5 . Prin urmare este foarte eficientă în Modelul Tobit Q Reg la $[H = 20]$. Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la cinci niveluri cuantilice [H=5] . Valoarea probabilității sale este (0.346) mai mică de 0.5. Este de asemenea importantă în construcția Modelului Tobit Q Reg la $[\mathrm{H} = 5]$. In general, x_4: Rezerve bancare are o importanță relativ mare în Modelul Tobit Q Reg la cele șase grupuri de nivele cuantilice Tobit. Putem depinde de aceată variabilă în cosntrucția modelului nostru.

x_5 **:Imprumuturi bancare**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la douăzeci și cinci de niveluri cuantilice [H=25]. Valoarea probabilității sale este (0.667) mai mare de 0.5 . Prin urmare Imprumuturi bancare este foarte eficientă în Modelul Tobit Q Reg la [25]. Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la 20 de nivele cuantilice Tobit [H=20]. Valoarea probabilității sale este (0.306) mai mică de 0.5. Prin urmare este foarte slabă la structura modelului la [H=20] . In general, x_5 :Imprumuturi bancare are o importanță relativ mică în Modelul Tobit Q Reg la cele șase grupuri de nivele cuantilice Tobit. Nu putem depinde de aceată variabilă în cosntrucția modelului nostru.
x_6 **: Cheltuieli de promovare**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la treizeci de nivele cuantilice Tobit [H=30]. Valoarea probabilității sale este (0.853), mai mare de 0.5 . Prin urmare este importantă în Modelul Tobit Q Reg la [30]. Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la cinci nivele cuantilice Tobit [H=5]. Valoarea probabilității sale este (0.802) mai are de 0.5. . Prin urmare este foarte eficientă în Modelul Tobit Q Reg la [H=5]. In general, x_6 : Cheltuieli de promovare are o importanță relativ mare în Modelul Tobit Q Reg la cele șase grupuri de nivele cuantilice Tobit. Putem depinde de această variabilă în cosntrucția modelului nostru.
x_7 **:Vârsta băncii**: Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la douăzeci de nivele cuantilice [H=20]. Valoarea probabilității sale este (0.684) mai mare de 0.5 . Este prin urmare activă în Modelul Tobit Q Reg la [H=20]. Valoarea probabilității mai mici al variabilei Vârsta băncii in Modelul Tobit Q Reg la zece nivele cuantilice Tobit [H=10]. Valoarea probabilității sale este (0.324) mai mică de 0.5. Prin urmare nu este activa in Modelul Tobit Q Reg la [10]. In general, x_7 :Vârsta băncii este activ așa că putem depinde de această variabilă în construcția modelului nostru.
x_8 **Numărul de filiale ale băncii** : Valoarea de probabilitate mai mare pentru această variabilă în modelul Tobit Q Reg la douăzeci și cinci de nivele cuantilice [H=25]. Valoarea probabilității sale este (0.903) mai mare de 0.5 . Prin urmare este foarte eficientă în Modelul Tobit Q Reg la [25]. Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la douăzeci de nivele cuantilice Tobit [H=20]. Valoarea probabilității sale este (0.865) mai mare de 0.5. Este foarte eficientă în construcția modelului la [H=20] . In general, x_8 Number of Banks Branches este foarte puternică în modelul nostru, la cele șase grupuri de nivele cuantilice Tobit . Nu o putem omite din modelul nostru.
x_9**: Datorii negative:** The Larger probability value to Datorii negative in composite Modelul Tobit Q Reg at twenty five composite Tobit cuantilice levels [H=25].Valoarea probabilității sale este (0.910) mai mare de 0.5 . Prin urmare este foarte eficientă în construcția modeluilui la[25]. Valoarea probabilității mai mici a acestei variabile în Modelul Tobit Q Reg la cinci nivele cuantilice Tobit [H=5]. Valoarea probabilității sale este (0.892) mai mare de 0.5 . De asemenea este foarte eficientă în construcția modelului la [5]. In general, x_9: Datorii negative is very strong in composite Modelul Tobit Q Reg la cele șase grupuri de nivele cuantilice Tobit . Nu putem ignora această variabilă din modelul nostru.

- Variabilele independente (x_1: Depozite bancare, x_2: Profit bancar, x_3: Capital bancar, x_6 : Cheltuieli de promovare, x_8 :Numărul filialelor băncii,x_9: Datorii negative) au valoarea probabilității mai mare de 0.5 in Modelul Tobit Q Reg la cele șase grupuri de nivele cuantilice Tobit. Prin urmare aceste variabile sunt foarte importante in modelul nostru. Și variabilele independente (x_4: Rezerve bancare, x_7 :Vârsta băncii) au valoarea probabilității mai mare de 0.5 in Modelul Tobit Q Reg la majoritatea celor sase grupuri de nivele cuantilice Tobit. Prin urmare aceste variabile independente sunt foarte puternice în modelul nostru. Insa variabilele independente (x_5 :Imprumuturi bancare) au valoarea probabilității mai mică de 0.5 in Modelul Tobit Q Reg la majoritatea nivelelor cuantilice Tobit. Prin urmare, pe acestea le putem ignora din modelul nostru.

Cercetări viitoare

Metodele propuse au o bună posibilitate în selectarea variabilelor și estimarea coeficienților într-un set de modele de regresie. Prin urmare, aceste metode pot fi extinse cu ușurință în mai multe moduri, după cum urmează:

- Metoda propusa noua regresie Bayesian Lasso cuantilică poate fi usor extinsa prin utilizarea unui amestec scalar uniform in loc de amestec scalar normal unde un amestec scalar de uniforme va crea o noua formula ierarhica Bayesian de Lasso adaptive. Tratamentul Bayesian nou așteptat duce la o distribuție aposteriori condiționată și simplă a probelor Gibbs. Abordările simulării și datele reale sunt utilizate pentru a testa performanța metodei propuse în comparație cu alte metode din același domeniu. Metoda preconizată propusă consideră o nouă adăugare în selecția variabilă și precizia predicției în modelul de regresie cuantilă la diferite nivele cuantilice.

- Metoda noastră new Lasso Bayesian Tobit Q Reg poate fi extinsă la noua regresie Bayesian adaptive Lasso Tobit cuantilice cu o nouă distribuție ierarhică a priori prin utilizarea unei uniforme de amestec scalar care generează o nouă ierarhie Bayesiană pentru distribuția aposteriori condiționată completă. Se așteaptă ca un nou Bayesian ierarhic să ne ofere un eșantion Gibbs atractiv și informativ. De asemenea, noul model Bayesian adaptabil Lasso Tobit Q Reg așteptat este foarte eficient în estimarea coeficienților și selecția variabilă în modelul de regresie Tobit cuantilice. Scenariile de simulare și setul de date real vor fi folosite pentru evaluarea noului Bayesian adaptiv Lasso Tobit Q Reg comparativ cu alte metode din același domeniu.

- Metoda propusă noul model Bayesian Lasso Tobit Q Reg poate fi extins cu ușurință la metoda Tobit Q Reg compozit Lasso Bayesian, cu o scară uniformă înainte de amestecare prin atribuirea unui amestec independent de scară de distribuții uniforme a priori parametrilor modelului. Această sugestie creează o nouă structură a distributiilor a priori tuturor parametrilor modelului. Această nouă stricțiune generează distribuții aposteriori condiționale informative care conduc la algoritmul de eșantionare Gibbs atractiv și eficient și acest algoritm este foarte robust până când variabila de răspuns are date mult mai cenzurate. Pentru a evalua această metodă propusă se va folosi metoda de simulare și datele reale.
- Regresia canalelor binare a fost dezvoltată de Manski (1975, 1985) și utilizată în clasificare, indicând dezavantajele proceselor frecvente date fiind dificultatea de optimizare pentru a estima parametrii și problema calculului intervalului de încredere față de parametri. Kordas (2006) a studiat modele cu variabila răspuns binar prin regresie cuanală și a concluzionat că această abordare conduce la o bună clasificare. Abordarea Bayesiană a fost adoptată de Benoit et al (2012) [11] pentru a evita dezavantajul menționat mai sus prin stabilirea unor presupuneri privind termenul de eroare. Miguéis și colab. (2013) a considerat abordarea propusă de (Benoit et al. (2012)) [11] pentru a evalua riscul de credit și a fost modelată de regresia binară cuanală. Romanul din acest studiu propus este modelul ierarhic Bayesian pentru a estima coeficienții modelului de regresie cuanelă compusă atunci când variabila de răspuns este binară. Pentru a selecta variabilele, în compoziția binară Lasso pentru regresia cuanlelor și pedeapsa Lasso adoptivă este derivată dintr-un cadru Bayesian.

Bibliografie

[1] Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. Statistical Modelling, 12(3), 279-297.

[2] Alhamzawi, R. and Yu, K. (2014). “Bayesian Lasso-mixed quantile regression.” *Journalof Statistical Computation and Simulation*, 84(4): 868–880. 2

[3]Alhamzawi, R. (2014). Bayesian elastic net tobit quantile regression. Communications inStatistics - Simulation and Computation 45 (7), 2409–2427.

[4] Alhamzawi, R., 2013, “*Tobit quantile regression with adaptive Lasso penalty*”, The 4th International Scientific Conference of Arab Statistics 450 ISSN, pp. 1681-6870).

[5] Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B 36, 99–102. 7

[6] Abrevaya, J., & Dahl, C. M. (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics*, *26*(4), 379-397.

[7] Arto Luoma (2014). Introduction to Bayesian analysis, University of Tampere, Finland

[8] Buchinsky, M. and J. Hahn (1998). An alternative estimator for censored quantile regression.*Econometrica 66*, 653–671.

[9] Bilias, Y., S. Chen, and Z. Ying (2000). Simple resampling methods for censored regression quantiles. *Journal of Econometrics 68*, 303–338.

[10] Baraniuk, R. G. (2007). “Compressive sensing.” *IEEE signal processing magazine*, 24(4).

[11] Benoit, D. F. and D. V. D. Poel (2012). Binary quantile regression: a Bayesian approach based on the asymmetric Laplace distribution. *Journal of AppliedEconometrics 27*, 1174–1188.

[12] Bradic, J., J. Fan, and W. Wang (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society, Ser. B 73*,325–349.

[13] Benoit, D. F., R. Alhamzawi, and K. Yu (2013). Bayesian Lasso binary quantile regression.*Computational Statistics 28* (6), 2861–2873.

[14] Cade, B. S. and Noon, B. R. (2003). “A gentle introduction to quantile regression for ecologists.” *Frontiers in Ecology and the Environment*, 1(8): 412–420. 1

[15] Candes, E. and Tao, T. (2007). “The Dantzig selector: statistical estimation when p is much larger than n.” *The Annals of Statistics*, 2313–2351. 2

[16] Candes, E. J. and Recht, B. (2009). “Exact matrix completion via convex optimization.”*Foundations of Computational mathematics*, 9(6): 717–772. 2

[17] Chernozhukov, V., & Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, *142*(1), 379-398.

[18] Dunia report (2014) . Iraqi Private sector banking

[19] Ikechukwu, I. O., & Boniface, U. U. (2016). Examination of the Relationship between Bank Age and Bank Retention Policy (A Study of Zenith Bank (Nig) Plc). *International Journal of Finance and Accounting*, *5*(5), 233-239.

[20] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.

[21] Fair, R. C. (1978). A theory of extramarital affairs. *The Journal of Political Economy*,45–61.

[22] Fohlin, C. (2014). A Brief History of Investment Banking from Medieval Times to the Present.

[23]Fadel Hamid Hadi Alhusseini, and Vasile Georgescu (2017). "Bayesian composite Tobit quantile regression." Journal of Applied Statistics: pp 1-13.

[24]Fadel Hamid Hadi Alhusseini (2017) " BAYESIAN QUANTILE REGRESSION WITH SCALEMIXTURE OF UNIFORM PRIOR DISTRIBUTIONS " International Journal of Pure and Applied Mathematics pp 77-91.

[25]Fadel Hamid Hadi Alhusseini (2017):. "New Bayesian Lasso in Tobit Quantile Regression." Romanian Statistical Review Supplement65.6 pp 213-229.

[26] Fadel. Hamid. Hadi Alhusseini (2016). Tobit Quantile Regression and Iraqi Banks' Profit. The young Economists journal. 27 ,141-152

[27] Fadel. Hamid. Hadi Alhusseini (2017) Using the Bayesian Technique to the Tobit Quantile Regression with Scale Mixture Uniform for Selecting Important Variables Affecting Iraqi Investment Banking. journal of applied quantitative methods ,volume 12 issue1.

[28] Fadel Hamid Hadi Alhusseini. " SELECTION OF VARIABLES INFLUENCING IRAQI BANKS DEPOSITS BY USING NEW BAYESIAN LASSO QUANTILE REGRESSION." Journal of Social and Economic Statistics 6.1 (2017): 45-59.

[29] Griffin, J. E. and Brown P. J. Brown(2010). Bayesian adaptive Lassos with non-convex penalization. Technical report. Institute of Mathematics, Statistics and Actuarial Science, University of Kent.

[30] Greene, W. (1999). Marginal effects in the censored regression model.*Economics Letters*, *64*(1), 43-49.

[31] Greene, W. (2010)."econometric analysis ".*seventh edition New York University*

[32] Gramacy, R. B., & Lee, H. K. H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, *103*(483), 1119-1130.

[33] Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory 11*,105–121.

[34] Huang, H. and Z. Chen (2015). Bayesian composite quantile regression. *Journal of Statistical Computation and Simulation 85* (18), 3744–3754.

[35]Hala Hijazi (2017) " ISLAMIC AND CONVENTIONAL BANKS PROFITABILITY MODELLING " Alkhawayn university , SCHOOL OF SCIENCE & ENGINEERING

[36] Iraq Central Bank (2013). *Financial stability report*.

[37] Ji, Y., N. Lin, and B. Zhang (2012). Model selection in binary and Tobit quantile regression using the Gibbs sampler. *Computational Statistics* & *Data Analysis 56*, 827–839.

[38] Koenker, R. and G. J. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

[39] Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, *91*(1), 74-89.

[40] Koenker, R., & Hallock, K. (2001). Quantile regression: An introduction.*Journal of Economic Perspectives*, *15*(4), 43-56.

[41] Kostov, P., & Davidova, S. (2013). A quantile regression analysis of the effect of farmers' attitudes and perceptions on market participation. *Journal of Agricultural Economics*, *64*(1), 112-132.

[42] Koenker, R. (2005). *Quantile Regression*. Cambridge Books. Cambridge University Press.

[43] Koenker, R., & Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*,*96*(454), 458-468.

[44] Kotz,S., Kozubowski, T. J., and Podg`orski, K. (2001). *The Laplace Distribution and Generalizations:A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Birkh¨auser,Boston.

[45] Koenker, R. (2011). quantreg: Quantile regression. R package version 4.71.

[46] Koenker, R. and V. D'Orey (1987). Algorithm AS 229: Computing regression quantiles. Journal of the Royal Statistical Society: Series C (Applied Statistics) 36, 383–393.

[47] Kai, B., R. Li, and H. Zou (2010). Local composite quantile regression smoothing: an efficientand safe alternative to local polynomial regression. *Journal of the Royal Statistical Society:Series B (Statistical Methodology) 72* (1), 49–69.

[48] Koenker, R. and Machado, J. A. (1999). "Goodness of fit and related inference processes for quantile regression." Journal of the american statistical association, 94(448): 1296–1310. 3

[49] Kozumi, H. and Kobayashi, G. (2011). "Gibbs sampling methods for Bayesian quantile regression." Journal of statistical computation and simulation, 81(11): 1565–1578. 3

[50] Koenker, R. (2013). quantreg: Quantile regression. R package version 5.05.

[51] Kobayashi, G. and H. Kozumi (2012). Bayesian analysis of quantile regression for censored dynamic panel data. *Computational Statistics 27* (2), 359–380.

[52] Li, Q., Xi, R., & Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, *5*(3), 533-556.

[53]Li, Y., & Zhu, J. (2008). L 1-norm quantile regression. Journal of Computational and Graphical Statistics, 17(1), 163-185.

[54] Lawrence, K. D., & Arthur, J. L. (Eds.). (1990). *Robust regression: analysis and applications*. New York: Marcel Dekker.

[55] Mallick, H. and Yi, N. (2014). "A new Bayesian Lasso." *STATISTICS AND ITS INTERFACE*, 7(4): 571–582. 2, 3, 4

[56] Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica: Journal of the Econometric Society*, 765–799.

[57]Nasr, S., Petersen, A., Van der Vossen, J., Hashad, N., Britton, R., Kulaksiz, S., & Huitfeld, E. (2011). *IRAQ: FINANCIAL SECTOR REVIEW*(No. 10869). The World Bank.

[58] Ors, E. (2006). The role of advertising in commercial banking. Finance and Economics Department . School of Management, Paris

[59] Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics 32*, 143–155.

[60] Park, T. and Casella, G. (2008). The Bayesian Lasso. Journal of the American Statistical Association 103, 681–686.

[61] Reich, B. J., H. D. Bondell, and H. J. Wang (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics 11*, 337–352.

[62] Reed, C. and K. Yu (2009). A partially collapsed Gibbs sampler for Bayesian quantile regression. Technical report, Brunel University, Department of Mathematical Sciences.

[63] Sun, W., J. G. Ibrahim, and F. Zou (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. Genetics 185, 349–359.

[64] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.

[65] Tibshirani, R. (2011). Regression shrinkage and selection via the Lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273-282.

[66] Tobin, J., 1958, "Estimation of relationships for limited dependent variables", Econometrica, 26, pp. 24-36.

[67] Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 801-817.

[68] Wei, Y., Pere, A., Koenker, R., and He, X. (2006). "Quantile regression methods for reference growth charts." *Statistics in medicine*, 25(8): 1369–1382. 1

[69] Wang, H., & He, X. (2007). Detecting differential expressions in GeneChip microarray studies: a quantile approach. *Journal of the American Statistical Association*, *102*(477), 104-112.

[70] White, E. N. (1986). Before the Glass-Steagall Act: An analysis of the investment banking activities of national banks. *Explorations in Economic History*, *23*(1), 33-55.

[71] Yuan, Y. and Yin, G. (2010). Bayesian quantile regression for longitudinal studies with non-ignorable missing data. Biometrics 66, 105–114.

[72] Yu, K. and Zhang, J. (2005). "A three-parameter asymmetric Laplace distribution and its extension, "*Communications in Statistics - Theory and Methods*, **34**, 1867–1879.

[73] Yue, Y. R., & Hong, H. G. (2012). Bayesian Tobit quantile regression model for medical expenditure panel survey data. *Statistical Modelling*, *12*(4), 323-346.

[74] Yu, K., W. Dang, H. Zhu, and R. Al Hamzawi (2013). Comment on article by spokoiny,wang and hardle. *Journal of Statistical Planning and Inference 7* (143), 1140–1144.

[75] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

[76] Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics 36*, 1108–1126.

[77] Zhao, Z. and Z. Xiao (2014). Efficient regressions via optimally combining quantile information. *Econometric theory 30* (06), 1272–1314.