# CENSORED AND QUANTILE REGRESSION ESTIMATION IN TWO-PART MODELS FOR CENSORED SEMI-CONTINUOUS DATA

**PhD Dissertation**

**İsmail YENİLMEZ**

**Eskişehir 2021**

# CENSORED AND QUANTILE REGRESSION ESTIMATION IN TWO-PART MODELS FOR CENSORED SEMI-CONTINUOUS DATA

**İsmail YENİLMEZ**

**PhD Dissertation**

**Department of Statistics**

**Programme in Statistical Theory**

**Supervisor: Prof. Dr. Yeliz MERT KANTAR**

**Eskişehir**

**Eskişehir Technical University**

**Institute of Graduate Programs**

**November 2021**

*This thesis has been supported by the project 20DRP007 which is approved by the SRP Committee. This thesis has been also supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK).*

## ABSTRACT

CENSORED AND QUANTILE REGRESSION ESTIMATION IN TWO-PART MODELS FOR CENSORED SEMI-CONTINUOUS DATA

İsmail YENİLMEZ

Department of Statistics

Programme in Statistical Theory

Eskişehir Technical University, Institute of Graduate Programs, November 2021

Supervisor: Prof. Dr. Yeliz MERT KANTAR

Data with an abundance of zeros arise on many occasions. Traditional statistical methods often cannot deal with this data which can be called zero-inflated data. The zero inflation caused by censored zero is often modeled by the Tobit model. However, when the distribution of the error term is not normal, the maximum likelihood estimator of the Tobit model is inconsistent. In this study, a generalization of the Tobit model based on the extended normal distribution with two additional shape parameters is proposed. In addition, the two-part model is used for modelling the zero inflation caused by real zero. The inflations caused by real zero on the left and the censoring point on the right are investigated for semi-continuous data in this study. The proposed generalized censored regression model is adjusted to the two-part model, moreover, the two-part quantile regression modelling framework is also modified for censored semi-continuous data. Two new models, the two-part generalized censored regression model and the two-part quantile regression model are introduced for censored semi-continuous data. For estimating the parameters of models, maximum likelihood estimators are obtained. The estimation process is conducted using the Expectation-Maximization (EM) algorithm. The performance of maximum likelihood estimators is evaluated by the Monte Carlo experiments. The proposed estimators are also applied to the Istanbul Chamber of Industry export data for an econometrical data example and wild boar dispersal data for an environmental data example.

**Keywords:** Zero-inflated data, Censored semi-continuous data, Two-part model, Generalized Tobit Model, Quantile regression.

# ÖZET

SANSÜRLÜ YARI SÜREKLİ VERİLER İÇİN İKİ PARÇALI MODELLERDE SANSÜRLÜ VE KANTİL REGRESYON TAHMİN EDİCİLERİ



İsmail YENİLMEZ

İstatistik Anabilim Dalı

İstatistik Teorisi Bilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Kasım 2021

Danışman: Prof. Dr. Yeliz MERT KANTAR



Çok sayıda sıfır içeren veriler birçok durumda ortaya çıkar. Geleneksel istatistiksel yöntemler sıfır şişirilmiş veri olarak adlandırılan bu verilerin analizinde çoğu zaman yetersiz kalır. Sansürlenmiş sıfırın neden olduğu sıfır enflasyon, sıklıkla Tobit model ile modellenir. Ancak hata teriminin dağılımı normal olmadığında, Tobit modelinin en çok olabilirlik tahmincisi tutarsızdır. Bu çalışmada, iki ek şekil parametresi ile genişletilmiş normal dağılıma dayalı Tobit modelinin genelleştirilmesi önerilmiştir. Ayrıca, gerçek sıfırın neden olduğu sıfır enflasyonunu modellemek için iki parçalı model kullanılır. Bu çalışmada yarı-sürekli veriler için soldaki gerçek sıfırın ve sağdaki sansür noktasının neden olduğu şişirmeler incelenmiştir. Önerilen genelleştirilmiş sansürlü regresyon modeli iki parçalı modele uyarlanmıştır. Ayrıca iki parçalı kantil regresyon modelleme çerçevesi de sansürlü yarı sürekli veriler için düzenlenmiştir. Sansürlü yarı sürekli veriler için iki yeni model; iki parçalı genelleştirilmiş sansürlü regresyon modeli ve iki parçalı kantil regresyon modeli tanıtılmıştır. Modellerin parametrelerini tahmin etmek için en çok olabilirlik tahmin edicileri elde edilmiştir. Tahmin süreci, Beklenti-Maksimizasyon (EM) algoritması kullanılarak gerçekleştirilmiştir. En çok olabilirlik tahmin edicilerinin performansı Monte Carlo deneyleri ile değerlendirilmiştir. Önerilen tahmin ediciler, ekonometrik veri örneği için İstanbul Sanayi Odası ihracat verilerine ve çevresel veri örneği için yaban domuzu dağılım verilerine de uygulanmıştır.



**Anahtar Sözcükler:** Sıfır şişirilmiş veri, Sansürlenmiş yarı sürekli veri, İki parçalı model, Genelleştirilmiş Tobit Model, Kantil regresyon.

## ACKNOWLEDGEMENTS

First of all, I would like to thank my esteemed and kind supervisor Prof. Dr. Yeliz MERT KANTAR, for her support, encouragement, and patience, who has carried out my consultancy during my postgraduate as well as during my doctoral period. Her contributions allowed my studies to go the extra mile and have been invaluable throughout the writing of this dissertation.

I would like to express my thanks to my patient and supportive mentor, Prof. Dr. Lea PETRELLA, who has supported me throughout the dissertation and made indescribable contributions to the writing of this dissertation. Her insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would like to offer special thanks to Prof Dr. İlhan USTA, for his valuable comments and suggestions from a theoretical perspective throughout the writing of this dissertation. His immense knowledge and plentiful experience have encouraged me throughout my academic studies.

I would like to express my endless thanks to my colleague from the Sapienza University of Rome Mr. Luca MERLO, with whom I could communicate without hesitation during my research period in Italy. I am grateful for his contributions to programming.

I would also like to thank the Department of Methods and Models for Economics, Territory, and Finance for welcoming me wonderfully throughout my research period at the Sapienza University of Rome.

Last but not least, I would like to express my sincere the Scientific and Technological Research Council of Turkey (TÜBİTAK) for the 2214 / A International Doctoral Research Fellowship Program run by Science Fellowships and Grant Programme Directorate.

And my biggest thanks to my family for the unconditional and loving support they have shown me throughout this research. Thank you, my little cute girls, Asya and Deniz. And my wife Begüm, the best person in the world, thanks for all your support.

İsmail YENİLMEZ

16/11/2021

## STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with "scientific plagiarism detection program" used by Eskişehir Technical University, and that "it does not have any plagiarism" whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

İsmail YENİLMEZ

## CONTENTS

## LIST OF TABLES

**Page**

## LIST OF FIGURES

## GLOSSARY OF SYMBOLS AND ABBREVIATIONS

| | | |
|---|---|---|
| ZIP | : | Zero-inflated Poisson |
| ZINB | : | Zero-inflated Negative Binomial |
| LDV | : | Limited Dependent Variable |
| DV | : | Dependent Variable |
| CR | : | Censored Regression |
| CDV | : | Censored Dependent Variable |
| LR | : | Linear Regression |
| QR | : | Quantile Regression |
| 2-PM | : | Two-part model |
| 2P-GCR | : | Two-part generalized censored regression model |
| 2P-FMQR | : | Two-part finite mixture quantile regression model |
| 2P-FMQRC | : | Two-part finite mixture quantile regression model for censored semi-continuous data |
| CLAD | : | Censored Least Absolute Deviation |
| STLS | : | Symmetrically Trimmed Least Squares |
| OLSE | : | Ordinary Least Squares Estimator |
| ML | : | Maximum Likelihood |
| MLE | : | Maximum Likelihood Estimator |
| Q-MLE | : | Quasi-Maximum Likelihood Estimator |
| NPML | : | Non-parametric Maximum Likelihood |
| CDF | : | Cumulative Distribution Function |
| PDF | : | Probability Density Function |
| AL | : | Asymmetric Laplace |
| EG | : | Exponentiated Generalized |
| EN | : | Extended Normal |
| EGT | : | Exponentiated Generalized Tobit |
| MCAR | : | Missing Completely at Random |
| CCA | : | Complete Case Analysis |
| iid | : | independent and identically distributed |
| ICI | : | Istanbul Chamber of Industry |

# 1. INTRODUCTION

The zero observations frequently arise in many fields (see, e.g., Martín-Fernández et al. 2003; Blasco-Moreno et al., 2019; Amore and Murtinu, 2021; Boulton and Williford, 2018; Chai and Bailey, 2008; Humphreys, 2013; Dow and Norton, 2003). Conventional statistical methods often cannot overcome data with an abundance of zeros. There are many types of zeros (false zeros, true zeros [censored and rounded zeros due to detection of limit, random zeros, structural zeros, real zeros, etc.], zeros associated with latent outcomes). The statistical models that deal with excess zeros diversify depending on the types of zeros. In other words, the analysing method to be chosen varies according to the zero sources and occurrence of zeros. Therefore, the types of zeros are the key point to determine the appropriate statistical model. Awareness of subtle differences in the types of zeros and determination of appropriate models for data with an abundance of zeros is crucial for the accuracy of the analysis. However, there is some terminological confusion about the methods used to distinguish and model zeros. In order to avoid confusion and give a brief review, types of zeros are summarized in the next section. Types of zeros, source of zeros, possible problems related to excess zeros, and models in terms of their ability to handle excess zeros are presented.

*Zero inflation*, which occurs as a result of excess zeros, has generally been discussed in the literature for two data types: *count data* and *semi-continuous data*. For further details, see also Neelon et al., 2016a; Neelon et al., 2016b. Zero inflation occurs when there are more zeros in the data than the distribution allows for. Zero inflation can also cause skewness and over-dispersion. Various terms such as *clumping at zero*, *spike-at-zero*, and *zero inflation* are used for the excess zeros and such terms are common in count and semi-continuous data.

*Count data* consist of integers that arise from counting rather than ranking. Count data that have an excessive number of zeros are often encountered in practice (Feng, 2021). The zero-inflated Poisson (ZIP) (Lambert, 1992), and the zero-inflated negative binomial (ZINB) (Greene, 1994) models are often used for zero-inflated count data. For further details on analysis of count data and zero inflation in count data, see Cameron and Trivedi, 2013. In addition, the *two-part model* (2-PM) as an alternative model specification for zero-inflated count data is introduced by Duan et al. (1983) and Mullahy (1986). In the 2-PMs, a binary choice model is fit for the probability of observing a positive-versus-zero outcome, thereafter, a model is fit for the positive outcomes conditional on positive outcomes (Belotti et al., 2015).

Non-negative continuous data with clumping at zero can be expressed as a mixture of discrete and continuous data. This hybrid data can be also called *semi-continuous data* (Shmueli et al., 2008). In general terms, a combination of a high mass at one point (often zero) and a continuous variable for the remaining part are defined as *semi-continuous data* (Buuren, 2018). The standard statistical methods are not appropriate for semi-continuous data (Shmueli et al., 2008) and they yield inefficient results. A number of statistical methods have been developed to tackle this issue (see, e.g., Belasco and Ghosh, 2012; Biswas et al., 2020). Since the 2-PMs include a joint distribution consisting of mixing a discrete point mass, with all mass at zero, and a discrete or continuous random variable (Merlo et al., 2021), the 2-PM can be used not only for zero-inflated count data analysis but also for semi-continuous data analysis.

The zero-inflated data can also be associated with censoring at zero (see, e.g., Eggers, 2015; Boulton and Williford, 2018). It is necessary to distinguish between censored, structural, and other types of zeros (Boulton and Williford, 2018). The censored dependent variable concept is a general definition used when data are observed under certain constraints. For analysing zero expenditures for durable goods, the censored normal regression (CR) model (hereafter, *Tobit model*) was developed by Tobin (1958) as a statistical regression model under normality assumption.

Consider the Tobit model for the fixed left-censored response values $y_i$:

$$y_i = \begin{cases} y_i^* = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i, \ z_i^* > c_L \\ c_L, \qquad\qquad otherwise \end{cases} \quad u_i \sim N(0, \sigma^2) \tag{1.1}$$

where $\mathbf{x}_i^{\mathrm{T}} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^{\boldsymbol{T}}$ is an $(n \times p)$ matrix of the explanatory variables, and $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \dots, \beta_p\right)^T$ is a $(p \times 1)$ dimensional vector of the regression coefficients for the parametric component of the model. $u_i$ represents random error term assumed to be independent and identically distributed (iid) with probability density function given as $f(u|.)$. $c_L$ is lower bound. In fact, in Tobin's seminal article, dependent variable observed under certain conditions are called *limited dependent variable* and the *censored dependent variable* or *censored regression model* expression is never used. Later, Tobit model has been overidentified with censored dependent variable, which is a sub-title of limited dependent variables.

In econometrics data, zeros are often associated with *limited dependent variables* (LDVs). It has been stated that most of the models that aim to overcome the zeros fall under the heading of models for LDVs (Humphreys, 2013). In fact, one of the first models that try to deal with zeros is the Tobit model proposed in Tobin's seminal work (Tobin 1958), and the Tobit model is also accepted as the first model that tries to deal with *semi-continuous* data (Min and Agresti, 2002; Gottard et al., 2013). However, the so-called Tobit model is frequently used in the presence of censored zero. Tobit model may not be appropriate for structural zeros as in semi-continuous data.

The standard Tobit model is calculated using a single equation for zero and positive outcomes as seen in Eq. (1.1). In other words, binary outcomes (For instance, the decision to buy durable goods or participation in consumption) and positive outcomes (amount paid for the goods or consumption amount) are taken into account by one equation. However, the 2-PMs use two equations for binary and positive outcomes. Therefore, single equation methods such as the Tobit model may not be appropriate for data with the mass point at zero. The 2-PMs are used as an alternative to the Tobit model, for instance, the probability of a binary outcome (0 or 1) is investigated by the probit model in the first part of the model and then the non-zero values are modeled by the truncated normal model in the second part (Cragg, 1971). In other words, binary and positive outcomes are modeled with two equations, and they are independently modeled by probit model and truncated normal regression model (Cragg, 1971). Cragg model, called Craggit, nests the Tobit model and forms the basis of the 2-PMs. The Heckit model is proposed by Heckman (1979) as another alternative to the Tobit model to deal with zeros. Heckit model, known as type-II Tobit model, also consists of two equations. Heckit is proposed as an alternative to Craggit (Heckman, 1979), the assumed independence between the two equations of Craggit is reversed. Thus, a more realistic approach has been obtained. In other words, Heckit has a modelling method that takes into account the dependency of the two equations. Heckit is often used for sample selection models.

The 2-PM introduced by Duan et al. (1983) is a non-selection model. In fact, the Heckit takes into account the potential outputs thus the Heckit differs slightly from Craggit and other 2-PMs. The difference between the Heckit and the 2-PMs is highlighted by Leung and Yu, (1996) and Dow and Norton, (2003). Especially, the difference between structural zero and censored zero and the difference between the estimation process of the Heckit and the 2-PMs despite their superficial similarity are emphasized by Belotti et al. (2015). The zeros in the Heckit may represent censored values or latent values, while zeros in 2-PMs can represent structural or real zeros. Thus, the difference between models determined depending on types of zeros is also presented by Belotti et al. (2015). Moreover, a number of articles (Duan et al., 1983, 1984; Manning et al., 1987) have criticized the sample selection models and suggested the usage of the 2-PMs. Hay and Olsen (1984) state that, contrary to the claims of Duan et al., (1983) the 2-PMs are built on unusual assumptions. The main discussion concerns the choice of the sample selection model and the 2-PM. The discussion, which is called the *cake debate*, takes place in the literature. As stated by Leung and Yu (1996), the debate and advocates should not be ignored. This issue, which has not yet been concluded in the literature, is remarkable, but this study does not deal with sample selection models, but only handles the modification of 2-PMs and new estimators for 2-PMs.

In addition, the expression *hurdle model*[1] encountered in the literature of 2-PMs is mostly used as a counterpart for the two-part modelling of count data (Belotti et al., 2015). It has also been mentioned that the *two-part model* represents two-part modelling for semi-continuous data. Following Belotti et al. (2015), the same concept has been used from here on to avoid conceptual confusion in this study.

Heretofore, spike-at-zero caused by censoring at zero and excessive zeros occurred by structural/real[2] zeros is mentioned. CR model is frequently used for the censored response variable and 2-PMs are frequently used for the response variable included structural/real zeros. The distinguishing of censored and structural/real zeros and analysis of the zeros is a complicated problem. Moreover, in the literature, the right-censored response variable with structural zero has been handled for count data. The analysis of the semi-continuous response variable with real zeros on the left and censored point on the right (the analysis of censored semi-continuous response variable) is also a problem.

[1] It should be noted that the *Hurdle* expression is also used for the Tobit (single hurdle model), Craggit (double hurdle model – independent hurdles), and Heckit (double hurdle model – dependent hurdles) (Rufino, 2016). Moreover, Craggit model used in the analysis of continuous data with the mass point at zero can be referred to as *two-stage*, *two-tier*, and *double hurdle model* (Burke, 2009).

[2] Structural zeros and real zeros represent zeros in count data and semi-continuous data, respectively.

Non-zero censored and zero-inflated response variable is addressed for count data: Santos and Bolfarine (2015), Saffari and Adnan (2011), Nguyen and Dupuy (2019), Yang and Simpson (2012) examine the non-zero censored and zero-inflated count data in the cross-sectional data setting. In contrast, the non-zero censored and zero-inflated data concept has rarely been discussed for semi-continuous data in the literature. As far as is known, the concept is handled in Freitas Costa et al., (2021). In other words, the investigation of the non-zero censored and zero-inflated data concept has recently gained attention. Data, including non-zero censored and excess zeros, and analysis of this data can be considered as a literature gap.

The main motivation of this study is to adjust the 2-PMs for censored semi-continuous data and to obtain maximum likelihood (ML) estimators of proposed models. The proposed models are able to take into account both extreme zeros on the left and censored values on the right in semi-continuous data to fill the aforementioned research gap.

The classical Tobit estimator of the parameters for the CR model is inefficient in case of non-normal error terms. To solve this problem, a generalization of the censored normal regression is initially proposed by the author. Then, generalized censored normal regression is extended to right-censored semi-continuous data (i.e., the generalized censored regression model based on the extended normal distribution with two additional shape parameters as $a$ and $b$ is adapted to the 2-PM process for right-censored semi-continuous data).

The conditional mean of the response variable is often used in the estimation in the framework of linear regression (LR). However, in the presence of excess zeros and censorship, assumptions of LR can be violated (Sauzet et al., 2019). Quantile regression (QR) can be considered as an alternative to the LR. QR, which stands out with seminal studies such as Koenker and Bassett (1978), Koenker and Machado (1999), and Koenker (2005), use the conditional median (or other quantiles) of the response variable instead of the conditional mean of the response variable. Thus, for modelling non-Gaussian, skewed, censored, truncated, and fat-tailed outcomes, QR can be utilized as an alternative. The advantages of QR are used in studies such as Yu and Stander (2007), Kozumi and Kobayashi (2011), and King and Song (2019) in estimating the censored regression models and 2-PMs. In addition, QR is used in the presence of both non-zero censored and zero-inflated data for the count data setting by Santos and Bolfarine (2015), Saffari and Adnan (2011), Nguyen and Dupuy (2019), and Yang and Simpson (2012). In this context, the two-part quantile regression modelling framework is adjusted for censored semi-continuous data in this study.

As a result, modifications of the 2-PMs have been proposed for data with non-zero censored values and excess zeros. CR and QR are used as base models for modification of the 2-PM to fulfill the mentioned aim of the study. The main reason for using CR is the superiority of CR in case of censored dependent variables, as well as its adaptability to real zeros. The reason for using QR is that it can analysis data according to different quantile levels, not based on the mean. The procedure based on the different quantiles makes using QR attractive for modelling non-Gaussian, skewed, censored, truncated, and fat-tailed outcomes. Therefore, QR has been utilized as an alternative for modelling non-Gaussian, skewed, censored outcomes.

In this study, (i) to investigate a two-part generalized censored regression model (2P-GCR), the model proposed by Moulton and Halsey (1995) is modified for censored semi-continuous data and the parameter of the modified model is estimated, (ii) to obtain a two-part quantile regression model for censored semi-continuous data (2P-FMQRC), the two-part finite mixture quantile regression model proposed by Merlo et al. (2021) is modified for censored semi-continuous data and parameter of the modified model is estimated. The likelihood function for 2P-FMQRC is constructed for censored semi-continuous data under the frequentist paradigm, since the minimization of the quantile loss function is equivalent to the maximization of the likelihood associated with the asymmetric Laplace (AL) density in terms of parameter estimates. Therefore, the ML estimation procedure of the two-part finite mixture quantile regression (2P-FMQR) model (Merlo et al., 2021) is followed and modified for the censored dependent variable in this study.

The spike-at-zero is governed with the help of the logit model in both models (2P-GCR; 2P-FMQRC). In 2P-GCR, the generalized censored regression model for the 2-PM is used in the second part, and in this way, the skewed and right-censored dependent variable is managed. In 2P-FMQRC, QR with the capability of compensating for the deterioration caused by censorship on the positive side is used in the second part. Moreover, in estimation, non-parametric ML estimation used to overcome the computational difficulties encountered in parameter estimation is calculated with the EM algorithm.

The rest of the paper is organized as follows. In Section 2, types of zeros, limited dependent variables, in which extreme zeros are handled from different perspectives, and subtitle of LDV, the censored regression model, are discussed in detail. Moreover, an extension of the generalized Tobit model for left-censored semi-continuous data is presented in Section 2. The two-part generalized censored regression (2P-GCR) and the two-part finite mixture quantile regression for censored dependent variable (2P-FMQRC) are presented and the estimations based on ML are considered in Section 3. Outputs of Monte Carlo simulations conducted under different conditions and the findings of the simulation experiments designed to compare the two proposed estimators are shared in Section 4. Section 5 presents the motivating datasets for application and discusses the main empirical results. Finally, all findings are discussed and presented in the Conclusion section.

# 2. ZERO INFLATION AND CENSORED RESPONSE DATA

In this section, zero types that can cause zero inflation, limited dependent variables associated with zeros, and censored regression used for censorship at zero or any point are presented.

## 2.1. Types of Zeros

The problem of the determination of the appropriate model for zero-inflated data may be solved by well-identified types of zeros. However, it is not always easy to distinguish zeros. Zero types can be confusing when the literature is examined, terminologically. Detailed studies for zero types are limited. In this study, zeros encountered in different disciplines are reviewed. Zero types and models are summarized with the help of Table 2.1 and Figure 2.1. In addition, some studies in the literature are summarized as follows:

Zeros are classified as *essential zeros* (absolute absence of the part in the observation) and *rounded zeros* (presence of a component below the detection limit) (Martín-Fernández et al. 2003).

For ecological data in the count data, *false zeros* (observer errors and the experimental design errors) and *true zeros* (structural zeros and random zeros) are deeply investigated. While it is stated that false zeros should be ignored in the analysis, structural zeros (emerged from ecological and environmental constraints) and random zeros (emerged from sampling variability) should be analysed. Structural zeros and random zeros are included in the analysis (Blasco-Moreno et al., 2019).

In strategy research, excess zeros are related to decision output (Yes/No) and the answer to the question of "How much?". It can cause a mass at the zero point. *True zeros* are associated with absence, and *non-true zeros* are associated with *missing-latent* and *censored* dependent variables (Amore and Murtinu, 2021).

In social sciences, excess zeros are examined as s*kewed continuous outcomes with many zeros* (Boulton and Williford, 2018). It is noted that extreme zeros are caused by *true absence* and *censoring due to the measurement limit*. While dealing with *true* and *censored zeros* causing right-skewed data, concepts such as *zero-inflated, floor effect, limited dependent variable, semi-continuous* are frequently used (Boulton and Williford, 2018). In the modelling phase, the suitability of the Tobit model in case of censorship and the 2-PM in case of semi-continuous data are emphasized (Boulton and Williford, 2018).

The occurrence of zeros is attributed to three circumstances for medical studies (Chai and Bailey, 2008). The measurement for the subject cannot occur over a certain period of time (*True zero*). *Censored zeros* are occurred due to failure to observe some positive values originating from the detection limit. Zeros can also occur due to technical reasons such as environmental or physical factors and are called them *missing or low* (Chai and Bailey, 2008).

While *zero-censored dependent variables* are often associated with models of LDVs (Humphreys, 2013), zeros associated with unobserved response variables or latent outcomes (it may be called *latent zeros*) are associated with the sample selection problem and are often solved with Heckit.

The analysis of *actual* and *potential zeros* is discussed in the applied health economics literature in terms of the proper use of the Heckit and the 2-PM. (Dow and Norton, 2003). The discussion of the use of Heckit and the 2-PMs dates back to the 1980s. Jones (2000) is a comprehensive source for the history of this discussion, referred to as "*cake debates*".

It is clear that determining the types of zeros and choosing the appropriate model for the analysis of zeros create confusion. To clarify this situation, the types of zeros are presented in Table 1.1. This table is an expanded version of the table presented by Blasco-Moreno et al. (2019).

A flowchart is also presented in Figure 1.1 for the zeros that can be encountered in the different fields, the problems that occur for the presence of excess zeros, and the methods to be used for modelling. This diagram is more comprehensive than the diagrams presented in Humphreys (2013) and Amore and Murtinu (2021) and can be a guide for researchers.

**Table 2.1.** *Source, types of zeros, and possible problems*

| Types of Zeros | Source | Possible Problems |
|---|---|---|
| **False Zeros** | Design-induced zeros such as poor experimental design | Analysis with incorrect measurements |
| | Zeros that the observer recorded as zero due to lack of experience | Analysis with incorrect measurements |
| | Zeros recorded as technically zero due to problems with measuring device | Analysis with incorrect measurements |
| **True Zeros** | Structural zeros representing the real absence for count data | Over-dispersion and zero inflation problems |
| | Real zeros representing the real absence for semi-continuous data | Over-dispersion and zero inflation problems |
| | Random zeros due to sampling variability | Spike-at-zero due to distributional assumptions |
| | Censored zeros due to the limit of detection | Information loss |
| | Zeros associated with choosing | Corner solution problem |
| | Zeros representing the joint decision where the binary part and the positive part are dependent | Inability to modelling two potentially distinct aspects of the case |
| | Sequential decision zeros where the binary part and the positive part are not dependent | Independence of hurdles |
| | Non-sequential decision zeros where the binary part and the positive part are dependent | Single stochastic process to determine whether the response value is zero or positive outcomes |
| **Latent Zeros** | Zeros associated with non-observed outcomes | Sample selection problem |

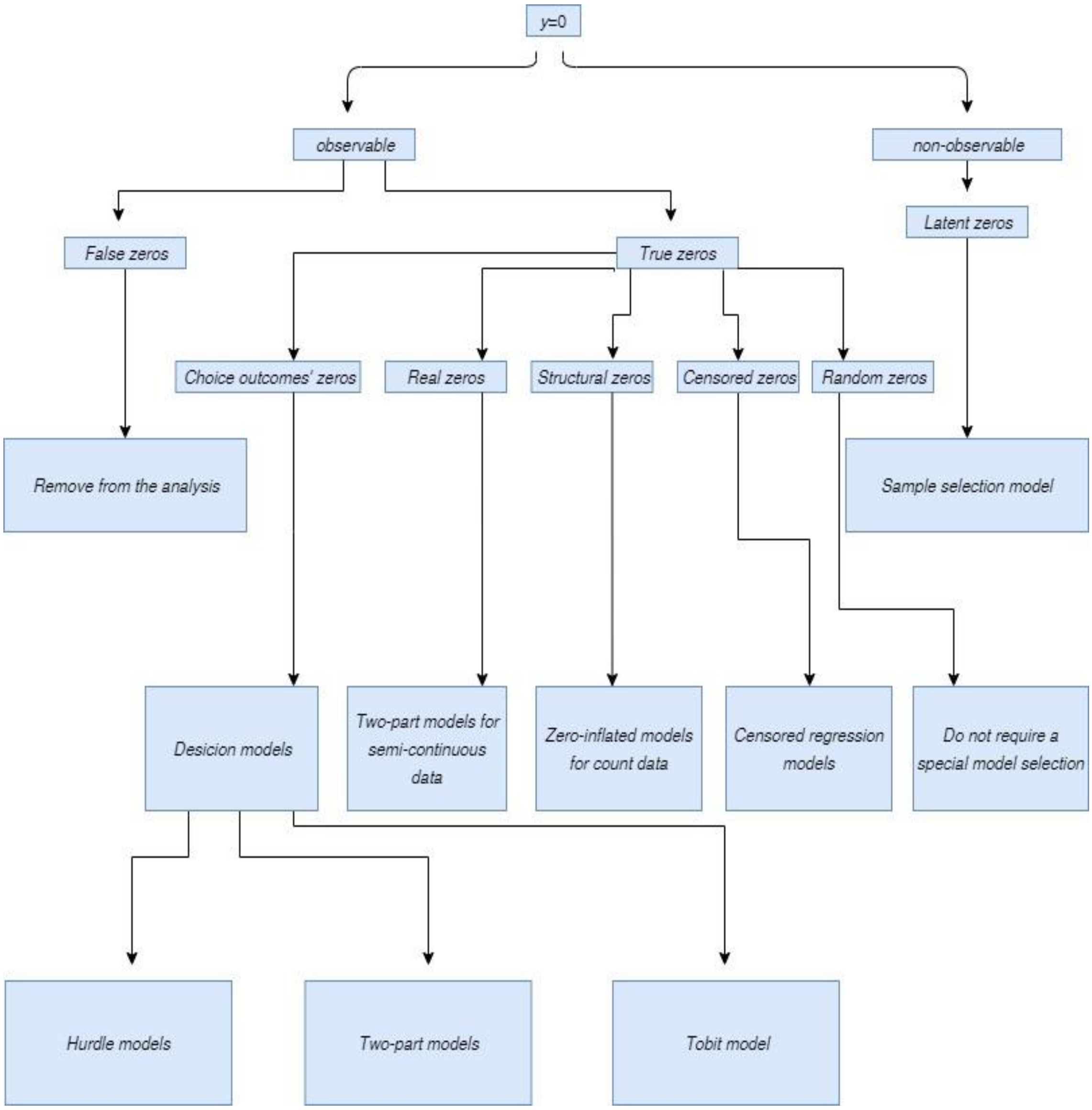



**Figure 2.1.** *Types of zeros and methods*

## 2.2. Limited Dependent Variables

In econometrics, excess zeros are often associated with LDVs. In this section, the scope of LDVs and its subtitle, the censored regression model, are discussed in detail. In statistical modelling, the stochastic linear equation is written as:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\boldsymbol{Y} = (y_1, y_2, \dots, y_n)'$ is $n \times 1$ vector of dependent variables (response variables), $\boldsymbol{X}$ represents $n \times (p+1)$ matrix of independent variables (explanatory - designed matrix), $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \dots, \beta_p\right)'$ is $(p+1) \times 1$ the vector of the coefficients, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ is $n \times 1$ vector of error terms $\left(\varepsilon_i \sim N(0, \sigma^2)\right)$. The dependent variable $\boldsymbol{Y}$, abbreviated as (DV), sometimes has a certain range or takes particular values. Such DV is called limited dependent variables (LDV). Discrete (dummy, ordered, categorical), non-negative, censored, truncated, count, duration, and sample selection variables have been handled within the LDVs framework in some studies. For instance, qualitative, truncated, and censored DVs have been dealt with regression and simultaneous equations in Maddala's seminal book related to limited-dependent and qualitative variables (Maddala, 1983). Another important source that dealt with limited dependent and categorical variables has handled variables as binary, nominal, censored, truncated, and counted (Long, 1997). Long (1997) has also emphasized that LDV models are non-linear models and they have certain problems in calculation and interpretation. Discrete and binary response models, multinomial and ordered response models, corner solution outcomes and censored regression models, sample selection issues, and count data are dealt with concerning each other. In another important book by Wooldridge (2012), LDV models and sample selection corrections are examined under the same heading. Wooldridge (2012) emphasized that binary dependent variable which takes only two values, discrete response variables such as the number of children, positive values such as hourly wage, and corner solution response variable which includes the decision of participation are examples of LDVs. Limited dependent and discrete variables are dealt with concerning each other by Johnston and DiNardo (1997). Truncated, censored, sample selection, and duration models are discussed by Greene (2003) in chapter LDV and duration models. In the next edition of Greene's effective book (2008), truncation and censoring, and sample selection are discussed together, while counts and duration models are examined together in another chapter. In the seventh edition of Greene's book (Greene, 2011), LDVs are discussed as truncation, censoring, and the 2-PMs. In addition, duration models are briefly discussed. These examples show how the topic of LDVs is addressed in efficacious econometrics books, and also it can be concluded that the term LDV has a dynamic structure. In other words, the literature is constantly changing and developing due to the exponential increase of the data and the lack of a certain form of data.

As a widely used concept, censored dependent variable (CDV), which is a special subject of LDVs, is the main subject of this study. CDVs are encountered in many fields. Statistical, econometric, survival, and reliability analysis include some of the methods that make various statistical inferences using censored response variables. However, CDVs do not have a settled terminological structure. For instance, CDV is defined as a continuous dependent variable that can take only a restricted – limited range of values in the influential econometrics book of Davidson and MacKinnon (2003). However, it is emphasized that another term matching this rough definition is truncation (Davidson and MacKinnon, 2003). Attention should be paid to some details so that these two terms are not confused. For instance, an upper limit is set, values higher than the upper limit are considered equal to the limit in the censorship mechanism, while values higher than the upper limit are considered absent in the truncation mechanism. A similar operation is executed for values smaller than the lower limit. In addition, it can be applied from both the lower and upper at the same time. The details and differences of the censorship and truncation procedure are presented in Yenilmez (2017) with examples. The censorship mentioned here is fixed censorship. There are also cases where the censorship point is random. The fixed censorship point is more common in statistics and econometrics, while the random censorship point is more frequently encountered in medical and engineering data such as survival analysis and reliability analysis. The literature has been developed for econometric analysis and survival analysis. For instance, Cox regression, the Kaplan-Meier estimator, and the accelerated failure time model are often used in survival analysis in the presence of censoring. On the other hand, the Tobit model is often used in the presence of censored response variables from an econometrical perspective. Furthermore, conventional estimators (such as maximum likelihood estimator (MLE), ordinary least-squares estimator (OLSE), etc.) are used in statistical analysis. In reliability analysis, various methods used in survival analysis and statistical tools are utilized. Although the procedures differ in the four main areas mentioned, the purpose is the same; analysing CDVs. The four areas of interest can be detailed. In this study, the literature for survival analysis and econometrical analysis is presented for the sake of brevity.

For clinical and medical studies, literature has been developed for the analysis of censored response variables in the context of survival analysis. Even today, the most commonly used estimators have been introduced for survival analysis in the seminal works of Kaplan-Meier (1958) and Cox (1972). The LR model has been extended for random right-censored response variables by Miller (1976). The inconsistency problem of Miller’s method is solved by Buckley and James (1979). This approach provides estimation without determining residual distribution for analysis of censored response variables. The relationship between random censoring and synthetic data has been investigated by Leurgans (1987) for LR. The accelerated failure time model which is an extended and weighted least-squares estimator is proposed for censored response variables by Stute (1999). In addition to these important studies, Klein and Moeschberger (1997) provide a comprehensive study, including estimators used in survival analysis for censored and truncated data. In addition, the Bayesian approach in survival analysis has been used by Ibrahim et al. (2001). After this important study, the Bayesian approach has been widely used in survival analysis. The quantile regression (QR) for survival analysis has used by Peng and Huang (2008). A comprehensive survey about the use of machine learning methods in survival analysis has been presented by Wang et al. (2019).

A seminal model for the estimation of LDVs is proposed by Tobin (1958) for econometrical studies. The household zero expenditures variables are considered as LDVs by Tobin (1958). The response variable containing extreme zeros can be considered a censored response variable from an econometric perspective. The model is called the Tobit model (Tobin's Probit) by Goldberger (1964) for analysis of censored dependent variables. The 2-PM extension of the Tobit model is investigated by Cragg (1971). Heckman's selection model is presented by Heckman (1979) as an alternative to the Tobit model. Amemiya (1984) classified Tobit models into five different types. The first and the second types of Tobit correspond to the classical Tobit model and Heckman's model, respectively. For the CR model, a Monte Carlo simulation is conducted to compare estimators by Paarsch (1984), thus the simulation setup for CR and determination of censorship levels and points in CR models are presented. Married women's hours of work are investigated with the Tobit model by Mroz (1987). The dataset of this seminal study, which can be called Mroz data, is frequently used in CR applications. The estimation of the Tobit model is required in which error terms are normally distributed. Alternative estimators which are less sensitive to normality assumptions are investigated. Censored least absolute deviations (CLAD) estimator is proposed by Powell (1984). The symmetrically trimmed least-squares (STLS) estimator is proposed as an alternative to the Tobit model by Powell (1986). CLAD and STLS are well-known non-density-based estimators.

ML estimation of the Tobit model and its alternatives are often used in econometrics for fixed CDV. Analysis of random censored response variables depends on the condition that censoring times are independent of event times, or that they are independent conditional on covariates in regression settings (Koenker, 2008). In medical research, the Tobit model is often used for data with fixed censorship as health expenditures, health economics, and health services research, etc., not for the random censorship part of medical studies. However, Gong and Schaubel (2018) propose a Tobit model based on the weighted maximum likelihood for survival times. The inverse probability of the censoring weighting procedure is used to overcome random censoring. In this way, the Tobit model becomes available for random censored response variables.

In order to determine the scope of the research, the analysis of the CDV whose general framework is presented above is examined for fixed censoring. In this study, a generalized censored regression model based on extended normal distribution, which covers a wide range of skew and symmetric distributions, is proposed as an alternative to the censored normal regression model. The proposed generalized censored regression model nests censored normal regression model (Tobit Model) (Tobin, 1958) and the Alpha-power Tobit (APT) (Martínez-Flórez et al., 2013). Moreover, the 2-PM extension of the generalized censored regression model and the ML estimator of the proposed model are presented in the third section for the censored semi-continuous data.

The preliminary for censored regression and the obtaining of the generalized censored regression model are presented in the next subsections.

### 2.3. Censored Regression Model and Estimation

CR is represented as:

$$y_i = \begin{cases} c_L & y_i^* < c_L \\ c_U & y_i^* > c_U \\ y_i^* & otherwise \end{cases} \qquad i = 1,2,\dots,n \qquad (2.2)$$

where $y_i^* = \boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta} + u_i$ and $u_i$ represent random error term assumed to be iid with probability density function given as $f(u|.)$. $x_i$ and $\beta$ denote explanatory variables and unknown parameters, respectively. $c_L$ and $c_U$ are lower and upper bound, respectively. In other words, $c_L$ and $c_U$ can be called the left censorship point and the right censorship point. $y_i$ can be accepted as the data observed under these constraints.

Conventional estimator OLSE is often used in presence of censored response data. OLSE can be obtained from Eq. (2.3) for censored response variables. In the estimation with OLSE, all censored and uncensored response variables (OLSC - OLSE that takes into account censored response variables) or only uncensored response variables (OLSUC - OLSE that takes into account only uncensored response variables) can be used. The OLSUC estimator presented in Eq. (2.4) gives estimates by neglecting the censored response variables. Although the OLSUC estimator seems more problematic, it is emphasized that both OLSC and OLSUC estimator results are biased and inconsistent by Karlsson and Laitila (2014), Moon (1989).

$$argmin_{\beta} \sum_{i=1}^{n} [Y_i^* - max(x_i^T \beta, c_L)]^2 \tag{2.3}$$

$$argmin_{\beta} \sum_{i=1}^{n} [(Y_i^* - x_i^T \beta) 1(y_i^* > c_L)]^2 \tag{2.4}$$

where $\mathbf{1}(.)$ is indicator function.

Another conventional estimator is MLE for censored dependent variables. In fact, the general CR model presented in Eq. (2.2) transforms to the Tobit model or censored normal model when the error distribution is normally distributed $u_i \sim N(0, \sigma^2)$. The MLE of the new model is known as the Tobit estimator (see, e.g., Arabmazar and Schmidt, 1982; Wilson et al., 2020). The ML estimates are also inconsistent when the distribution of error terms is non-normal, therefore, various estimators are proposed as an alternative to the ML estimation of Tobit model in case of violation of the normality assumption (see, e.g., Powell, 1984; Powell, 1986; Acitas et al., 2021; McDonald and Nguyen, 2015; Yenilmez et al., 2018). The likelihood function is written through probability density and cumulative distribution functions of normal distribution. The likelihood function for the MLE to be written depending on the probabilities presented in Eq. (2.5) - (2.6) is handled in Eq. (2.7).

$$P(y_i^* < c_L) = P(y_i = c_L) = \Phi(c_L) \tag{2.5}$$

$$P(y_i^* > c_U) = P(y_i = c_U) = 1 - \Phi(c_U) \tag{2.6}$$

$$L(x; \Theta) = K_1 [\Phi(\, c_L)]^{n_{c_L}} [1 - \Phi(c_U)]^{n_{c_U}} \prod_{i=1}^{n - n_{c_L} - n_{c_U}} \phi(x_i) \tag{2.7}$$

where $n$ is the size of sample, $n_{c_L}$, $n_{c_U}$ and $K_1$ denote the numbers of left-censored observation, numbers of right-censored observation, and ordering constant that do not depend on the parameters, respectively. Eq. (2.7) is presented as a general form for left/right-censored response variables.

Finding the Tobit estimates may not be possible when the function is not concave. The problem of maximizing a non-concave function can be solved by special arrangements. Re-parameterization is proposed by Olsen (1978). The following transformations are used in this operation that solves the problem of maximizing a non-concave function: Location parameter ($\beta$) is written as $\beta = \frac{\delta}{\gamma}$ and scale parameter $\sigma^2$ is re-parameterized as $\sigma^2 = \frac{1}{\gamma^2}$. After the re-parameterization, a global solution is found by arranging likelihood and log-likelihood functions.

It should be noted that the Tobit estimator is inefficient and inconsistent results when the distribution of error terms is not normal (Caudill, 2012). Powell's STLS estimator is used as another alternative to the Tobit estimator. Its results are particularly attractive in case of symmetrical error distributions. For left-censored response variables, STLS estimates are obtained from Eq. (2.8)

$$argmin_{\beta} \sum_{i=1}^{n} \left[ (Y_i^* - \max(\frac{1}{2} Y_i^*, x_i^T \beta) \right]^2 + \sum_{i=1}^{n} I(y_i^* > 2x_i^T \beta) \left[ \left[\frac{1}{2} Y_i^*\right]^2 - \max(c_L, x_i^T \beta)^2 \right] \tag{2.8}$$

It should be emphasized that the results of the STLS estimator are adversely affected by skewed error distributions. STLS estimator gives inefficient results in case of skewed data.

The violation of the distribution assumptions can be associated with outlier(s). Robust estimators are attractive for the estimation of the CR model in this case. Some of the studies in the literature on this subject are as follows: Estimators based on the expectation function or the conditional expectation function for uncensored observations are used in Stapleton and Young (1984) for robustness. A class of high breakdown point estimators for the LR model has been proposed by Salibian-Barrera and Yohai (2008) for censored response variables. Consistent estimators using a bounded loss function have been defined and solved. In the presence of left truncation and right censoring on the observed response variables, M-estimators have been used by Lai and Ying (1994). For this purpose, a general missing information principle is utilized. On the other hand, the robustness of the new Tobit estimators based on distributions is remarkable. For further details, see Arellano-Valle et al. (2012), Garay et al. (2017) and Guzmán et al. (2020).

In the next sub-section, the generalized censored regression is introduced. Moreover, the generalized censored model is used as the base model for the 2P-GCR proposed in Section 3.

### 2.4. The Distribution Assumption for Censored Regression Model

An alternative method used in CR is a quasi-maximum likelihood estimator (Q-MLE) based on various distributions. Q-MLE is also known as a partially adaptive estimator. For censored response variables, Q-MLE based on generalized normal distribution Type I (which is also called exponential power distribution or generalized error distribution) is proposed by Yenilmez (2017a) and examined by Yenilmez and Kantar (2017b) for different simulation scenarios. This estimator is also examined for heteroscedasticity and non-normality by Yenilmez and Kantar (2017c). Q-MLE based on generalized logistic distribution is proposed by Kantar et al. (2017) and examined by Yenilmez et al. (2018) for different factors in simulation. Maximum entropy distributions are used by Usta et al. (2018) in the Q-MLE procedure for the CR model. Q-MLE based on Moyal distribution is considered by Yenilmez et al. (2019a) for censored response variables. An alternative estimation method for parameters of the CR model based on alpha skew logistic distribution is proposed by Yenilmez and Kantar (2019b). The obtaining parameter estimation for the censored dependent variable with Q-MLE is associated with the distributional assumption in CR.

For $y_i^* = \boldsymbol{x_i^T \beta} + u_i$, $u_i$ represent random error terms assumed to be iid with probability density function given as $f(u|.)$. Under these conditions, the expression presented in Eq. (2.7) is generalized by $f_.(.)$ and $F_.(.)$, the basic and general likelihood function framework for Q-MLE is found as:

$$L(x;\theta) = [F_.(c_L)]^{n_{c_L}}[1 - F_.(c_U)]^{n_{c_U}} \prod_{i=1}^{n-n_{c_L}-n_{c_U}} f_.(x_i) \qquad (2.9)$$

The Log-likelihood function for Q-MLE is obtained as:

$$logL(x;\theta) = \sum_{j=1}^{n_{c_L}} \ln F(c_L) + \sum_{k=1}^{n_{c_U}} \ln(1 - F(c_U)) + \sum_{i=1}^{n-n_{c_L}-n_{c_U}} f(x_i) \tag{2.10}$$

The initial two parts are the contribution of the left and right-censored response variables to the likelihood function. The last part is the contribution of uncensored response variables to the likelihood function. An extension of the Tobit estimator based on the power-normal distribution has been proposed by Martínez-Flórez et al. (2013). Student-t error distribution is used by Arellano-Valle et al. (2012) for the CR model. The distribution assumption in CR is handled in-depth by McDonald and Nguyen (2015). The scale mixtures of normal distribution and scale mixtures of skew-normal distribution (some skew distributions) are used for CR by Garay et al. (2017) and Mattos et al. (2018), respectively. In addition, the skew scale mixtures of normal distributions are used for CR by Guzmán et al. (2020). The class of log-symmetric error distribution for the CR model is used and a solution is proposed for problems related to asymmetry and bimodality by Saulo et al. (2021).

In this study, a generalized normal distribution is used to obtain a generalized censored regression model. In the next sub-section, the generalized Tobit model based on the Exponentiated Generalized (EG) family (Cordeiro et al., 2013) is proposed and the ML estimator of the model is obtained.

### 2.4.1. Exponentiated Generalized Tobit Model

#### *2.4.1.1. Exponentiated generalized family and extended normal distribution*

Cumulative distribution function (CDF) of EG family is shown as:

$$F(x) = [1 - [1 - G(x)]^a]^b \tag{2.11}$$

where $a > 0$ and $b > 0$; $G(x)$ represents CDF. $a$ and $b$ are additional shape parameters. The EG family presents wide ranges and various forms of tails such as heavier and lighter thanks to these two additional parameters. Lima et al. (2019) have emphasized that EG can be applied to censored response variables and it can be used in various fields such as engineering and biology thanks to structure of the flexibility of EG.

The probability density function (PDF) is given as:

$$f(x) = ab[1 - G(x)]^{a-1}[1 - [1 - G(x)]^a]^{b-1}g(x) \tag{2.12}$$

If $G(x)$ and $g(x)$ are taken as $\Phi\left(\frac{x-\mu}{\sigma}\right)$ and $\phi\left(\frac{x-\mu}{\sigma}\right)$, respectively, the CDF and PDF of extended normal (EN) distribution is obtained and given as:

$$F_{EN}(x) = \left[1 - \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^a\right]^b \tag{2.13}$$

$$f_{EN}(x) = \frac{ab}{\sigma}\left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{a-1}\left[1 - \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^a\right]^{b-1}\phi\left(\frac{x-\mu}{\sigma}\right) \tag{2.14}$$

For $a = b = 1$, $x_i \sim EN(\mu, \sigma, a, b)$ transform to normal distribution $x_i \sim EN(\mu, \sigma, 1{,}1) = N(\mu, \sigma)$. For $a = 1$, power-normal distribution is obtained $x_i \sim EN(\mu, \sigma, a, b) = PN(\mu, \sigma, \delta)$ where $\delta$ is shape parameter. For $b = 1$, Lehmann type II-G (LTII-G) family is obtained $x_i \sim EN(\mu, \sigma, a, b) = LTII - G(\mu, \sigma, \gamma)$ where $\gamma$ is shape parameter. In addition, Table 2.1 presents the transition between distributions.

**Table 2.2.** *Members of the Extended Normal distribution*

| **Distributions** | **Restricted parameters** | **Probability Density Function** |
|---|---|---|
| **Normal** | $a = b = 1$ | $f_N(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$ |
| **Power-Normal** | $a = 1$ | $f_{PN}(x) = \frac{b}{\sigma}\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{b-1}\phi\left(\frac{x-\mu}{\sigma}\right)$ |
| **Lehmann type II-G** | $b = 1$ | $f_{LTII}(x) = \frac{a}{\sigma}\left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{a-1}\phi\left(\frac{x-\mu}{\sigma}\right)$ |

As seen from Figure 2.1, two shape parameters of the EN distribution provide flexibility. Location and scale parameters are fixed $(\mu, \sigma) = (0, 1)$ in the Figure 2.1. In the next section, exponentiated generalized Tobit (EGT) model based on EN distribution is introduced and its some statistical properties such as the MLE and score vector are presented.

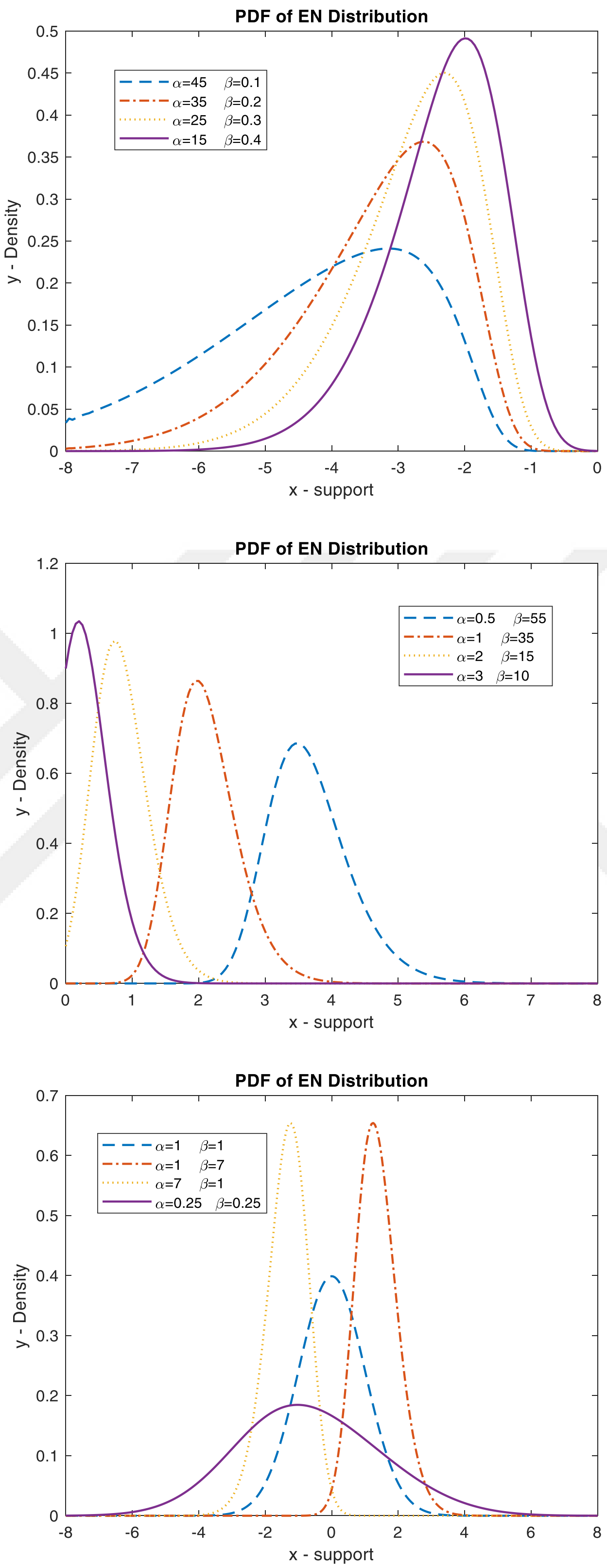



**Figure 2.2.** *PDF of Extended Normal distribution for different shape parameters*

### ***2.4.1.2. Estimation of exponentiated generalized Tobit model***

Exponentiated Generalized Tobit (EGT) model is defined as:

$$y_i = \begin{cases} c_L & y_i^* < c_L \\ c_U & y_i^* > c_U \\ y_i^* = \boldsymbol{x_i^T \beta} + u_i & otherwise \end{cases} \quad i = 1,2,\dots,n \tag{2.15}$$

$\beta$ and $x_i$ denote a $k \times 1$ vector of unknown parameters and $n \times (k+1)$ matrix of independent variables (explanatory - designed matrix), respectively. The EGT is defined based on the EN distribution $\left(u_i \sim EN(\mu, \sigma, a, b)\right)$. Obviously, classical Tobit model is obtained when additional shape parameters are taken as 1 ($a = b = 1$). Moreover, Lehmann type II-G (LTII-G) Tobit estimator based on LTII-G distribution is introduced for $b = 1$. As far as is known, LTII-G Tobit (LTII-GT) model has been used for the first time. EGT also includes Alpha-power Tobit (APT) model for $a = 1$. The proposed models are more flexible than the classical Tobit model thanks to the additional two parameters of the EGT. It makes the EGT an alternative that can be used both in case of violation of the normality assumption and the presence of outlier(s).

### ***2.2.1.3. Maximum likelihood estimation of EGT model***

Likelihood function of EGT is as:

$$L(y;\mu,\sigma,a,b) = [F_{EN}(\,c_L)]^{n_{c_L}}[1 - F_{EN}(c_U)]^{n_{c_U}} \prod_{i=1}^{n-n_{c_L}-n_{c_U}} f_{EN}(y_i) \tag{2.16}$$

The Log-likelihood function is obtained as:

$$logL(y;\mu,\sigma,a,b) = \sum_{j=1}^{n_{c_L}} \ln F_{EN}(\,c_L) + \sum_{k=1}^{n_{c_U}} \ln(1 - F_{EN}(c_U)) + \sum_{i=1}^{n-n_{c_L}-n_{c_U}} f_{EN}(y_i) \tag{2.17}$$

The initial two parts are the contribution of the left and right-censored response variables to the likelihood function. The last part is the contribution of uncensored response variables to the likelihood function.

Interpretations of results on the performance of estimators for singly right-censored and doubly censored response variables have not been presented here for the sake of brevity as they are similar to interpretations of results for singly left-censored. Moreover, the left censoring for fixed censorship is often handled in econometrics. Left censoring can also arise in the laboratory results in the field of medical research. Fixed censorship from the left occurs because values below a certain threshold cannot be observed in laboratory measurements. This situation is also referred to as the *limit of detection* in medical research. For generalization of censored regression, left censoring is handled in this context. The likelihood and log-likelihood function for left-censored regression is written as follows:

$$L(y;\mu,\sigma,a,b)=\prod_{j=1}^{n_{c_L}}F_{EN}(\,c_L)\prod_{i=1}^{n-n_{c_L}}f_{EN}(y_i) \tag{2.18}$$

$$logL(y;\mu,\sigma,a,b)=\sum_{j=1}^{n_{c_L}}\ln F_{EN}(\,c_L)+\sum_{i=1}^{n-n_{c_L}}lnf_{EN}(y_i) \tag{2.19}$$

Within this framework, score vector $u(\Theta)$ corresponding to $\Theta=(\mu,\sigma,a,b)$ is given as follows:

$$\begin{aligned}U(b)&\\&=n_{c_L}(1-(1-\Phi)^a)^b\ln(1-(1-\Phi)^a)\\&+\sum_{i=1}^{n-n_{c_L}}\frac{\ln(1-(1-\Phi_i)^a)b+1}{b}\end{aligned} \tag{2.20}$$

$$\begin{aligned}U(a)&=-n_{c_L}b[1-\Phi]^a\ln([1-\Phi])\,[1-[1-\Phi]^a]^{b-1}\\&+\sum_{i=1}^{n-n_{c_L}}\frac{([1-\Phi_i]^a\,ln([1-\Phi_i]b-[1-\Phi_i])\,a+[1-\Phi_i]^a-1}{([1-\Phi_i]^a-1)a}\end{aligned} \tag{2.21}$$

$$U(\sigma) = -n_{c_L}\frac{ab(y-\mu)\phi(1-(1-\Phi)^a)^{b-1}(1-\Phi)^{a-1}}{\sigma^2}$$
$$+\sum_{i=1}^{n-n_{c_L}}\frac{\phi_i(1-(1-\Phi_i)^a)^{1-b}(1-\Phi_i)^{1-a}\sigma}{ab}$$
$$\times\left(\begin{array}{c} -\frac{ab(1-(1-\Phi_i)^a)^{b-1}(1-\sigma\Phi_i)^{a-1}}{\sigma^2} \\ -\frac{a^2(b-1)b(y_i-\mu)\phi_i(1-(1-\Phi_i)^a)^{b-2}(1-\Phi_i)^{2a-2}}{\sigma^3} \\ +\frac{(a-1)ab(y_i-\mu)\phi_i(1-(1-\Phi_i)^a)^{b-1}(1-\Phi_i)^{a-2}}{\sigma^3}\end{array}\right)$$
$$-\sum_{i=1}^{n-n_{c_L}}\frac{(y_i-\mu)\phi_i'\phi_i ln\left(\frac{ab(1-(1-\Phi_i)^a)^{b-1}(1-\Phi_i)^{a-1}}{\sigma}\right)}{\sigma^2}$$

(2.22)

$$U(\mu) = -\frac{ab\phi\left(\frac{y-\mu}{\sigma}\right)(1-(1-\Phi)^a)^{b-1}(1-\Phi)^{a-1}}{\sigma}$$
$$\sum_{i=1}^{n-n_{c_L}}\left\{\frac{\sigma\phi_i(1-(1-\Phi_i)^a)^{1-b}(1-\Phi_i)^{1-a}}{ab}\right.$$
$$\times\left(\begin{array}{c}\frac{(a-1)ab\phi_i(1-(1-\Phi_i)^a)^{b-1}(1-\Phi_i)^{a-2}}{\sigma^2} \\ -\frac{a^2(b-1)b\phi_i(1-(1-\Phi_i)^a)^{b-2}(1-\Phi_i)^{2a-2}}{\sigma^2}\end{array}\right)$$
$$\left.-\frac{\ln\left(\frac{ab(1-(1-\Phi_i)^a)^{b-1}(1-\Phi_i)^{a-1}}{\sigma}\right)\phi_i{}'\phi_i}{\sigma}\right\}$$

(2.23)

where $\Phi = \Phi\left[\frac{c_L-\mu}{\sigma}\right]$, $\Phi_i = \Phi_i\left[\frac{y_i-\mu}{\sigma}\right]$, $\phi = \phi\left[\frac{c_L-\mu}{\sigma}\right]$ and $\phi_i = \phi_i\left[\frac{y_i-\mu}{\sigma}\right]$. Thus, the components of score vector are found. Estimates of parameters are obtained by solving equations $u(\Theta) = 0$ for all components of parameter space $\Theta = (\mu, \sigma, a, b)$. In addition, the Hessian matrix is obtained as a $4 \times 4$ matrix and presented as follows:

$$U(bb) = n_{c_L}\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^b$$

$$\times\, ln^2\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right) - (n-n_{c_L})\frac{1}{b^2}$$

(2.24)

$$U(ba) = n_{c_L}\frac{\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a ln\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)}{\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1}$$

$$\times\frac{\left(bln\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)+1\right)}{\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1}$$

$$\times\frac{\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^b}{\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a ln\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)}{\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a-1}$$

(2.25)

$$U(b\sigma) = -n_{c_L}\frac{a(y-m)\left(bln\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)+1\right)\phi\left[\frac{c_L-\mu}{\sigma}\right]}{\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)\sigma^2}$$

$$\times\frac{\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^b\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a}{\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)\sigma^2}$$

$$-\sum_{i=1}^{n-n_{c_L}}\frac{a(y_i-\mu)\phi_i\left[\frac{y_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a-1}}{\left(1-\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a\right)\sigma^2}$$

(2.26)

$$U(b\mu) = -n_{c_L}\frac{a\left(bln\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)+1\right)\phi\left[\frac{c_L-\mu}{\sigma}\right]}{\sigma\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)}$$

$$\times \frac{\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^b\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a}{\sigma\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)}$$

$$-\sum_{i=1}^{n-n_{c_L}}\frac{a\phi_i\left[\frac{y_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a-1}}{\sigma\left(1-\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a\right)}$$

(2.27)

$$U(aa)=n_{c_L}\frac{b\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a ln^2\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)}{\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)^2}$$

$$\times\frac{\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^b\left(b\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)}{\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)^2}$$

$$-\sum_{i=1}^{n-n_{c_L}}\frac{\left(b\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a ln\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)\right)a}{\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a-1\right)a}$$

$$-\sum_{i=1}^{n-n_{c_L}}\frac{\left(-ln\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)\right)a}{\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a-1\right)a}$$

$$-\sum_{i=1}^{n-n_{c_L}}\frac{\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a-1}{\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^a-1\right)a}$$

(2.28)

$$U(a\sigma)=-n_{c_L}\frac{b(x-\mu)\phi\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^b}{\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)\sigma^2}$$

$$\times\frac{\left(\left(abln\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)+1\right)\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-aln\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)-1\right)}{\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)\sigma^2}$$

$$\times \frac{\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}}{\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)\sigma^{2}}$$

$$+\sum_{i=1}^{n-n_{c_L}} \frac{(x_i-\mu)\phi_i\left[\frac{y_i-\mu}{\sigma}\right]}{\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]-1\right)\sigma^{2}}$$

$$+\sum_{i=1}^{n-n_{c_L}} \frac{\left(b\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{2a}+1\right)}{\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]-1\right)\sigma^{2}}$$

$$+\sum_{i=1}^{n-n_{c_L}} \frac{\left(\left((a-ab)\ln\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)-b-1\right)\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a}\right)}{\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]-1\right)\sigma^{2}}$$

(2.29)

$$U(a\mu)=-n_{c_L}\frac{b\phi\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}\right)^{b}}{\sigma\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)}$$

$$\times \frac{\left(\left(ab\ln\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)+1\right)\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}-a\ln\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)-1\right)}{\sigma\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)}$$

$$\times \frac{\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}}{\sigma\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)}$$

$$+\sum_{i=1}^{n-n_{c_L}} \frac{\phi_i\left[\frac{y_i-\mu}{\sigma}\right]\left(b\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{2a}\right)}{\sigma\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]-1\right)}$$

$$+\sum_{i=1}^{n-n_{c_L}} \frac{\left(\left((a-ab)\ln\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)-b-1\right)\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a}+1\right)}{\sigma\left(\left(1-\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]\right)^{a}-1\right)^{2}\left(\Phi_i\left[\frac{y_i-\mu}{\sigma}\right]-1\right)}$$

(2.30)

$U(\sigma\sigma)$

$$= n_{c_L}\frac{2ab(x-\mu)\phi\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-1}\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a-1}}{\sigma^3}$$

$$- n_{c_L}\frac{ab(x-\mu)^2\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-1}\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a-1}\Phi\left[\frac{c_L-\mu}{\sigma}\right]}{\sigma^4}$$

$$+ n_{c_L}\frac{a^2(b-1)b(x-\mu)^2\phi^2\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-2}\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{2a-2}}{\sigma^4}$$

$$- n_{c_L}\frac{(a-1)ab(x-\mu)^2\phi^2\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-1}\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a-2}}{\sigma^4}$$

$$+ \sum_{i=1}^{n-n_{c_L}}\frac{1}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)\sigma^2}$$

$$\times\left\{\left(\frac{a(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{a-1}\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^2}\right)\sigma\right.$$

$$+\left(\frac{(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)}{\sigma^2}\right)\sigma$$

$$+\left(\frac{a(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{a-1}}{\sigma^2}\right)\sigma$$

$$+\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a$$

$$\left.\times\left(\frac{(x_i-\mu)\big((ab-1)x_i+(1-ab)\mu\big)\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^2}-\frac{(\mu-x_i)(x_i-m)\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^2}\right)\right\}$$

$$+ \sum_{i=1}^{n-n_{c_L}}\frac{\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)s^2}$$

$$\times\left\{\left(-\frac{(x_i-\mu)^2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a}{\sigma^2}\right)\right.$$

$$+\left(\frac{a(x_i-\mu)^2\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{a-1}}{\sigma^2}\right)$$

$$+\left(\frac{(\mu-x_i)(x_i-\mu)\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^2}\right)\Bigg\}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{(x_i-\mu)\big((1-a)x_i+(a-1)\mu\big)\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)\sigma^4}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left((x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a\right)}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)\sigma^4}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left((\mu-x)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)\sigma^4}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{a(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)\sigma^4}$$

$$\times\Bigg\{\Bigg(\Big((ab-1)x_i+(1-ab)\mu\Big)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]$$

$$+(\mu-x_i)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\Bigg)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{a-1}$$

$$-(x_i-\mu)^2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]\Bigg\}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a+1}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)\sigma^2}$$

$$-\sum_{i=1}^{n-n_{c_L}}\frac{a(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{a-1}}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)\sigma^4}$$

$$\times\Bigg\{\Bigg(\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a+1\Bigg)\sigma$$

$$-\left(\left(\left((x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a+(m-x_i)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)\right.$$

$$-\left(\left(\left((ab-1)x_i+(1-ab)\mu\right)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right.\right.$$

$$\left.\left.+(\mu-x_i)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a\right)$$

$$\left.-\left(\left((1-a)x_i+(a-1)\mu\right)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)\right)\Bigg\}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2\sigma^4}$$

$$\times\Bigg\{\left(\left((x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a+(\mu-x_i)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)$$

$$+\left(\left(\left((ab-1)x_i+(1-ab)\mu\right)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right.\right.$$

$$\left.\left.+(\mu-x_i)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a\right)$$

$$+\left(\left((1-a)x_i+(a-1)\mu\right)c\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+(x_i-\mu)\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)\Bigg\}$$

(2.31)

$$U(\sigma\mu)$$

$$=-n_{c_L}\frac{ab\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-1}\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a-1}\Phi\left[\frac{c_L-\mu}{\sigma}\right](x-m)}{\sigma^3}$$

$$+\frac{a^2(b-1)b\phi^2\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-2}\left(1-s\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{2a-2}(x-\mu)}{\sigma^3}$$

$$-\frac{(a-1)ab\phi^2\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-1}\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a-2}(x-\mu)}{\sigma^3}$$

$$+\frac{ab\phi\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^{b-1}\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^{a-1}}{\sigma^2}$$

$$
-\sum_{i=1}^{n-n_{c_L}} \mu \frac{\Phi_i{}^2\left[\frac{x_i-\mu}{\sigma}\right]}{s^3\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}
$$

$$
\times\left\{\left(\left(\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-ab+1\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}\right.\right.
$$

$$
+\left(-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+ab+a-2\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a+\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-a+1\right)
$$

$$
\left(-\left(-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+ab-1\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}\right.
$$

$$
\left.+\left(4\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-ab-a+2\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+a-1\right)\right\}
$$

$$
-\sum_{i=1}^{n-n_{c_L}} \frac{\mu}{s^3\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}
$$

$$
\times\left\{-\left(\left((1-ab)\phi_i{}^2\left[\frac{x_i-\mu}{\sigma}\right]+\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}\right.\right.
$$

$$
+\left(\left((a-a^2)b+a^2+a-2\right)\phi_i{}^2\left[\frac{x_i-\mu}{\sigma}\right]-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a
$$

$$
\left.\left.+(1-a)\phi_i{}^2\left[\frac{x_i-\mu}{\sigma}\right]+\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]\right)\right\}
$$

$$
-\sum_{i=1}^{n-n_{c_L}} \frac{\Phi_i{}^2\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^3\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}
$$

$$
\times\left\{\left(-x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+(ab-1)x_i\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}\right.
$$

$$
+\left(2x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+2\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+(-ab-a+2)x_i\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a
$$

$$
\left.-x\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+(a-1)x_i\right\}
$$

$$
-\sum_{i=1}^{n-n_{c_L}} \frac{\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^3\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}
$$

$$
\times\Bigg\{\Bigg((1-ab)\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+2x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+2\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]
$$

$$
+(1-ab)x_i\Bigg)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}
$$

$$
+\Bigg((ab+a-2)\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]-4x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-4\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]
$$

$$
+(ab+a-2)x_i\Bigg)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{a}
$$

$$
+(1-a)\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+2x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+2\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]+(1-a)x_i\Bigg\}
$$

$$
-\sum_{i=1}^{n-n_{c_L}} \frac{\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}}{\sigma^3\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}
$$

$$
\times\left((ab-1)x_i{\phi_i}^2\left[\frac{x_i-\mu}{\sigma}\right]+(ab-1)\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]-x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)
$$

$$
-\sum_{i=1}^{n-n_{c_L}} \frac{\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{a}}{\sigma^3\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}
$$

$$
\times\Bigg(\left((a^2-a)b-a^2-a+2\right)x_i\,\phi^2\left(\frac{x-\mu}{\sigma}\right)+(-ab-a+2)\sigma\phi\left(\frac{x-\mu}{\sigma}\right)
$$

$$
+2x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+2\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]\Bigg)
$$

$$
-\sum_{i=1}^{n-n_{c_L}} \frac{(a-1)x_i{\phi_i}^2\left[\frac{x_i-\mu}{\sigma}\right]+(a-1)\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]-x_i\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-\sigma\phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^3\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}
$$

(2.32)

$$U(\mu\mu) = -n_{c_L}\frac{ab\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)^b\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a}{\sigma^2\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi\left[\frac{c_L-\mu}{\sigma}\right]-1\right)^2}$$

$$\times\left\{\left(\left(\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a-1\right)\Phi^2\left[\frac{c_L-\mu}{\sigma}\right]\right)\right.$$

$$-\left(\left(1-\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)$$

$$-\left((1-ab)\phi^2\left[\frac{c_L-\mu}{\sigma}\right]\left(1-\Phi\left[\frac{c_L-\mu}{\sigma}\right]\right)^a\right)$$

$$\left.-\left((a-1)\phi^2\left[\frac{c_L-\mu}{\sigma}\right]\right)\right\}$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{{\Phi_i}^2\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^2\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}$$

$$\times\left(\begin{array}{c}\left(\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-ab+1\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}\\ +\left(-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+ab+a-2\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a+\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-a+1\end{array}\right)$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^2\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}$$

$$\times\left(\begin{array}{c}\left(-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+ab-1\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}\\ +\left(4\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]-ab-a+2\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]+a-1\end{array}\right)$$

$$+\sum_{i=1}^{n-n_{c_L}}\frac{\left((1-ab){\phi_i}^2\left[\frac{x_i-\mu}{\sigma}\right]+\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^{2a}}{\sigma^2\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}$$

$$+\frac{(1-a){\phi_i}^2\left[\frac{x_i-\mu}{\sigma}\right]+\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]}{\sigma^2\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}$$

$$+\sum_{i=1}^{n-n_{c_L}} \frac{\left(\left((a-a^2)b+a^2+a-2\right){\phi_i}^2\left[\frac{x_i-\mu}{\sigma}\right]-2\phi_i'\left[\frac{x_i-\mu}{\sigma}\right]\right)\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a}{\sigma^2\left(\left(1-\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]\right)^a-1\right)^2\left(\Phi_i\left[\frac{x_i-\mu}{\sigma}\right]-1\right)^2}$$

(2.33)

# 3. TWO-PART MODELS FOR CENSORED SEMI-CONTINUOUS DATA

Semi-continuous data can have a complex structure with excess zeros and censoring (Freitas Costa et al., 2021). *Censored semi-continuous* data arise in this case. Modelling the censored semi-continuous data and parameter estimations of the model have some difficulties. The classical methods for semi-continuous data cannot deal with censoring. Censoring can be seen as an additional difficulty. In this study, 2-PMs[3] frequently used for semi-continuous data are modified from two different aspects: CR and QR. Therefore, excess zeros and censoring can be governed in 2-PMs.

Excess zeros from the left and censorship from the right are taken into account in this study. The main reason for this situation is that in the case of both zeros (real zero and censored zero) on the left side, the distinction cannot be made completely. In other words, the distinction between non-censored zero and censored zero is not clear (Boulton and Williford, 2018).

The 2P-GCR and the 2P-FMQRC are proposed and explained in detail for censored semi-continuous data. For the 2P-GCR, the generalized censored regression model is extended to a two-part estimation procedure of Moulton and Halsey (1995). ML estimation of the 2P-GCR is obtained. For the 2P-FMQRC, a generalization of the method proposed by Merlo et al. (2021) is modified for censored response variables. For the ML estimation of the 2P-FMQRC, the ML estimation of the 2P-FMQR based on asymmetric Laplace (AL) distribution has been used in terms of the QR frequentist paradigm in this study. Following Merlo et al. (2021), non-parametric ML estimation is used to overcome the computational difficulties encountered in parameter estimation, and a solution is presented with the EM algorithm.

In the first part of the proposed 2-PMs, the logit model is used, and the generalized Tobit model and QR model are used in the second part as the base models for obtaining the 2P-GCR and the 2P-FMQRC, respectively. In the next sub-sections, the 2P-GCR and the 2P-FMQRC are detailed.

---

[3] The expression *hurdle model* is mostly used as a counterpart for the two-part modelling of count data (Belotti et al., 2015). It has also been mentioned that the *two-part model* represents two-part modelling for semi-continuous data. Following Belotti et al. (2015), the same concept has been used from here on to avoid conceptual confusion in this study.

### 3.1. Two-Part Generalized Censored Regression Model

For the 2P-GCR, the generalized censored regression model proposed in this study is used for right-censored response variables and is modified for non-censored zeros. In this section, the 2P-GCR model is proposed in the scope of the generalized censored model in the presence of non-censored zero inflation.

Cragg (1971) proposes the 2-PM to relax the tail-probability constraint of the Tobit model specification (Chai and Bailey, 2008). The probability density function of outcome $y_i$ under Cragg's two-part specification is expressed as:

$$g(y_i) = p_i I_i + (1 - p_i) f^+(y_i) \tag{3.1}$$

where $I_i = \mathbf{1}(y_i = 0)$, $p_i$ are the probability determining the relative contribution made by the point distribution to the overall mixture distribution and, $f^+$ is density function of positive part. This model is a hierarchical model, in other words, the occurrence of limited relative subject and amount of non-limit outcomes are not estimated, sequentially. In the first part, the probability of decision/participation (according to the data handled by the researcher, the situation represented by binary-dichotomous output variables may change.) value (0 or 1) is investigated with the help of probit, and then the non-zero values model. Truncated normal regression is used in Craggit model for second part.

A generalized 2-PM proposed by Moulton and Halsey (1995) allows for the possibility that some limit response variables are the result of interval censoring from $f^+$. Therefore, an observed zero can be either a realization from the point distribution or a partial observation from $f^+$ with a value lying somewhere in $(0, T)$ for a small pre-specified constant $T$. Moulton and Halsey's model is shown as:

$$g(y_i) = [p_i + (1 - p_i) F^+(T)] I_i + (1 - p_i) f^+(y_i) \tag{3.2}$$

where $f^+$ is a PDF with positive support and $F^+$ is CDF. While $p_i$ represents the probability determining the relative contribution made by the point distribution to the overall mixture distribution in Craggit and, Moulton and Halsey's models, $p_i$ is used for zeros in the Poisson regression framework for count data settings by Saffari and Adnan (2011); Nguyen and Dupuy (2019).

Moulton and Halsey's model is modified for right-censored response variables. Therefore, the modified model accounts for zero-inflated right-censored semi-continuous data. The modified model is shown as:

$$g(y_i) = [p_i + (1 - p_i)F^+(y_i)]I_i + (1 - p_i)f^+(y_i) + (1 - p_i)(1 - F^+(y_i)) \tag{3.3}$$

In Craggit and Moulton and Halsey's models, the occurrence of limit response variables is determined by $p_i$. In this study, the occurrence of limit response variables and the size of the non-limit response variables are determined by following Craggit and Moulton and Halsey's models, however, the right fixed censor point and the number of censored observations are not tied to a probability. This situation is similar to the other proposed model (the 2P-FMQRC). In the next sub-section, ML estimation of the 2P-GCR is presented.

### 3.1.1. ML Estimation for Two-Part Generalized Censored Model

ML estimates of the 2-PM based on the generalized censored regression model are obtained in this section. For the positive part, the generalized censored distribution is used. In other words, $F_{EN}(.)$ and $f_{EN}(.)$ are used instead of $F^+(.)$, $f^+(.)$, respectively.

Notice that for modelling positive values, to match the support of the density and positive part ($R^+$), the logarithmic transformation of the random variable $(t_i)$ is considered. Standard $EN(a, b)$ can be used for sake of brevity ($(\mu, \sigma) = (0,1)$). All results can be extended location-scale extension of $EN(a, b)$ (i.e., $T = \mu + \sigma Z$). In this context, logarithmic transformation of standard $EN(a, b)$ can be obtained ($Z = log(\varepsilon)$), thus the distribution of error term is obtained as the log-extended normal distribution $(\varepsilon_i \sim LEN(a, b))$.

Log-linear model is written as:

$$log(y_i) = \boldsymbol{z}_i'\boldsymbol{\beta} + \varepsilon_i \tag{3.4}$$

where $z_i = (z_{i1}, \dots, z_{im})$ represents the $m$ dimensional set of explanatory variables, $\beta$ is the parameter vector and $\varepsilon_i \sim LEN(a, b)$. However, $E(\varepsilon_i)$ is not equal to zero under $\varepsilon \sim LEN(a, b)$ $(E(\varepsilon_i) \neq 0)$. In other words, for the expectation of the random variable regressed under the error distribution $(\varepsilon_i \sim LEN(a, b))$ is obtained as:

$$E(y_i) \neq \boldsymbol{z}_i'\boldsymbol{\beta} \tag{3.5}$$

A correction is neccesary to slope parameter. New parameter can be defined as:

$$\ddot{\beta}_0 = \beta_0 + E(\varepsilon_i) \tag{3.6}$$

In this case, for $\ddot{\beta} = (\ddot{\beta}_0, \beta_0, \dots, \beta_m)$, expectation is obtained as:

$$E(y_i) = \boldsymbol{z}_i'\ddot{\boldsymbol{\beta}} \tag{3.7}$$

For estimation of the 2P-GCR, contribution of zeros, censored values and others to likelihood function depends on binary random variable $(d_i)$ and its occurrence probability $(p_i)$ is as:

$$d_i = 1(y_i = 0) \tag{3.8}$$

$$p_i = \Pr(y_i = 0) = \Pr(d_i = 1) \tag{3.9}$$

For (3.3), the binary logistic model is obtained as:

$$logit(\, p_i \,) = \boldsymbol{x}_i'\boldsymbol{\gamma} \tag{3.10}$$

where $x_i = (x_{i1}, \dots, x_{im})$ represents the $m$ dimensional set of explanatory variables, $\gamma$ is the parameter vector. Note that for $p_i = 0$ for all $i$, two-part generalized censored model reduces to censored normal model.

The likelihood function is written as:

$$L = [p_i \, + (1 - \, p_i \,)F_{LEN} \, (0)]^{d_i \delta_i} [\, (1 - \, p_i \,) \, f_{LEN}(y_i \,)]^{(1 - \, d_i \,)\delta_i}$$
$$\times \left[(1 - \, p_i \,)\left(1 - \, F_{LEN} \, (y_i \,)\right)\right]^{\,(1 - \, d_i \,)(1 - \, \delta_i \,)} \tag{3.11}$$

where $\delta_i = \mathbf{1}(y_i < c_i)$ and $c_i$ is fixed censoring values possibly specific to each observation.

The log-likelihood function is written as:

$$logL = \sum_{i} d_i \delta_i log\left([p_i + (1 - p_i) F_{LEN}(0)]\right)$$

$$+ (1 - d_i)\delta_i log\big((1 - p_i) f_{LEN}(y_i)\big)$$

$$+(1 - d_i)(1 - \delta_i) log\Big((1 - p_i)\big(1 - F_{LEN}(c_U)\big)\Big)$$

(3.12)

where $c_U$ is right (upper) fixed censored point. Score vector $u(\Theta)$ can be obtained. The components of score vector are found as:

$$U(b) = (1-p)\ln\left(1 - \left(F_{LEN_0} + 1\right)^a\right)\left(1 - \left(F_{LEN_0} + 1\right)^a\right)^b$$

$$-\frac{(p-1)\left(\ln\left(1 - \left(1 - F_{LEN_i}\right)^a\right)b + 1\right)}{b}$$

$$+(p-1)\ln\left(1 - \left(1 - F_{LEN_U}\right)^a\right)\left(1 - \left(1 - F_{LEN_U}\right)^a\right)^b$$

(3.13)

$$U(a) = -b(1-p)\ln\left(F_{LEN_0} + 1\right)\left(F_{LEN_0} + 1\right)^a\left(1 - \left(F_{LEN_0} + 1\right)^a\right)^{b-1}$$

$$-\frac{(p-1)\left(\left(b\left(1 - F_{LEN_i}\right)^a \ln\left(1 - F_{LEN_i}\right) - \ln\left(1 - F_{LEN_i}\right)\right)a + \left(1 - F_{LEN_i}\right)^a - 1\right)}{\left(\left(1 - F_{LEN_i}\right)^a - 1\right)a}$$

$$+ b(1-p)\ln\left(1 - F_{LEN_U}\right)\left(1 - F_{LEN_U}\right)^a\left(1 - \left(1 - F_{LEN_U}\right)^a\right)^{b-1}$$

(3.14)

$$U(\sigma) = \frac{ab\mu(1-p) f_{LEN_0}\left(F_{LEN_0} + 1\right)^{a-1}\left(1 - \left(F_{LEN_0} + 1\right)^a\right)^{b-1}}{\sigma^2}$$

$$+\frac{(p-1)\left(\left(\left(\left(1 - F_{LEN_i}\right)^a - 1\right)F_{LEN_i} - \left(1 - F_{LEN_i}\right)^a + 1\right)\sigma\right)}{\left(\left(1 - F_{LEN_i}\right)^a - 1\right)\left(F_{LEN_i} - 1\right)\sigma^2}$$

$$+\frac{(p-1)\left(\left((x-\mu) f_{LEN_i}\left(1 - F_{LEN_i}\right)^a + (\mu - x) f_{LEN_i}\right)F_{LEN_i}\right)}{\left(\left(1 - F_{LEN_i}\right)^a - 1\right)\left(F_{LEN_i} - 1\right)\sigma^2}$$

$$+\frac{\left(\left((ab-1)x+(1-ab)\mu\right)f_{LEN_i}+(\mu-x)\phi_i\right)\left(1-F_{LEN_i}\right)^a}{(p-1)^{-1}\left(\left(1-F_{LEN_i}\right)^a-1\right)\left(F_{LEN_i}-1\right)\sigma^2}$$

$$+\frac{\left((1-a)x+(a-1)\mu\right)f_{LEN_i}+(x-\mu)f_{LEN_i}}{(p-1)^{-1}\left(\left(1-F_{LEN_i}\right)^a-1\right)\left(F_{LEN_i}-1\right)\sigma^2}$$

$$+\frac{ab(c-\mu)(1-p)f_{LEN_U}\left(1-\left(1-F_{LEN_U}\right)^a\right)^{b-1}\left(1-F_{LEN_U}\right)^{a-1}}{\sigma^2}$$

(3.15)

$$U(\mu)=-\frac{ab(1-p)f_{LEN_0}\left(F_{LEN_0}+1\right)^{a-1}\left(1-\left(F_{LEN_0}+1\right)^a\right)^{b-1}}{\sigma}$$

$$+\frac{(p-1)\left(\left(f_{LEN_i}\left(1-F_{LEN_i}\right)^a-f_{LEN_i}\right)F_{LEN_i}\right)}{\sigma\left(\left(1-F_{LEN_i}\right)^a-1\right)\left(F_{LEN_i}-1\right)}$$

$$+\frac{(p-1)\left(\left((ab-1)f_{LEN_i}-f_{LEN_i}\right)\left(1-F_{LEN_i}\right)^a\right)}{\sigma\left(\left(1-F_{LEN_i}\right)^a-1\right)\left(F_{LEN_i}-1\right)}$$

$$+\frac{(p-1)\left((1-a)f_{LEN_i}+f_{LEN_i}\right)}{\sigma\left(\left(1-F_{LEN_i}\right)^a-1\right)\left(F_{LEN_i}-1\right)}$$

$$+\frac{ab(1-p)f_{LEN_U}\left(1-\left(1-F_{LEN_U}\right)^a\right)^{b-1}\left(1-F_{LEN_U}\right)^{a-1}}{\sigma}$$

(3.16)

where $\Theta$ represents parameters and $F_{LEN_0}=F_{LEN}\left[\frac{\mu}{\sigma}\right]$, $F_{LEN_i}=F_{LEN_i}\left[\frac{x_i-\mu}{\sigma}\right]$, $F_{LEN_U}=F_{LEN_U}\left[\frac{c_U-\mu}{\sigma}\right]$, $f_{LEN_0}=f_{LEN}\left[\frac{\mu}{\sigma}\right]$, $f_{LEN_i}=f_{LEN_i}\left[\frac{x_i-\mu}{\sigma}\right]$, and $f_{LEN_U}=f_{LEN_U}\left[\frac{c_U-\mu}{\sigma}\right]$. Estimates of parameters are obtained by solving equations $u(\Theta)=0$ for all components of parameter space $\Theta=(\gamma,\beta,\sigma,a,b)$. In addition, the Hessian matrix is obtained. These processes are as in the ML estimation of EGT. For further details, see Appendix.

A set of individual-specific random-effects $\boldsymbol{b}_i$, which is used for over-dispersion and problematic regression structure, is included in the model. It is stated that the random-effects can be used in single-level data (Aitkin, 1999). In addition, the process can be reduced to cross-sectional data (Alfo et al., 2017). $\boldsymbol{b}_i = (\boldsymbol{b}_{i0}, \boldsymbol{b}_{i1})$ is added for the 2-PM. The regression model can be described by a slight extension of Eq. (3.9):

$$logit(\, p_i \,) = \boldsymbol{x}_i'\boldsymbol{\gamma} + \boldsymbol{v}_i'\boldsymbol{b}_{i0}$$

where $\boldsymbol{v}_i$ is a subset of $\boldsymbol{x}_i$. The likelihood function has been modified since the random effect is also included in the model. The computational difficulties that occurred are solved by numerical approaches, similar to the process presented in the sub-section.

## 3.2. Two-Part Finite Mixture Quantile Regression Estimation for Censored Semi-continuous Data

In this section, the ML estimation of the 2P-FMQRC is addressed in scope of frequentist paradigm for the censored semi-continuous data. Let $y_i$, $i = 1, \ldots, N$, be a semi-continuous variable for unit $i$ and let $\boldsymbol{b}_i = (\boldsymbol{b}_{i0}, \boldsymbol{b}_{i1})$ be a time-constant, individual-specific, random effects vector having distribution $f_b(\cdot)$ with support $\boldsymbol{\mathcal{B}}$ where $\mathbb{E}[\boldsymbol{b}_i] = 0$ is used for parameter identifiability. The role of the random coefficients $b_i$ is to capture unobserved heterogeneity and within subject dependence. In the 2-PM, the probability distribution of the outcome variable $y_i$ can be written as:

$$f(y_i) = p_i^{d_i}[(1 - p_i)g(h(y_i)|y_i > 0)]^{1-d_i} \qquad (3.17)$$

with $d_i = \mathbf{1}(y_i = 0), \quad p_i = \Pr(y_i = 0) = \Pr(d_i = 1)$

where $d_i$ denotes the occurrence variable for unit $i, \mathbf{1}(\cdot)$ is the indicator function, $g(\cdot)$ is the density function for the positive outcome given that $y_i > 0$ and $h(\cdot)$ is a (monotone) transformation function of $y_i$.

The model is completed by defining the linear predictors for the binary and the positive parts of the model. It is assumed that the spike-at-zero is governed by a binary logistic model such that:

$$\text{logit}(p_i \mid \boldsymbol{s}_i, \boldsymbol{b}_{i0}) = \boldsymbol{s}_i'\boldsymbol{\gamma} + \boldsymbol{c}_i'\boldsymbol{b}_{i0} \qquad (3.18)$$

where $\boldsymbol{s}_i = (s_{i1}, \ldots, s_{im})$ is the $m$ dimensional set of explanatory variables, $\gamma$ is the parameter vector and $c_i$ is a subset of $s_i$.

The minimization of the quantile loss function is associated with the maximization of the likelihood based on the asymmetric Laplace (AL) density in terms of parameter estimates. The positive outcomes are modeled with the quantile regression approach. Following Yu and Moyeed (2001), the AL distribution is used.

The functional form of the AL distribution is written as:

$$g\big(y_i;\mu_i(\tau),\sigma(\tau)\big)=\frac{\tau(1-\tau)}{\sigma(\tau)}\exp\left\{-\rho_\tau\left(\frac{y_i-\mu_i(\tau)}{\sigma(\tau)}\right)\right\}, \tag{3.19}$$

where $\mu_i(\tau)$ represent the $\tau$-th quantile, with $\tau \in (0,1)$, of $y_i$, $\sigma(\tau)>0$ is the scale parameter and $\rho_\tau(\cdot)$ denotes the quantile loss function of Koenker and Bassett (1978):

$$\rho_\tau(u)=u(\tau-\mathbf{1}(u<0)) \tag{3.20}$$

Because the positive values are modeled, matching the support of the AL density requires choosing a monotone transformation of the conditional quantile function of the positive values of $y_i$.

The logarithmic transformation of the positive values of $y_i$ is considered. In particular, it is assumed that for a given $\tau$, conditionally on $b_{i1}$ and after log-transforming the outcome variable, $\tilde{y}_i=\log(y_i)$, the conditional density $g(\tilde{y}_i|y_i>0,x_i,b_{i1})$ in Eq. (3.17) is an AL distribution as given in Eq. (3.19) whose location parameter is defined by the linear model:

$$\mu_i(\tau)=\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)+\boldsymbol{z}_i'\boldsymbol{b}_{i1}(\tau) \tag{3.21}$$

where $z_i$ is a subset of covariates of $x_i$.

Suppose that $y_i$ may be censored and instead $y_i$, $y_i^\star=\min(y_i,c_i)$ is observed and $\delta_i=\mathbf{1}(y_i<c_i)$, where $c_i$ are fixed censoring values possibly specific to each observation. This type of data can be called right-censored semi-continuous data.

For a fixed quantile level $\tau$, the likelihood function can be written as:

$$L(\Phi_\tau)=\prod_{i=1}^{N}\int_{\mathcal{B}}\big(f(y_i^\star)^{\delta_i}Pr(y_i\geq y_i^\star)^{(1-\delta_i)}\big)f_b(\boldsymbol{b}_i)d\boldsymbol{b}_i \tag{3.22}$$

$$L(\Phi_\tau)=\prod_{i=1}^{N}\int_{\mathcal{B}}\big(f(y_i^\star)^{\delta_i}S(y_i^\star)^{(1-\delta_i)}\big)f_b(\boldsymbol{b}_i)d\boldsymbol{b}_i \tag{3.23}$$

where $\Phi_\tau = \{\boldsymbol{\gamma}(\tau), \boldsymbol{\beta}(\tau), \sigma(\tau)\}$ denotes the global set of model parameters, $f(y_i^\star)$ has been defined in Eq. (3.17), $S(y_i^\star)$ is defined as:

$$S(y_i^\star) = (1 - p_i)\Big(1 - F\big(\, y_i^\star \mid \mu_i(\tau), \sigma(\tau)\,\big)\Big) \qquad (3.24)$$

and $F(y_i^\star \mid \mu_i(\tau), \sigma(\tau))$ is the cumulative distribution function of the AL density in Eq. (3.19) for the positive outcomes.

Some computational difficulties are encountered in the solution phase for Eq. (3.22). Calculation of the likelihood function has been an ongoing research topic in the literature for a long time (Gupta and Mehra, 1974; Fair and Taylor, 1983; Rabe et al. 2005; Zhao and Severini 2017). A parametric assumption based on the distribution or simulation methods such as Monte Carlo experiments designs contains conflicting purposes. For distribution-based methods, a large sample means computational difficulty, while for Monte Carlo-based methods, a small sample means convergence problem. In parallel with the procedure followed in Laird (1978), Non-parametric Maximum Likelihood (NPML) estimation was examined in this study. The distribution $f_b(\cdot)$ in Eq. (3.22) was not specified parametrically and an approximation was conducted by using a discrete distribution on $G < N$ locations $\boldsymbol{b}_k(\tau) = (\boldsymbol{b}_{0k}(\tau), \boldsymbol{b}_{1k}(\tau))$. In other words, $b_i(\tau)$is approximated as:

$$\boldsymbol{b}_i(\tau) \sim \sum_{k=1}^{G} \pi_k\,(\tau)\delta_{b_k(\tau)} \qquad (3.25)$$

where the probability $\pi_k(\tau)$ is defined by

$$\pi_k(\tau) = \Pr(b_i(\tau) = b_k(\tau)) \qquad (3.26)$$

$i = 1, \ldots, N$ for $k = 1, \ldots, G$ and $\delta_{b_k(\tau)}$ is a one-point distribution putting a unit mass at $b_k(\tau)$.

The likelihood function in Eq. (3.23) is rewritten as:

$$L(\Phi_\tau) = \prod_{i=1}^{N} \left\{ \sum_{k=1}^{G} \left( p_{ik}^{d_i \delta_i} [(1 - p_{ik}) g_{ik}]^{(1-d_i)\delta_i} S(y_{ik}^\star)^{(1-\delta_i)} \right) \pi_k \right\}$$

(3.27)

where $\Phi_\tau = \{\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_G, \pi_1, \ldots, \pi_G\}$ represent the parameter vector.

For the each $k$-th cluster, the spike-at-zero and positive outcomes are also estimated by the logistic model, $\text{logit}(p_{ik} \mid \boldsymbol{s}_i, \boldsymbol{b}_{0k}) = \boldsymbol{s}_i'\boldsymbol{\gamma} + \boldsymbol{k}_i'\boldsymbol{b}_{0k}$, and linear mixed quantile with AL density in Eq. (3.19) having location parameter given by $\mu_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{b}_{1k}$, respectively.

Eq. (3.27) represents the likelihood function of the 2P-FMQRC models. A finite bivariate mixture model can be estimated based on Eq. (3.27). The model-based clustering allows dividing in $G$ homogeneous sub-populations which differ for the values of the regression parameters. Advantages of finite mixtures are the classification of subjects in $G$ clusters which are characterized by homogeneous values of regression parameters and accommodation of extreme departures from the homogeneous regression model. In this modelling framework, it is crucial to choose the number of mixture components. It should be noted that there are conflicting objectives at this stage. Increasing the number of support points always improves the fit of the model. However, this improvement has to be traded off against the increase in the number of parameters. The choice must be made in such a way as to produce interpretable and meaningful results.

### 3.2.1. ML Estimation Based on the EM Algorithm for 2P-FMQRC

ML estimates based on the EM algorithm are obtained in this section. For the finite mixture representation in Eq. (3.27), the stages are as follows:

- Classification: Each unit $i$ can be conceptualized as drawn from one of $G$ distinct groups: $w_{ik}$ denotes the indicator variable that is equal to 1 if the $i$-th unit belongs to the $k$-th component of the finite mixture, and 0 otherwise.
- Regulation of log-lihelihood function: The component membership $w_{ik}$ is considered missing data and the log-likelihood function is written as:

$$\ell_c(\Phi_\tau) = \sum_{i=1}^{N}\sum_{k=1}^{G} w_{ik}\left\{log\left(p_{ik}^{d_i\delta_i}[(1-p_{ik})g_{ik}]^{(1-d_i)\delta_i}S(y_{ik}^{\star})^{(1-\delta_i)}\right) + log(\pi_k)\right\} \tag{3.28}$$

- Step E: $w_{ik}$ is handled by the conditional expectation of $w_{ik}$ given the observed data and the parameter estimates at the $r$-th iteration $\widehat{\Phi}_\tau^{(r)} =$

$\{\hat{\boldsymbol{\gamma}}^{(r)}, \hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\sigma}}^{(r)}, \hat{\boldsymbol{b}}_1^{(r)}, \dots, \hat{\boldsymbol{b}}_G^{(r)}, \hat{\pi}_1^{(r)}, \dots, \hat{\pi}_G^{(r)}\}$ in the presence of the unobserved group-indicator $w_{ik}$.

At the $(r+1)$-th iteration of the algorithm, $w_{ik}$ is replaced by its conditional expectation $\hat{w}_{ik}^{(r+1)}$ using the following update equation:

$$\hat{w}_{ik}^{(r+1)} = \mathbb{E}\left[w_{ik} \middle| y_{it}, s_i, x_i, \hat{\Phi}_\tau^{(r)}\right] = \frac{f_{ik}^{(r)} \hat{\pi}_k^{(r)}}{\sum_{l=1}^{G} f_{il}^{(r)} \hat{\pi}_l^{(r)}} \tag{3.29}$$

where $f_{ik}^{(r)} = p_{ik}(\hat{\Phi}_\tau^{(r)})^{d_i} \left[\left(1 - p_{ik}(\hat{\Phi}_\tau^{(r)})\right) g_{ik}(\hat{\Phi}_\tau^{(r)})\right]^{1-d_i}$.

- Step M and iteration: Using $\hat{w}_{ik}^{(r+1)}$, solutions for M-step are updated by maximizing $\mathbb{E}[\ell_c(\Phi_\tau) \mid y_i, \boldsymbol{s}_i, \boldsymbol{x}_i, \hat{\Phi}_\tau^{(r)}]$ with respect to $\Phi_\tau$ using numerical optimization techniques.
- Convergence: The E- and M-steps are alternated until convergence (convergence criterion equal to $10^{-5}$).

Convergency to local maximum and computational difficulty problems are overcome by multi-start strategy and closed form of the ML estimator.

In the simulation section, a sub-simulation study is designed to investigate the performance of proposed estimators LTII-GT and EGT against traditional CR estimators through simulation before using the EGT in the two-part estimation.

The main simulation study is conducted to examine the performance of the 2P-GCR and the 2P-FMQRC for different scenarios under determined conditions.

## 4. SIMULATIONS

In this section, the performance of the ML estimators of generalized censored regression models proposed in Section 2 and ML estimators of 2-PMs proposed in Section 3 are investigated.

The LTII-GT and EGT estimators are evaluated against the conventional censored regression estimators.

The 2P-GCR and the 2P-FMQRC are compared with the help of statistical criteria and the finite sample properties are examined.

### 4.1. The Performance Evaluation for ML Estimators of the Proposed Generalized Censored Regression Model: LTII-GT and EGT

The simulation study is designed to investigate the performance of LTII-GT and EGT estimator introduced in Section 2, relative to the OLSC, OLSUC, TSLS, Tobit ML estimator (TOBIT), Olsen's estimator (OLSENT), Alpha-power Tobit estimator (APT). Simulation experiments are conducted to investigate the presence of outliers and non-normal error terms through Monte Carlo scenarios. The effect of non-normal error terms and outlier(s) in the simulation design is handled with different levels of factors such as sample size and censorship level. The presence of the 5% outlier is examined to investigate the effect of outliers in this study. In addition, it is reasonable and fair to choose the normal distribution as the error distribution to compare the estimators proposed in this study as an alternative to traditional estimators such as the Tobit model, which can show its best performance under the assumption of normality. Furthermore, the mixture normal distribution as the error distribution is used to represent non-normality in this study. Thus, violation of the normality assumption and the effect of outlier are examined. The simulation experiments conditions and the levels of the conditions are summarized in Table 4.1.

**Table 4.1.** *Conditions and determined values for evaluation of the performance of generalized censored regression estimators*

| Conditions | Determined values |
|---|---|
| Sample size | $n = (50, 250, 500)$ |
| Distribution of error terms | Normal, Mixture Normal |
| Degree of censoring level | $CL(\%) = (5, 20, 35)$ |
| Presence of outliers | $OL(\%) = (5)$ |

In the simulation, 1000 data sets are generated and analysis and computational operations are conducted on MatLab software.

Censored dependent variable is modeled with the help of the piecewise function. The function is shown as:

$$y_i = \begin{cases} y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i, & if \;\; y_i^* > censoring\ point \\ 0, & otherwise \end{cases} \quad i = 1,\ldots,n \tag{4.1}$$

where $x_i$ is distributed as uniform distribution $U(0,1)$. $\varepsilon_i$ is iid with normal distribution. $\beta_0$ and $\beta_1$ are taken as 0 and 1 without loss of generality, respectively. Since $\beta_0$ can be written as a function of $\beta_1$, the Bias and mean square error (MSE) values of $\beta_1$ are presented for sake of brevity. Bias and MSE values used to compare the estimators are presented.

$$MSE(\hat{Z}) = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{Z}_i - Z\right)^2 \tag{4.2}$$

$$Bias(\hat{Z}) = \left(\frac{1}{1000}\sum_{i=1}^{1000}\hat{Z}_i\right) - Z \tag{4.3}$$

The censoring point is determined so as to obtain different censoring levels (5%, 20%, and 35%). Stages of determining the left censorship point and obtaining the censored data are presented as:

- Sort data,
- Determine the corresponding value (censoring point) according to the censorship percentage in the sorted data
- Replace the values below the determined value with the censoring point.

It should be noted that the complement of the percentage $(1 - percentage\ of\ censoring\ level)$ of censorship is used for determining the right censoring point. For the censoring of the data, the data above the determined right censoring point is replaced with the censoring point. The sorted version of the data is used only to determine the censoring point. In addition, the procedure depending on the true regression coefficients (Paarsch, 1984) can be used for obtaining censored data.

The results obtained for two different distributions within the scope of the simulation scheme presented above are presented in Table 4.2 and Table 4.3, respectively.

While the upper part of the tables (Panel A) does not contain outliers, the lower part (Panel B) contains 5% outliers.

According to Table 4.2 containing the results carried out under the normal distribution;

- For the small sample size, the OLSC estimator is superior for all levels of censorship. For $n = 250$, the OLSC estimator maintains its superiority in the medium sample size.
- Performances of Tobit, APT, LTII-GT, and EGT estimators are close to each other. The APT, LTII-GT, and EGT estimators cannot produce MSE values as small as the Tobit estimator for the small sample size. This situation can be associated with the number of parameters. However, the performance of the APT, LTII-GT, and EGT estimators increases as the number of samples increases. This situation can be explained by the convergence depending on the increase in the number of samples.
- It can be observed that the estimators based on the normal distribution and the generalization of the normal distribution have close results under the normal distribution. This is an expected result.
- TSLS estimator gives better results according to Bias. It can be stated that the TSLS estimator is still a valid alternative in case of the symmetrical error distribution.

- The EGT estimator is often superior in all sample sizes and all censoring levels in presence of outliers. The EGT estimator is followed by APT and LTII-GT estimators.

According to Table 4.3 containing the results carried out under the mixture normal distribution;

- For all factors, the LTII-GT estimator is superior except low censorship $\times$ large sample sizes.
- In case of the absence of outliers, TSLS estimator is superior for low and high censorship levels according to Bias and LTII-GT estimator is better than its competitor for medium censorship level in terms of Bias criteria.

- In the presence of outliers, the LTII-GT estimator is often superior according to the MSE.
- Under the asymmetric distribution, the results of conventional estimators are significantly affected. Generalized estimators can compensate for this situation.

The findings valid for the two scenarios are as follows:

- The generalized censored estimators are more successful than the other conventional estimators.
- As the number of samples increases, MSE decreases, while as the censorship level increases, MSE increases. In addition, when outliers are added to the scenario, the MSE values of all estimators increase.
- In the presence of outliers and non-normal error terms, it is seen that conventional estimators are seriously affected, but generalized estimators can also cope with these situations.

**Table 4.2.** *Simulation results under Normal distribution for evaluation of the performance of estimators*

| | Degree of censoring level | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1$ | 5% | | | | | | 20% | | | | | | 35% | | | | | |
| N(0.1) | n=50 | | n=250 | | n=500 | | n=50 | | n=250 | | n=500 | | n=50 | | n=250 | | n=500 | |
| Panel A | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| OLSC | 0.1561 | 0.2036 | 0.0449 | 0.0437 | 0.0350 | 0.0344 | 0.0297 | 0.2291 | 0.1762 | 0.0648 | 0.2010 | 0.0591 | 0.3465 | 0.2493 | 0.3541 | 0.1510 | 0.3525 | 0.1382 |
| OLSUC | 0.1814 | 0.2475 | 0.1783 | 0.0704 | 0.1640 | 0.0484 | 0.3563 | 0.3239 | 0.3840 | 0.1879 | 0.4039 | 0.1820 | 0.5139 | 0.4409 | 0.5399 | 0.3286 | 0.5333 | 0.3040 |
| TSLS | -0.0060 | 0.2504 | 0.0035 | 0.0453 | -0.0109 | 0.0261 | -0.0182 | 0.2551 | -0.0214 | 0.0494 | 0.0081 | 0.0369 | 0.0135 | 0.2649 | 0.0136 | 0.0608 | 0.0080 | 0.0484 |
| TOBIT | -0.0263 | 0.2504 | -0.0038 | 0.0456 | -0.0118 | 0.0252 | -0.0340 | 0.2633 | -0.0219 | 0.0490 | 0.0080 | 0.0275 | 0.0043 | 0.2899 | 0.0100 | 0.0550 | 0.0087 | 0.0298 |
| OLSENT | -0.0519 | 0.2807 | -0.0100 | 0.0476 | -0.0131 | 0.0266 | -0.0497 | 0.2851 | -0.0293 | 0.0507 | 0.0056 | 0.0278 | -0.0162 | 0.3085 | 0.0066 | 0.0579 | 0.0095 | 0.0306 |
| APT | -0.0340 | 0.2624 | -0.0054 | 0.0464 | -0.0101 | 0.0251 | -0.0507 | 0.2807 | -0.0260 | 0.0511 | 0.0075 | 0.0281 | -0.0164 | 0.3069 | 0.0038 | 0.0575 | 0.0053 | 0.0309 |
| LTII-GT | -0.0308 | 0.2681 | -0.0051 | 0.0463 | -0.0105 | 0.0250 | -0.0428 | 0.2836 | -0.0239 | 0.0510 | 0.0077 | 0.0280 | -0.0053 | 0.3116 | 0.0048 | 0.0568 | 0.0067 | 0.0310 |
| EGT | -0.0333 | 0.2706 | -0.0058 | 0.0465 | -0.0112 | 0.0251 | -0.0450 | 0.2877 | -0.0241 | 0.0505 | 0.0073 | 0.0280 | -0.0026 | 0.3112 | 0.0063 | 0.0575 | 0.0063 | 0.0311 |
| | Outliers Cases | | | | | | | | | | | | | | | | | |
| Panel B | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| OLSC | 0.0554 | 1.0966 | 0.0210 | 0.2396 | 0.0298 | 0.1433 | -0.1252 | 1.1316 | -0.1185 | 0.2734 | -0.1476 | 0.1570 | -0.2685 | 1.4652 | -0.2554 | 0.3329 | -0.3155 | 0.2582 |
| OLSUC | -0.1031 | 1.5150 | -0.1210 | 0.3145 | -0.1732 | 0.2029 | 0.1082 | 1.6118 | 0.1386 | 0.3841 | 0.1051 | 0.2050 | 0.2515 | 2.4208 | 0.2847 | 0.6059 | 0.3070 | 0.3911 |
| TSLS | -0.4249 | 1.9613 | -0.4064 | 0.5077 | -0.4903 | 0.4538 | -0.3870 | 2.1821 | -0.4187 | 0.5528 | -0.4658 | 0.5095 | -0.4030 | 2.6806 | -0.4661 | 0.5676 | -0.4540 | 0.5299 |
| TOBIT | -0.4067 | 1.7407 | -0.3889 | 0.4460 | -0.4600 | 0.3881 | -0.6303 | 2.2980 | -0.7110 | 0.9454 | -0.7589 | 0.8074 | -0.8458 | 3.7864 | -1.0848 | 1.7991 | -1.1011 | 1.6122 |
| OLSENT | 0.1965 | 1.3251 | 0.2849 | 0.8318 | 0.3077 | 0.4253 | 0.1800 | 1.3625 | 0.3526 | 1.1759 | 0.3542 | 0.8504 | 0.2614 | 2.3643 | 0.4129 | 1.2176 | 0.4158 | 1.1956 |
| APT | -0.0512 | 0.3289 | -0.0721 | 0.1813 | -0.8742 | 0.1197 | -0.0853 | 0.3965 | 0.0987 | 0.2011 | -0.0201 | 0.1589 | -0.0351 | 0.4021 | -0.0813 | 0.2134 | 0.0169 | 0.1799 |
| LTII-GT | -0.0719 | 0.3568 | -0.0293 | 0.1911 | -0.0388 | 0.1678 | -0.0964 | 0.5244 | 0.0174 | 0.2346 | -0.0137 | 0.1978 | -0.2145 | 0.5781 | -0.0978 | 0.2712 | -0.0251 | 0.2279 |
| EGT | -0.0691 | 0.3077 | -0.0176 | 0.1699 | 0.0112 | 0.0856 | -0.0581 | 0.3286 | 0.0128 | 0.1857 | -0.0291 | 0.1501 | -0.1789 | 0.3921 | -0.0555 | 0.2098 | 0.0257 | 0.1527 |

**Table 4.3.** *Simulation results under Mixture Normal distribution for evaluation of the performance of estimators*

| | Degree of censoring level | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1$ | 5% | | | | | | 20% | | | | | | 35% | | | | | |
| MixNor | n=50 | | n=250 | | n=500 | | n=50 | | n=250 | | n=500 | | n=50 | | n=250 | | n=500 | |
| Panel A | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| OLSC | 0.1817 | 0.2904 | 0.0533 | 0.0637 | 0.0449 | 0.0378 | 0.3478 | 0.3279 | 0.1757 | 0.0748 | 0.1826 | 0.0640 | -0.0082 | 0.3518 | 0.3464 | 0.1604 | 0.3352 | 0.1361 |
| OLSUC | 0.0241 | 0.2881 | 0.0440 | 0.0556 | 0.0481 | 0.0325 | 0.4059 | 0.6178 | 0.3819 | 0.2219 | 0.3891 | 0.1862 | 0.8437 | 1.3492 | 0.8232 | 0.7985 | 0.8069 | 0.7210 |
| TSLS | -0.0410 | 0.4881 | 0.0127 | 0.0919 | -0.0454 | 0.0531 | 0.0475 | 0.5062 | -0.0317 | 0.0939 | 0.0481 | 0.0695 | 0.0119 | 0.6462 | 0.3090 | 0.0981 | 0.0287 | 0.0727 |
| TOBIT | -0.0413 | 0.4004 | 0.0374 | 0.0688 | 0.0275 | 0.0405 | -0.1966 | 0.4191 | -0.2100 | 0.1080 | -0.2052 | 0.0738 | -0.5328 | 0.8328 | -0.5516 | 0.4125 | -0.5712 | 0.3850 |
| OLSENT | 0.0147 | 0.4013 | 0.1643 | 0.0881 | 0.1622 | 0.0622 | -0.1675 | 0.4179 | -0.1043 | 0.0938 | -0.0786 | 0.0844 | -0.4151 | 0.6085 | -0.2704 | 0.1419 | -0.2599 | 0.1022 |
| APT | 0.0576 | 0.4530 | 0.1697 | 0.1274 | 0.1366 | 0.0913 | 0.0832 | 0.4937 | 0.1395 | 0.1418 | 0.1617 | 0.0982 | -0.3266 | 0.7045 | -0.2238 | 0.2466 | -0.1746 | 0.1749 |
| LTII-GT | -0.0224 | 0.1129 | 0.0100 | 0.0201 | 0.1033 | 0.0140 | -0.2488 | 0.2347 | 0.1216 | 0.0708 | 0.0192 | 0.0567 | 0.0023 | 0.3650 | -0.2130 | 0.0768 | -0.2189 | 0.0630 |
| EGT | -0.0599 | 0.2108 | 0.0911 | 0.0568 | 0.0645 | 0.0306 | -0.0279 | 0.3728 | -0.0400 | 0.0733 | -0.0302 | 0.0603 | -0.3456 | 0.4530 | -0.2468 | 0.1013 | -0.2283 | 0.0796 |
| | Outliers Cases | | | | | | | | | | | | | | | | | |
| Panel B | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| OLSC | -0.2892 | 1.1677 | -0.3518 | 0.5923 | -0.3492 | 0.3096 | -0.1023 | 1.6794 | -0.2083 | 0.7032 | -0.1813 | 0.5136 | 0.0840 | 2.6518 | -0.1411 | 1.3691 | -0.0274 | 0.9774 |
| OLSUC | -0.2634 | 1.3513 | -0.3350 | 0.6118 | -0.3338 | 0.3079 | 0.1973 | 3.0560 | 0.0715 | 0.6295 | 0.1227 | 0.3268 | 0.8205 | 6.7358 | 0.7826 | 1.7604 | 0.7448 | 1.0990 |
| TSLS | -0.3200 | 1.4820 | -0.4771 | 0.9528 | -0.4714 | 0.5356 | -0.3605 | 2.9120 | -0.4539 | 1.0164 | -0.4345 | 0.7063 | -0.2590 | 3.7181 | -0.4014 | 1.7889 | -0.4338 | 0.7685 |
| TOBIT | -0.3721 | 1.5222 | -0.4140 | 0.6884 | -0.4115 | 0.3792 | -0.8948 | 3.5517 | -0.2415 | 0.8448 | -0.2842 | 0.4272 | -1.7717 | 4.2273 | -2.4499 | 1.6180 | -2.5754 | 0.8500 |
| OLSENT | 0.2525 | 1.3636 | 0.4069 | 0.8489 | 0.4587 | 0.2541 | -0.0095 | 1.4160 | 0.1540 | 0.9645 | 0.2036 | 0.8225 | -0.2357 | 1.5109 | -0.9677 | 1.1238 | -0.0820 | 0.9344 |
| APT | 0.0790 | 1.4415 | -0.8006 | 0.2539 | -0.0251 | 0.1338 | 0.0040 | 1.6372 | -0.0198 | 0.6541 | -0.0239 | 0.1839 | -0.3725 | 1.8391 | -0.4603 | 0.9562 | -0.2529 | 0.2004 |
| LTII-GT | -0.0924 | 1.3146 | -0.2024 | 0.1565 | 0.0520 | 0.1126 | -0.2620 | 1.6274 | -0.2943 | 0.1628 | -0.3121 | 0.1374 | -0.6335 | 1.8141 | -0.8090 | 0.2893 | -0.8628 | 0.1587 |
| EGT | -0.0073 | 1.3670 | -0.1520 | 0.1957 | -0.0201 | 0.1127 | -0.1539 | 1.6846 | -0.0235 | 0.2200 | -0.0198 | 0.1405 | -0.3559 | 1.8791 | -0.0510 | 0.2690 | -0.0204 | 0.1502 |

## 4.2. The Performance Evaluation for ML Estimators of the Proposed Two-Part Models for Censored Semi-continuous Data: 2P-GCR and 2P-FMQRC

In this simulation study, to evaluate the finite sample properties and to examine the performance of the proposed models: the 2P-GCR and the 2P-FMQRC, a Monte Carlo experiment is conducted for censored semi-continuous data. The proposed estimators are examined for the case of encountering both censorship and excess zero at cross-sectional data with different Monte Carlo scenarios. In other words, the effect of censoring is investigated in the simulation design dealing with different levels of factors such as distributions of the error term, number of individuals, time, censoring percentile, and quantile level for semi-continuous data. The factors considered in the simulation study and the levels of these factors are summarized in Table 4.

**Table 4.4.** *Conditions and determined values for evaluation of the performance of two-part models estimators*

| Conditions | Determined values |
|---|---|
| Distributions | $\mathcal{N}-\mathcal{N},\ \chi^2-\mathcal{T}$ |
| Sample sizes | $n = (100, 200)$ |
| Degrees of censoring level | $CL(\%) = (5, 20, 35)$ |
| Zero-inflated data percentile | $ZI(\%) = (30)$ |
| Quantile Levels for FMQRM | $\tau = 0.1, 0.5, 0.9$ |

$\mathcal{N}-\mathcal{N}$ represents Scenario-1 and consists of $\epsilon_i \sim \mathcal{N}(0,1)$ and $b_i \sim \mathcal{N}_2(0,\Sigma)$. $\chi^2-\mathcal{T}$ represents Scenario-2 and consists of $\epsilon_i \sim \chi^2(2)$ and $b_i \sim \mathcal{T}_2(0,\Sigma)$. Three explanatory variables $x_{1i} \sim \mathcal{N}(0,1)$, $x_{2i} \sim \mathcal{N}(0,1)$ and $x_{3i} \sim Ber(0.5)$ were generated. Observations were obtained from the following the 2-PM:

$$logit(p_i) \quad = \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i} + b_{i0}, \quad i = 1, \dots, N$$

$$\log(y_i) \mid y_i > 0 = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + b_{i1} + \epsilon_i, \quad i = 1, \dots, N$$

where $\boldsymbol{b}_i = (\boldsymbol{b}_{i0}, \boldsymbol{b}_{i1})$ are possibly correlated random intercepts that capture individual departures from the fixed effects $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ and $\beta = (\beta_1, \beta_2, \beta_3)$, respectively. The true values of the model parameters are assumed as:

$$\gamma = (-1.5, 0.5, -2), \quad \beta = (2, 1, -1)$$

$$\Sigma = \begin{bmatrix} 0.25 & 0.0625 \\ 0.0625 & 0.25 \end{bmatrix}$$

For model parameters, estimation outputs are evaluated in terms of the Bias and the MSE, which are defined in Eq. (4.2) and Eq. (4.3), respectively.

In the simulation, 1000 data sets are generated and analysis and computational operations are conducted on MatLab and RStudio.

The main purpose of this simulation study is to reveal the performance of the introduced method under different conditions such as low, moderate, and heavy censorship; extreme and moderate quantile levels; different sample sizes and different distributional assumptions. It should also be noted that the extreme zeros rate (approximately 30% of the whole data) was kept under control with the true values of the binary model parameters. Corresponding outcomes are given in the following Table 4.5-4.10.

It can be concluded from Tables 4.5-4.10;

For the estimates of 2P-FMQRC when quantile levels compared to each other,

- Generally, under the assumption of the first distribution $\mathcal{N}-\mathcal{N}$, the MSE values are close at different quantile levels for the binary model, while the median level estimates for the positive model give lower MSE values than other estimates obtained for extreme quantiles.
- As with the first distribution assumption, under the assumption of the second distribution $\chi^2-\mathcal{T}$, MSE values are quite close at different quantile levels for the parameter estimates of the binary model. For the first extreme quantile level $\tau = 0.1$, the MSE values obtained for the parameter estimates of the positive part are smaller than the other quantile levels.
- From the first and second results, it can be concluded that the different quantile levels estimates of the binary model are not seriously affected by the distribution assumption, while the estimates based on different quantile levels for positive data are affected. The importance of the estimation based on different quantile levels for the skewed and over-dispersion data is revealed.
- The proposed method is able to estimate the true parameters for both the binary and positive part, correctly.

For comparative results for 2P-GCR and 2P-FMQRC estimates,

- In general, it can be said that according to the Bias, the 2P-GCR estimator is superior to the 2P-FMQRC estimator. While the 2P-FMQRC estimator is superior to the 2P-GCR estimator at different quantile levels especially under the assumption of $\mathcal{N}-\mathcal{N}$ distribution according to the MSE.
- The 2P-GCR estimator gives results close to the results of the 2P-FMQRC estimator due to the increased censorship level and sample size, but the 2P-FMQRC estimations retain their superiority over the 2P-GCR
- Under the assumption of $\chi^2-\mathcal{T}$ distribution, the 2P-FMQRC estimator is often superior to the 2P-GCR estimator at all censorship levels.
- The 2P-FMQRC estimates maintain their superiority over the 2P-GCR estimates in terms of MSE under the assumption of $\chi^2-\mathcal{T}$distribution while the direction of producing low MSE values in the 2P-FMQRC estimates shift. In other words, while it could produce the lowest MSE values in the median quantile under the assumption of $\mathcal{N}-\mathcal{N}$ distribution, this situation shift towards the lower quantile estimates under the assumption of $\chi^2-\mathcal{T}$distribution due to the changing distributional assumption.
- Compared to the first distribution scenario, there is no significant change for the MSE values of the 2P-GCR and the 2P-FMQRC estimators under the assumption of $\chi^2-\mathcal{T}$ distribution. However, the MSE values of both estimators generally increased depending on the assumption of varying distribution.

- As $N$ increases, the MSE values decrease for all estimators.
- As the censoring level increases, the MSE values increase for all estimators.

**Table 4.5.** *Simulation results for n=100, 5% censoring*

| Scenario | $\mathcal{N}-\mathcal{N}$ | | | | | | | | $\chi^2-\mathcal{T}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$-th Quantile | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | |
| Criteria | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| $\gamma_1$ | -2.171 | 0.104 | -4.108 | 0.105 | -1.458 | 0.107 | 0.420 | 0.221 | -2.667 | 0.232 | -2.912 | 0.227 | -1.387 | 0.226 | 0.245 | 1.146 |
| $\gamma_2$ | -1.230 | 0.107 | -1.891 | 0.106 | 0.978 | 0.108 | 0.374 | 0.421 | -8.344 | 0.155 | -2.441 | 0.160 | -6.495 | 0.157 | 0.344 | 1.308 |
| $\gamma_3$ | -4.785 | 0.239 | -5.288 | 0.237 | -0.992 | 0.240 | 0.294 | 0.203 | -9.563 | 0.394 | -10.974 | 0.398 | -9.574 | 0.396 | 0.414 | 1.203 |
| $\beta_1$ | 0.589 | 0.280 | 3.596 | 0.220 | -1.231 | 0.251 | -0.339 | 0.349 | -1.985 | 0.221 | -0.091 | 0.283 | -10.574 | 0.554 | -0.286 | 0.553 |
| $\beta_2$ | 2.670 | 0.221 | 1.251 | 0.185 | 2.270 | 0.205 | 0.397 | 0.224 | -2.321 | 0.191 | -1.395 | 0.276 | -11.652 | 0.338 | 0.407 | 0.431 |
| $\beta_3$ | 3.184 | 0.351 | -1.845 | 0.331 | -1.072 | 0.384 | -0.196 | 0.089 | -7.847 | 0.401 | 1.108 | 0.451 | -11.541 | 0.687 | -0.415 | 0.320 |

0.1, 0.5, and 0.9 represent quantile levels

**Table 4.6.** *Simulation results for n=200, 5% censoring*

| Scenario | $\mathcal{N}-\mathcal{N}$ | | | | | | | | $\chi^2-\mathcal{T}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$-th Quantile | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | |
| Criteria | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| $\gamma_1$ | -1.985 | 0.098 | -3.763 | 0.097 | -1.290 | 0.101 | 0.381 | 0.111 | -1.545 | 0.139 | -2.144 | 0.204 | -7.915 | 0.145 | -0.133 | 0.516 |
| $\gamma_2$ | -0.901 | 0.100 | -1.012 | 0.103 | 0.704 | 0.101 | 0.357 | 0.258 | -3.832 | 0.062 | -2.200 | 0.147 | -5.526 | 0.072 | 0.255 | 0.660 |
| $\gamma_3$ | -3.642 | 0.221 | -3.887 | 0.219 | -0.887 | 0.215 | 0.259 | 0.111 | -1.772 | 0.110 | -3.145 | 0.132 | -2.012 | 0.140 | -0.154 | 0.536 |
| $\beta_1$ | 0.318 | 0.178 | 2.457 | 0.154 | -0.980 | 0.190 | -0.330 | 0.218 | -1.547 | 0.148 | -0.145 | 0.194 | -11.178 | 0.391 | -0.211 | 0.188 |
| $\beta_2$ | 2.011 | 0.168 | 1.109 | 0.127 | 2.015 | 0.149 | 0.323 | 0.195 | -0.635 | 0.135 | -1.116 | 0.168 | 0.184 | 0.225 | 0.554 | 0.336 |
| $\beta_3$ | 2.401 | 0.299 | -1.506 | 0.292 | -0.914 | 0.298 | -0.185 | 0.051 | -5.147 | 0.275 | 1.074 | 0.345 | -3.814 | 0.463 | -0.301 | 0.297 |

0.1, 0.5, and 0.9 represent quantile levels

**Table 4.7.** *Simulation results for n=100, 20% censoring*

| Scenario | $\mathcal{N}-\mathcal{N}$ | | | | | | | | $\chi^2-\mathcal{T}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$-th Quantile | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | |
| Criteria | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| $\gamma_1$ | -1.351 | 0.133 | -3.244 | 0.135 | -1.685 | 0.130 | 0.207 | 0.250 | -2.692 | 0.234 | -4.657 | 0.237 | -2.108 | 0.242 | 0.093 | 1.338 |
| $\gamma_2$ | -4.734 | 0.150 | -4.114 | 0.150 | -3.540 | 0.153 | 1.088 | 1.519 | -2.149 | 0.160 | -9.621 | 0.166 | -6.548 | 0.167 | 1.259 | 3.969 |
| $\gamma_3$ | -2.569 | 0.259 | -4.685 | 0.260 | -1.740 | 0.261 | 0.015 | 0.215 | -8.458 | 0.405 | -11.555 | 0.425 | -9.800 | 0.427 | 0.137 | 1.401 |
| $\beta_1$ | 1.856 | 0.284 | 3.679 | 0.223 | -1.837 | 0.255 | -1.190 | 1.777 | -1.925 | 0.233 | 0.806 | 0.308 | -9.553 | 0.569 | 0.979 | 2.488 |
| $\beta_2$ | 1.477 | 0.234 | 2.662 | 0.192 | 7.540 | 0.223 | 0.366 | 0.245 | -2.923 | 0.208 | 1.136 | 0.283 | -9.857 | 0.357 | 0.478 | 0.512 |
| $\beta_3$ | 7.445 | 0.365 | -0.340 | 0.348 | 0.624 | 0.414 | -0.510 | 0.329 | -6.109 | 0.407 | 3.175 | 0.481 | -13.783 | 0.708 | 0.738 | 0.951 |

0.1, 0.5, and 0.9 represent quantile levels

**Table 4.8.** *Simulation results for n=200, 20% censoring*

| Scenario | $\mathcal{N}-\mathcal{N}$ | | | | | | | | $\chi^2-\mathcal{T}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$-th Quantile | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | |
| Criteria | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| $\gamma_1$ | -0.944 | 0.120 | -2.146 | 0.113 | -1.302 | 0.116 | 0.198 | 0.145 | -7.779 | 0.143 | -6.392 | 0.217 | -7.799 | 0.234 | 0.016 | 0.958 |
| $\gamma_2$ | -3.010 | 0.141 | -3.475 | 0.139 | -2.471 | 0.145 | 1.071 | 1.357 | -4.063 | 0.065 | -2.718 | 0.151 | -3.789 | 0.134 | 11.571 | 2.146 |
| $\gamma_3$ | -1.901 | 0.223 | -3.941 | 0.225 | -1.507 | 0.217 | 0.068 | 0.121 | -1.965 | 0.118 | -10.512 | 0.299 | -1.859 | 0.311 | -0.174 | 0.647 |
| $\beta_1$ | 1.837 | 0.251 | 2.754 | 0.192 | -1.520 | 0.195 | -1.106 | 0.574 | 0.655 | 0.166 | 0.746 | 0.199 | -10.080 | 0.447 | -0.563 | 1.144 |
| $\beta_2$ | 1.098 | 0.217 | 2.146 | 0.180 | 5.510 | 0.214 | 0.245 | 0.204 | -0.404 | 0.150 | 1.024 | 0.172 | -7.985 | 0.298 | 0.352 | 0.403 |
| $\beta_3$ | 5.441 | 0.314 | -0.255 | 0.314 | 0.543 | 0.307 | -0.509 | 0.287 | 3.839 | 0.294 | 3.054 | 0.390 | -6.471 | 0.501 | -0.677 | 0.811 |

0.1, 0.5, and 0.9 represent quantile levels

**Table 4.9.** *Simulation results for n=100, 35% censoring*

| Scenario | $\mathcal{N}-\mathcal{N}$ | | | | | | | | $\chi^2-\mathcal{T}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$-th Quantile | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | |
| Criteria | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| $\gamma_1$ | -3.781 | 0.146 | -2.566 | 0.146 | -1.374 | 0.143 | 0.353 | 0.339 | -3.478 | 0.255 | 0.954 | 0.251 | 0.857 | 0.248 | 0.091 | 1.078 |
| $\gamma_2$ | -4.477 | 0.151 | -4.808 | 0.153 | -1.532 | 0.155 | 1.354 | 2.316 | -9.938 | 0.163 | -4.985 | 0.169 | -5.817 | 0.170 | 1.454 | 5.171 |
| $\gamma_3$ | -5.473 | 0.279 | -2.791 | 0.280 | -1.811 | 0.282 | -0.006 | 0.234 | -10.854 | 0.425 | -9.545 | 0.430 | -9.102 | 0.429 | -0.264 | 1.070 |
| $\beta_1$ | -1.818 | 0.306 | 4.016 | 0.280 | 1.559 | 0.306 | -1.666 | 3.434 | -0.674 | 0.308 | 2.286 | 0.340 | -8.547 | 0.647 | -1.743 | 5.309 |
| $\beta_2$ | -3.013 | 0.259 | 4.407 | 0.227 | -1.923 | 0.239 | 0.363 | 0.277 | -6.327 | 0.222 | 8.471 | 0.298 | -3.431 | 0.398 | 0.624 | 0.781 |
| $\beta_3$ | 0.605 | 0.480 | -1.078 | 0.446 | -10.194 | 0.417 | -0.623 | 0.533 | -3.307 | 0.514 | -0.394 | 0.487 | -16.843 | 0.764 | -0.829 | 1.319 |

0.1, 0.5, and 0.9 represent quantile levels

**Table 4.10.** *Simulation results for n=200, 35% censoring*

| Scenario | $\mathcal{N}-\mathcal{N}$ | | | | | | | | $\chi^2-\mathcal{T}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$-th Quantile | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | | 2P-FMQRC 0.1 | | 2P-FMQRC 0.5 | | 2P-FMQRC 0.9 | | 2P-GCR | |
| Criteria | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| $\gamma_1$ | -2.717 | 0.141 | -0.703 | 0.127 | -0.967 | 0.126 | 0.351 | 0.222 | -7.813 | 0.150 | -6.393 | 0.233 | -8.037 | 0.241 | 0.144 | 0.994 |
| $\gamma_2$ | -1.735 | 0.147 | -0.230 | 0.146 | -0.339 | 0.149 | 1.337 | 2.007 | -4.036 | 0.071 | -2.563 | 0.166 | -5.499 | 0.151 | 1.367 | 2.942 |
| $\gamma_3$ | -6.178 | 0.259 | -3.218 | 0.254 | -3.282 | 0.257 | 0.001 | 0.137 | -2.220 | 0.148 | -6.310 | 0.347 | -1.875 | 0.392 | -0.198 | 0.745 |
| $\beta_1$ | -1.827 | 0.268 | -2.138 | 0.199 | -7.988 | 0.212 | -1.613 | 2.991 | -0.051 | 0.184 | 2.014 | 0.301 | -7.054 | 0.514 | -0.884 | 2.278 |
| $\beta_2$ | -1.557 | 0.247 | 0.806 | 0.197 | -1.744 | 0.227 | 0.265 | 0.251 | -5.288 | 0.158 | 5.938 | 0.247 | -5.029 | 0.316 | 0.352 | 0.578 |
| $\beta_3$ | 0.152 | 0.371 | -7.442 | 0.324 | -7.221 | 0.347 | -0.024 | 0.349 | 0.960 | 0.319 | 5.935 | 0.401 | 10.021 | 0.661 | -0.811 | 1.143 |

0.1, 0.5, and 0.9 represent quantile levels

Figures are presented to make the tables more understandable. For legibility and efficient usability of the area in the figures, 2P-FMQRC for low (0.1) quantile level, 2P-FMQRC for median (0.5) quantile level, 2P-FMQRC for high (0.9) quantile level, and 2P-GCR are written in place of 2P-QR.1, 2P-QR.5, 2P-QR.9, and 2P-CR, respectively. In Figure 4.1-4.2, B1-3 and D1-3 represent $\beta_i$ and $\gamma_i$ coefficients $(i = 1,2,3)$, respectively. When the MSE values are examined from Figure 4.1-4.2, the performance of each estimator can be seen. Therefore, the change in MSE values can be followed for each censorship level. The results of the estimations based on the quantiles (2P-QR.1, 2P-QR.5, and 2P-QR.9) are in harmony despite the varying level of censorship. The estimators, on the other hand, have structural breaks for some coefficient estimates.

The binary part gives close results at different quantile levels for each scenario. However, as the censorship level increase in both the binary and positive parts, the MSE increase. This increase is seen distinctly for the positive part. For the binary part, the increments are more modest.

Figure 4.3-4.6 is presented to make the bias comparison more effective. The low bias values of the 2P-GCR estimator are noteworthy. Moreover, the 2P-GCR does not move sharply when the conditions in the simulation (censorship levels, scenarios, and sample sizes) change, unlike quantile-based estimators. In inspecting the boxplots, it can be seen that the boxplots of the quantile-based estimators are wider than the boxplots of the 2P-GCR.

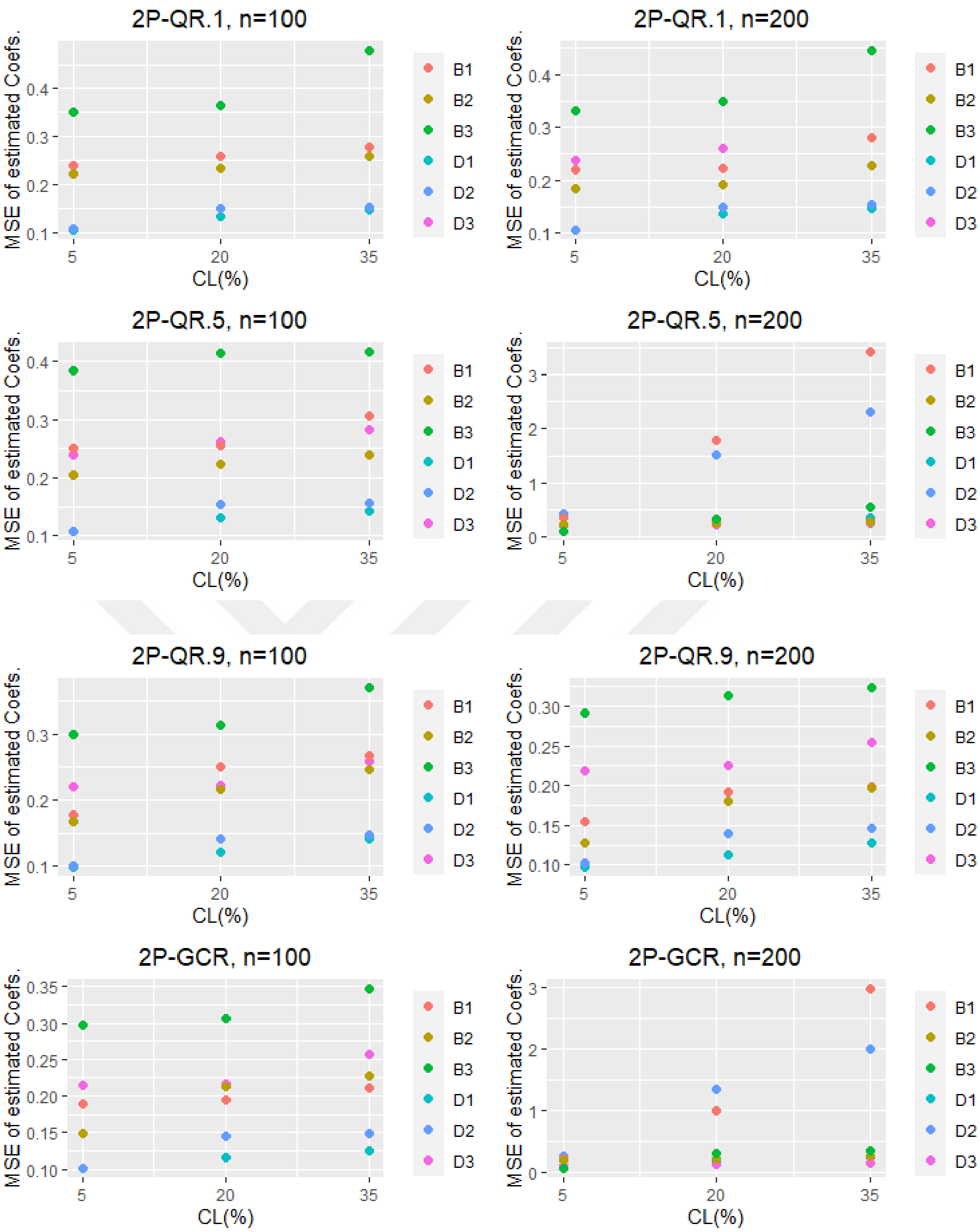



**Figure 4.1.** *Scatter plots of MSE for Scenario-1*

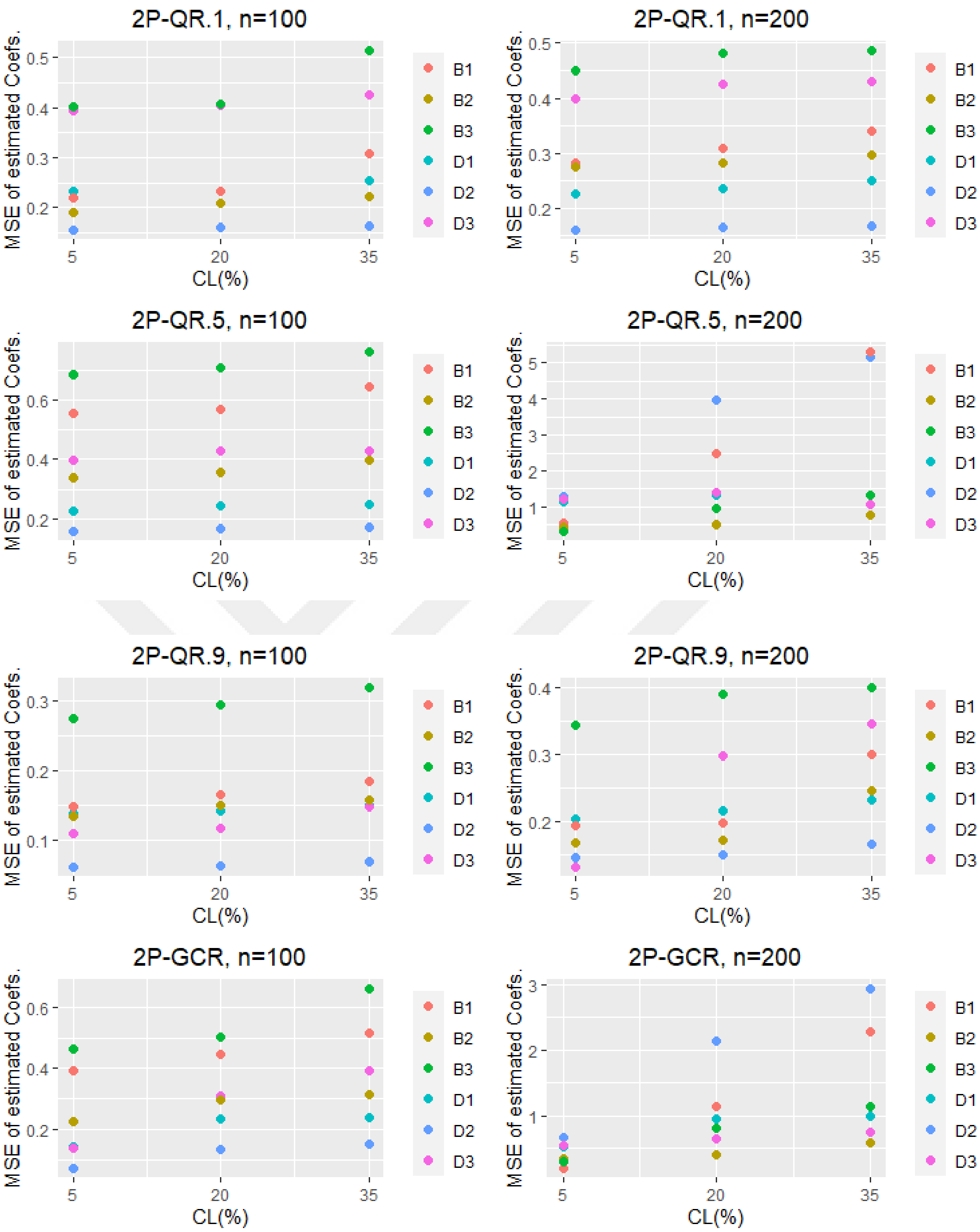



**Figure 4.2.** *Scatter plots of MSE for Scenario-2*

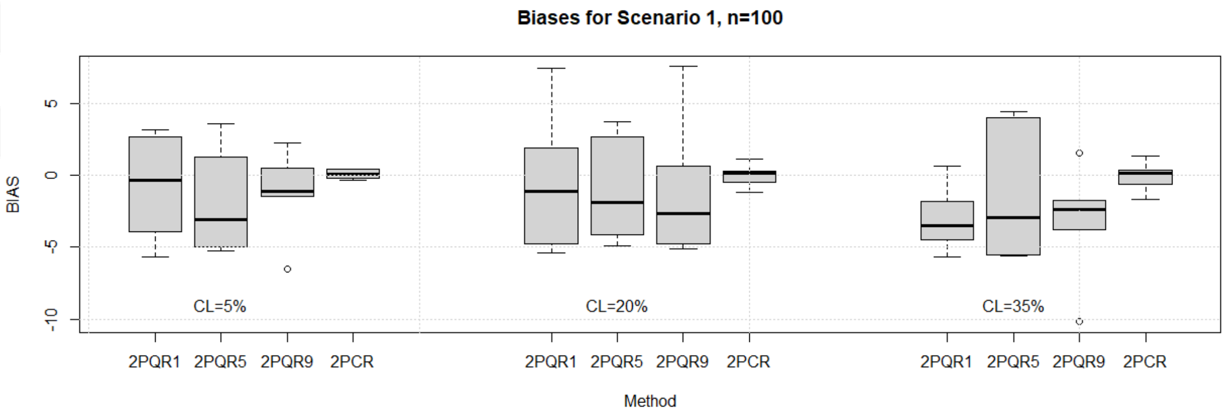



**Figure 4.3.** *Boxplots of Biases for Scenario-1 when n=100*

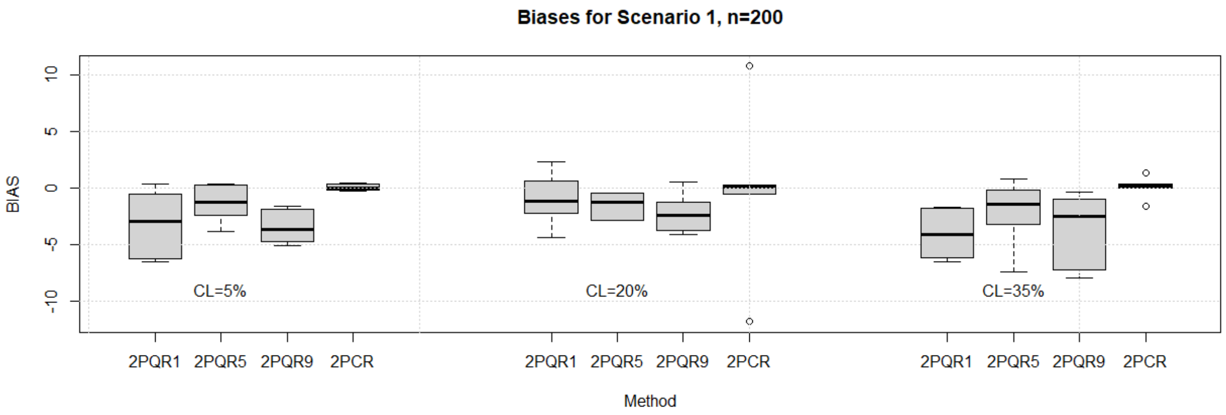



**Figure 4.4.** *Boxplots of Biases for Scenario-1 when n=200*

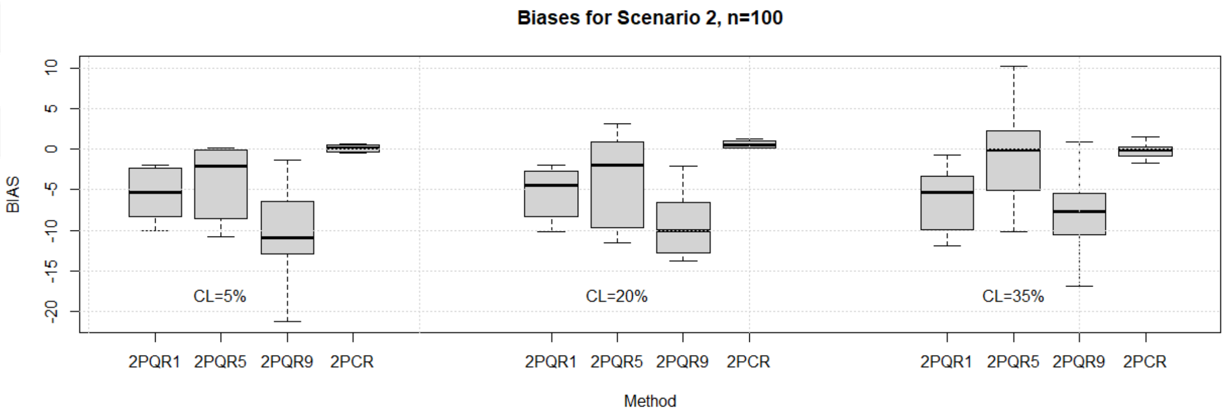



**Figure 4.5.** *Boxplots of Biases for Scenario-2 when n=100*

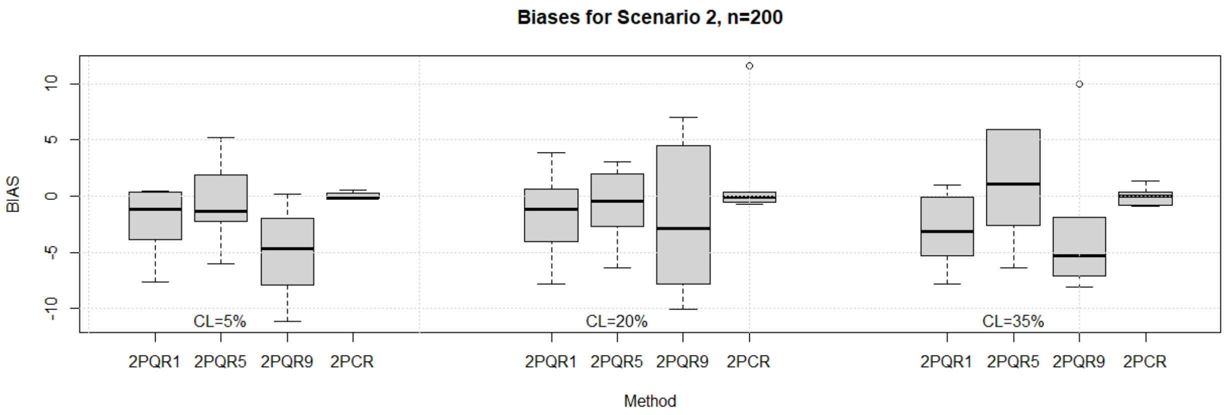



**Figure 4.6.** *Boxplots of Biases for Scenario-2 when n=200*

The overall performance of the estimators based on MSE can be seen from Figure 4.7-4.9. 2P-QR.1($\beta_i$), 2P-QR.5($\beta_i$), 2P-QR.9($\beta_i$), and 2P-CR($\beta_i$) denote the results for coefficient estimations of $\beta_i$ $(i = 1,2,3)$.

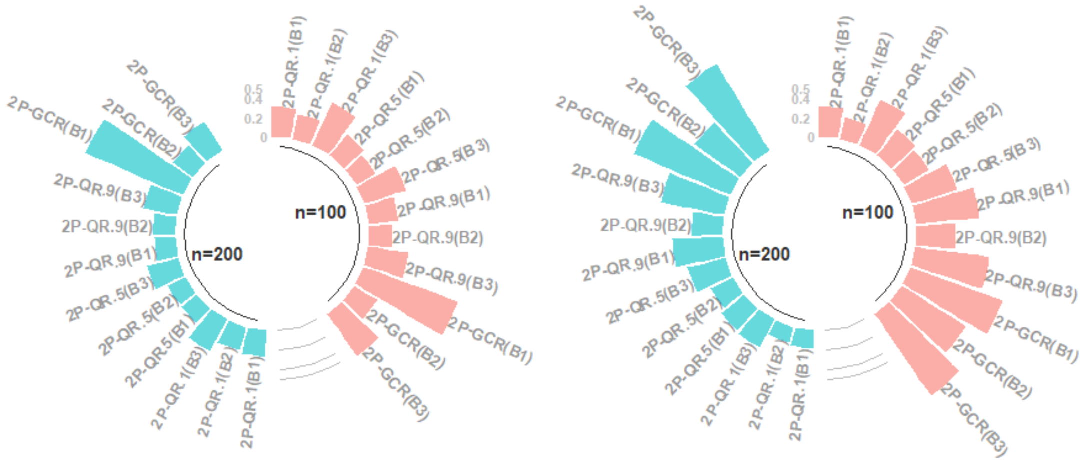
As can be seen from Figure 4.7-4.9, the MSE of the estimators increases as the censorship level increases. In addition, the estimations based on the QR (2P-QR.1, 2P-QR.5, and 2P-QR.9) and the CR (2P-GCR) are in harmony despite the varying level of censorship. A similar situation can be seen in Figure 4.1-4.2. In Figure 4.7-4.9, it can also be seen side by side MSE values for each estimator for different coefficients $\left(\beta_i\ (i = 1,2,3)\right)$. Contrary to the aforementioned harmony, MSE values for each beta obtained from different estimators are out of this harmony. Attention herein should be paid to the data consisting of three parts depending on the changes in the conditions. As a result, it should be noted that each method has the advantage for the different cases.


(a) For Scenario-1: $\mathcal{N} - \mathcal{N}$, (b) For Scenario-2: $\chi^2 - \mathcal{T}$

**Figure 4.7.** *Circular barplots of MSE for $\beta_i$ under 5% CL*


(a) For Scenario-1: $\mathcal{N}-\mathcal{N}$, (b) For Scenario-2: $\chi^2-\mathcal{T}$

**Figure 4.8.** *Circular barplots of MSE for $\beta_i$ under 20% CL*




(a) For Scenario-1: $\mathcal{N}-\mathcal{N}$, (b) For Scenario-2: $\chi^2-\mathcal{T}$

**Figure 4.9.** *Circular barplots of MSE for $\beta_i$ under 35% CL*

## 5. APPLICATIONS

Non-zero right censorship and non-censored zeros can be also observed in the data from different sides. In this context, the right-censored semi-continuous data is handled in the cross-sectional data in this study. Censored semi-continuous data can be encountered in many fields such as employment-related earnings and wages data, health expenditures, insurance, ecological dataset, etc (see, Galvao et al. (2013); Lopez et al. (2016); Liu and Ying (2007); Vock et al. (2011); Lourens et al. (2017); Su and Luo (2017); Lin et al. (2018); Freitas Costa et al. (2021)). When these studies are examined, it is clear that most of the studies in the literature deal with random censorship, and those which deal with fixed censorship use LR methods. In this study, fixed censored semi-continuous data is tackled with the proposed two-part methods (two-part estimator based on censored regression under LR and two-part estimator based on censored modification under QR). In this context, the applications to real-life data related to finance and ecological datasets are conducted to illustrate the performance of the methods.

### 5.1. ICI Data and Export Limiting

The Istanbul Chamber of Industry (ICI) data is used to examine the performance of the proposed methods. Turkey's top 1000 industrial enterprises data is available on the official website of the ICI. Production-based sales are the basic criterion for ranking. Data maintenance operations such as data cleaning and data tidying are required to make the data available for use. When one of three missing mechanisms - the case of missing completely at random (MCAR) also arises, complete case analysis (CCA) is the most commonly used approach to handling missing data (Stavseth et al., 2019). These procedures are followed and also some enterprises are excluded from data for studying with the smooth data in the analysis. After these steps, the data from 2002 for 118 enterprises is examined in this study.

The goal of the study is to understand how available covariates influence enterprises' export (EXP). The considered covariates are price index (PEI), gross value added (in producers' prices) (GVA), equity capital (CAP), the average number of wage workers (EMP), asset turnover ratio (ATR) (to calculate the ATR, revenue is divided by average total assets), and foreign ownership index (FRI). Some possible covariates presented in raw data at the website do not take into account due to multicollinearity. In addition, the applications with artificially censored zero-inflated data are common to show the performance of the proposed method. For further details, see i.e. Saffari and Adnan (2011); Nguyen and Dupuy (2019). In parallel with the literature, the artificial censorship procedure is used to demonstrate the validity of the proposed method. The data are censored from the right (censoring level = 19%). The determined censoring point is 80,000 US Dollars (81st percentile of export data). Export data values greater than the censoring point are also taken as equal to this threshold. While the restriction of exports was widely discussed until the 90s (see i.e. Weinblatt, (1986)), it was not discussed as much as in the recent past due to the commercial reflections of the globalizing world. However, countries have had limited export, especially for medical devices in the Covid-19 pandemic period. Therefore, the restriction of export is brought to the agenda again and it has been recently discussed in the literature (see i.e. Pelc, 2020; Lawrence, 2020). A guideline for export prohibitions and restrictions regarding the Covid-19 pandemic is published by The World Trade Organization (WTO)[4]. Another report is published by the Congressional Research Service (CRS)[5], one of the influential think tanks. In this context, the export analysis is seen as an up-to-date topic and the artificial censorship applied to use ICI data for the proposed method is reasonable considering the Covid-19 crisis.

### 5.1.1. Estimation of Artiffcially Censored ICI Export Data

The descriptive statistics of the data are presented in Table 5.1. In addition, the summary of the censored export (EXPC) values obtained by artificial censorship of exports is also presented in Table 5.1. The tail of the distribution having many occurrences far from the zeros is long with the maximum export being 783,806 US Dollars. The observed long tail, high kurtosis, and skewness indicate the presence of over-dispersion and potential outliers in the data. From Figure 5.1, it can be deduced that the data have spike-at zeros. It should be noted that zero rates of data are also approximately 21%. Zero export can indicate no international business activity, and it may mean that enterprises are working for the domestic market in the relevant period. Figure 5.1 also shows the peak, which is the result of the right censorship. As can be seen in Figure 5.2, the logarithmic transformations of positive and censored positive values do not follow

[4] For more information, visit the https://www.wto.org/english/tratop_e/covid19_e/export_prohibitions_report_e.pdf

[5] For more information, visit the https://crsreports.congress.gov/product/pdf/IF/IF11551

Gaussian distribution. In addition, The piecewise representation of the censored semi-continuous response variable is presented in Eq. (5.1).

$$y_i^{\star} = \begin{cases} y_i^{\star} = 80{,}000 & y_i > 80{,}000 \\ y_i^{\star} = y_i & otherwise \end{cases} \tag{5.1}$$

**Table 5.1.** *Descriptive statistics of ICI data*

| | Mean | St. Dev. | Min | Pct(25) | Pct(50) | Pct(75) | Max |
|---|---|---|---|---|---|---|---|
| **EXP** | 61878.310 | 122108.600 | 0 | 3642.8 | 24642 | 62758.2 | 783806 |
| **EXPC** | 32608.620 | 30784.280 | 0 | 3642.8 | 24642 | 62758.2 | 80000 |
| **PEI** | 84.112 | 4.614 | 75.026 | 81.036 | 83.140 | 87.629 | 93.784 |
| **GVA** | 147495691 | 830102686 | 332029 | 14200000 | 32000000 | 72500000 | 9000000000 |
| **CAP** | 124178390 | 233153082 | 2800000 | 26000000 | 51000000 | 120000000 | 1400000000 |
| **EMP** | 1466.873 | 2562.294 | 46 | 364 | 637 | 1237.5 | 18586 |
| **ATR** | 1.423 | 0.743 | 0.34 | 0.917 | 1.245 | 1.733 | 4.357 |
| **FRI** | 0.134 | 0.263 | 0 | 0 | 0 | 0.1 | 1 |

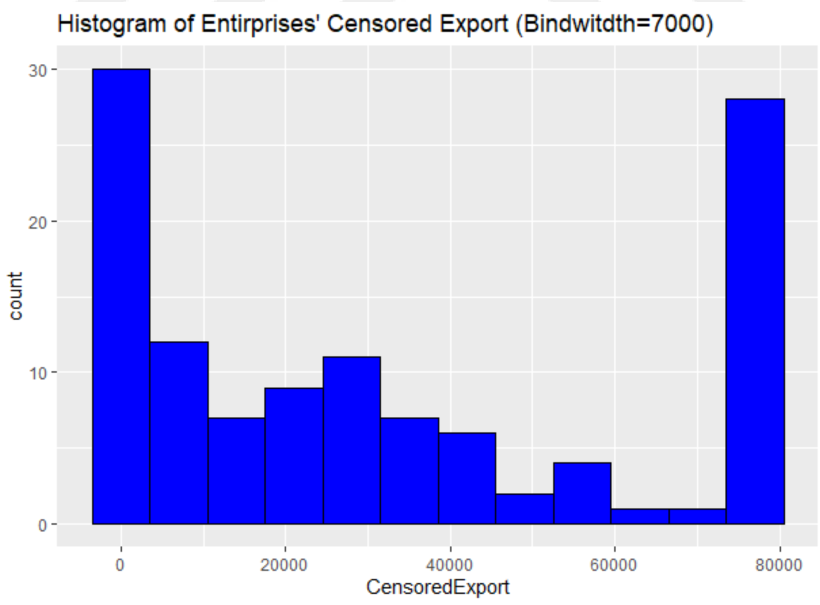



**Figure 5.1.** *Histogram of censored export data*

The right censoring level is approximately 20%. Maximum likelihood estimates of two-part generalized censored regression are obtained. Moreover, two-part finite mixture quantile regression model for censored semi-continuous data at different quantile levels of interest $\tau = (0{:}1;\ 0{:}5;\ 0{:}9)$ is estimated for a varying number of mixture components $G = (2, \ldots, 7)$. A sequence of $K = 100$ values of $\gamma$ for the 10-fold cross-validation procedure is considered. The number of components is unknown a-priori so the number of mixture components is determined according to the Bayesian Information Criterion (BIC) presented in Eq. (5.2).

$$BIC = -2\, log(L) + v_f\, log(N) \tag{5.2}$$

where $v_f$ denotes the number of free model parameters and $N$ is the number of enterprises. 250 bootstrap samples are simulated from the estimated model parameters and the samples are refitted. Standard errors (SEs) and t-statistics are obtained. In the whole process, the covariates for the binary and positive parts are the same and also it is assumed that random intercepts are sufficient to capture heterogeneity between subjects.

The binary and positive parts of the model are estimated and parameter estimates and t -statistics (in parentheses) are presented in Table 5.2. Panel A shows the estimates for the binary part of the model, Panel B shows the estimates for the positive part of the model, and Panel C presents the model fit summary.

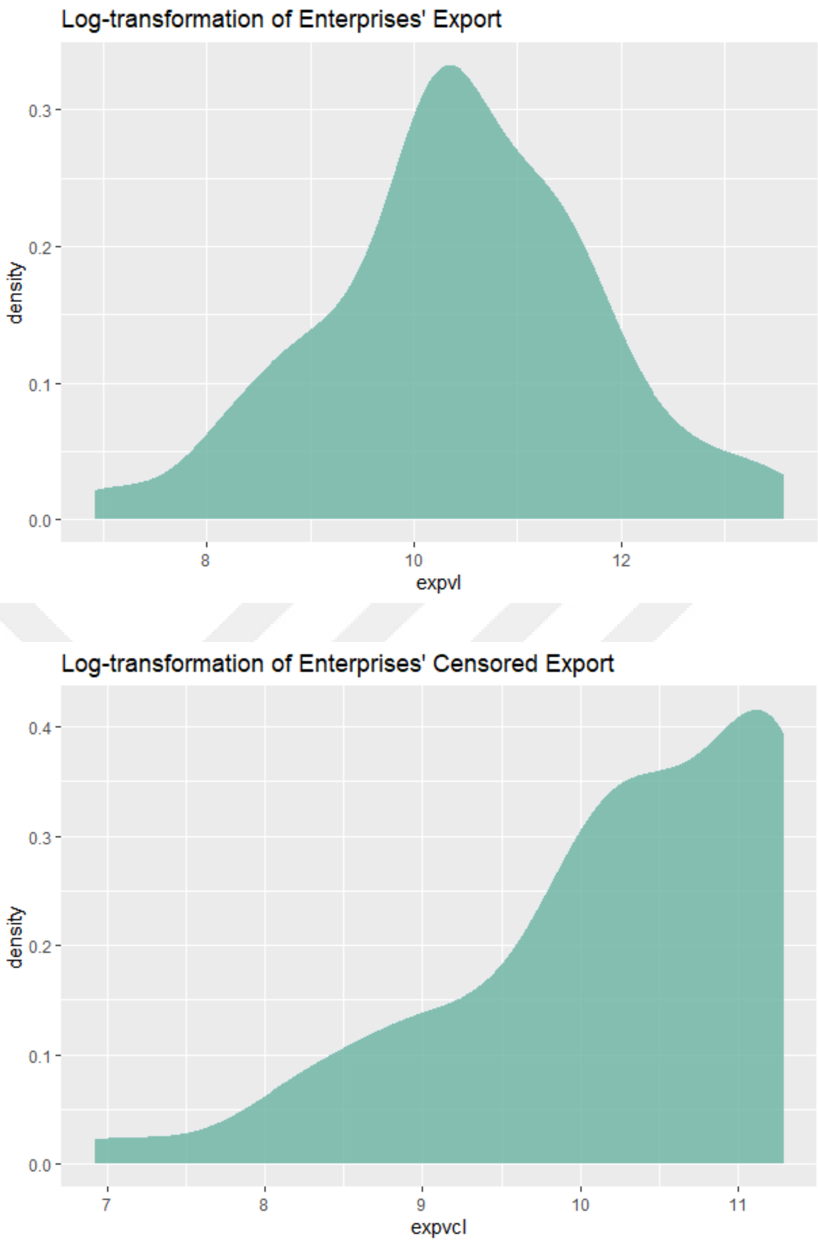



**Figure 5.2.** *Distribution of the logarithm of export values (up) and distribution of the logarithm of censored export values (down)*

**Table 5.2.** *The coefficient estimates of two-part models for ICI data: Estimates for the binary part (Panel A), estimates for the positive part (Panel B), and the model fit summary (Panel C)*

| Estimators | 2P-FMQRC 0.1 | 2P-FMQRC 0.5 | 2P-FMQRC 0.9 | 2P-GCR |
|---|---|---|---|---|
| **Panel A: Binary Estimation** | | | | |
| **PEI** | 0.214* (6.758) | 0.126* (4.512) | 0.197* (5.175) | 0.150* (4.972) |
| **GVA** | 0.125 (0.541) | 0.106* (3.175) | 0.214 (0.954) | 0.147 (0.317) |
| **CAP** | 0.241* (4.156) | 0.201* (3.984) | 0.197* (3.199) | 0.187* (4.536) |
| **EMP** | 0.233* (4.756) | 0.207* (3.544) | 0.192* (3.214) | 0.218* (3.642) |
| **ATR** | 0.350* (4.578) | 0.297 (1.141) | 0.278 (1.415) | 0.283 (1.470) |
| **FRI** | 0.114* (3.045) | 0.152* (2.950) | 0.147* (2.541) | 0.179* (3.470) |
| **Panel B: Positive Part Estimation** | | | | |
| **PEI** | 0.522* (4.548) | 0.547* (4.014) | 0.417* (3.980) | 0.457* (4.105) |
| **GVA** | 0.452 (1.547) | 0.398 (2.144) | 0.449 (1.248) | 0.367 (1.648) |
| **CAP** | 0.541* (4.366) | 0.601* (3.952) | 0.577* (4.149) | 0.412* (3.811) |
| **EMP** | 0.125* (7.145) | 0.152* 10.196) | 0.133* (6.224) | 0.145* (5.132) |
| **ATR** | 0.201* (1.524) | 0.224 (1.147) | 0.224 (1.452) | 0.198 (1.561) |
| **FRI** | 0.203* (2.454) | 0.180* (2.129) | 0.211* (2.457) | 0.229* (2.578) |
| $\boldsymbol{\sigma}$ | 0.081* (14.179) | 0.071* (13.457) | 0.090* (9.781) | 0.120* (4.247) |
| **Panel C: Model fit summary** | | | | |
| $\boldsymbol{logL}$ | -985.244 | -801.468 | -1146.628 | -1219.711 |
| **AIC** | 1982.489 | 1614.936 | 2305.256 | 2451.422 |
| **BIC** | 1999.131 | 1631.579 | 2321.889 | 2558.064 |

*t -statistics in parentheses and* $*: p < 0.05$

The estimated coefficients marked with an asterisk are statistically significant at a level of at least 5%. Moving onto the analysis of the binary outcomes, PEI, CAP, EMP, and FRI are statistically significant determinants of export. In addition, GVA is associated with the export for median (0.5) level in the binary part. According to Table 5.2 Panel B, PEI, CAP, EMP, and FRI are statistically significant determinants of export. ATR is also statistically significant for lower (0.1) quantile level. The results reveal the importance of different quantile levels that allow for the interpretation of coefficients estimates for ICI data.

### 5.2. Wild Boar Dispersal Data

Dispersal accounts for three distinct subpopulations of animals: a segment that will not disperse, i.e., have zero distance; a segment that will disperse measurably (event), and animals whose dispersal is unknown (right-censored) (Freitas Costa et al., 2021). It is emphasized that such dispersal data is unfortunately not fully shared in the literature by Freitas Costa et al., (2021). As a solution to this situation, artificially produced data that mimics a real situation is presented. This data directly fits the scope of this thesis in that it includes zero inflation due to real zeros on the left and censorship due to measurement limitations on the right.

The goal of the application is to understand how covariates influence wild boar dispersal (DIST). The considered covariates are age of wild boars (AGE) and gender of wild boars (SEX). In addition, ZERO and DELTA presented in the data set are indicators for zeros and observed values in the data, respectively. Descriptive statistics of the relevant data are presented in Table 5.3. The histogram of wild boar data is presented in Figure 5.3. The spike-at zeros can be seen in Figure 5.3. The zero rate of data is approximately 15%. The zero rates of data can be also deduced from the mean value of the ZERO in Table 5.3. Zero distance can indicate no movement. Figure 5.3 also shows the other peak, which is the result of the right censorship and represents the limitation of measurement. The right censoring is approximately 10% and this ratio can be also deduced from the complement of the mean value of the DELTA in Table 5.3.

**Table 5.3.** *Descriptive statistics of wild boar data*

| | N | Mean | St. Dev. | Min | Pct(25) | Pct(50) | Pct(75) | Max |
|---|---|---|---|---|---|---|---|---|
| **DIST** | 400 | 8.643 | 9.741 | 0 | 1.153 | 4.299 | 13.451 | 30.000 |
| **ZERO** | 400 | 0.155 | 0.362 | 0 | 0 | 0 | 0 | 1 |
| **DELTA** | 400 | 0.902 | 0.297 | 0 | 1 | 1 | 1 | 1 |
| **AGE** | 400 | 13.874 | 14.742 | 0.07 | 3.216 | 8.796 | 18.435 | 77.808 |
| **SEX** | 400 | 0.5 | 0.501 | 0 | 0 | 0.5 | 1 | 1 |

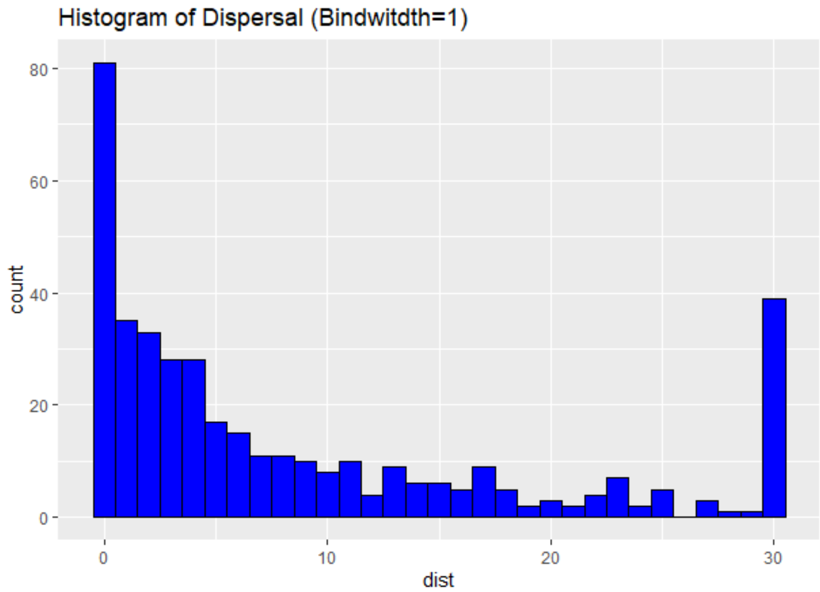



**Figure 5.3.** *Histogram of wild boar data*

The piecewise representation is presented in Eq. (5.3).

$$y_i^\star = \begin{cases} y_i^\star = 30 & y_i > 30 \\ y_i^\star = y_i & otherwise \end{cases}$$

(5.3)

An estimation procedure parallel to that of the first application is used. Maximum likelihood estimates of the 2P-GCR is obtained. Moreover, the 2P-FMQRC estimates for different quantile levels of interest $\tau = (0{:}1;\ 0{:}5;\ 0{:}9)$ is obtained. The maximum likelihood estimates for the binary (Panel A) and positive (Panel B) parts are presented and also $t$ -statistics (in parentheses) are presented in Table 5.4.

**Table 5.4.** *The coefficient estimates of two-part models for wild boar data: Estimates for the binary part (Panel A), estimates for the positive part (Panel B), and the model fit summary (Panel C)*

| $\tau$-th Quantile | 2P-FMQRC 0.1 | 2P-FMQRC 0.5 | 2P-FMQRC 0.9 | 2P-GCR |
|---|---|---|---|---|
| **Panel A: Binary Estimation** | | | | |
| **Age** | 0.211* (6.356) | 0.195* (8.143) | 0.203* (10.146) | 0.257* (4.199) |
| **Sex** | 0.149* (7.457) | 0.164* (11.734) | 0.153* (6.553) | 0.294* (16.243) |
| **Panel B: Positive Part Estimation** | | | | |
| **Age** | 0,114* (5.413) | 0.412* (3.179) | 0.614 (4.169) | 0.314* (8.146) |
| **Sex** | 0.453* (8.744) | 0.539* (10.048) | 0,360* (6.479) | 0.115* (3.689) |
| $\boldsymbol{\sigma}$ | 0.038* (6.304) | 0.045* (3.048) | 0.012* (11.154) | 0.051* (2.259) |
| **Panel C: Model fit summary** | | | | |
| $\boldsymbol{logL}$ | -744.4893 | -623.4194 | -937.735 | -1182.920 |
| **AIC** | 1546.979 | 1304.839 | 1927.470 | 2422.849 |
| **BIC** | 1662.731 | 1420.591 | 2031.248 | 2517.671 |

*t -statistics in parentheses and* $*: p < 0.05$

By looking at the binary estimates in Table 5.4 Panel A, one can see that all coefficients are statistically different from zero at the 5% significance level. In the general framework, coefficients have the expected signs and significance. The results are generally consistent with those of Freitas Costa et al., (2021). It can be concluded from Table 5.4 Panel B, AGE and SEX are statistically different from zero at the 5% significance level except for high quantile (0.9) for AGE. When Table 5.4 Panel C is examined, it is seen that the estimators are close in terms of model fit, in particular, the consistency between the quantile estimators is remarkable. However, the median (0.5) quantile level estimation provides the best fit.

# 6. CONCLUSIONS

The spike-at-zeros have been discussed in the literature for a long time. Tobin has paid attention to zeros for such as expenditures of durable goods and the Tobit estimator is developed by Tobin (1958) for zero household expenditures. Tobin has treated all zeros as *limited* in his seminal study. This is a substantial development considering that zeros were often not even included in the analysis until then.

Non-negative continuous data with clumping at one point (often zero) can be expressed as a mixture of discrete and continuous data and called semi-continuous data (Buuren, 2018). Tobit model is the first model that deals with semi-continuous data (Min and Agresti, 2002; Gottard et al., 2013). However, single equation methods such as Tobit model for data with the mass point at zero may not be appropriate, and the 2-PM provides a method to account for the mass of zeros (Belotti et al., 2015).

Non-zero censorship and non-censored zeros can be also observed in the data from different sides. In this context, the zero inflation and right censorship in the Poisson regression framework for cross-sectional count data are examined by Saffari and Adnan (2011); Nguyen and Dupuy (2019). However, the non-zero censoring can be seen as a literature gap for semi-continuous data. In order to fill the aforementioned research gap, two new models associated with 2-PMs are proposed in this study. Moreover, ML estimators of the proposed models are obtained. The two proposed models are based on CR and QR, respectively.

In order to be more flexible for changing distribution assumptions and to be robust to outliers, a generalized censored regression model based on EN distribution has been proposed for the two-part censored regression model. In other words, the single equation Tobit model used in the extreme zero cases has been firstly generalized. The simulation results show that the generalized censored regression model is a flexible model and it is preferable to the traditional models used in the presence of censored response variables. The maximum likelihood estimation of the generalized censored regression model is superior to the traditional estimators in terms of MSE and Bias. The generalized normal distribution used has four parameters that provide this flexibility. The distribution used in the proposed generalized censored regression also nests important distributions such as the standard normal and power-normal distributions for specific values therefore it includes models such as the censored normal regression (Tobit model) and the Alpha-power Tobit (APT) model.

The censored normal regression model (Tobit model) is used for censored response variables and Tobit model is used under the normality assumption. Therefore, the use of the generalized censored regression model in the 2-PMs framework can take advantage in the case of zero-inflated right-censored continuous (censored semi-continuous) dependent variables. The QR is based on different quantile levels, not based on the mean, therefore, the QR can be used for modelling non-Gaussian, skewed, and censored response variables. The adaptability to real zeros and censored variables of QR can take advantage in the censored semi-continuous data. In this context, the QR in the 2-PMs framework is considered.

As a result, the generalized censored regression model is adapted to the two-part estimator (2P-GCR). In addition, the 2P-FMQR model is adapted to censored regression (2P-FMQRC) in this study. The proposed two-part estimators have been compared. The ML estimation of the 2P-GCR model based on LR gives satisfactory results in terms of Bias, and occasionally also presents preferable results in terms of MSE. On the other hand, QR-based 2P-FMQRC ML estimates stand out as a preferable estimator in terms of MSE for different simulation scenarios despite the inflations at the data. Applications for real-life data in addition to simulation results support the findings. The non-parametric ML method is used in the ML estimation procedure. Calculation difficulties are also solved with the EM algorithm.

From perspectives of types of regression analysis, the two-part mean regression estimation is represented by 2P-GCR estimation and the two-part quantile regression is represented by 2P-FMQRC. As a result of the comparison of these two estimators, it is revealed that the estimation based on quantiles is superior in general. The results highlight the importance of considering a quantile regression approach because such effects could not be detected by the classical mean regression. However, the low bias of the 2P-GCR estimator in the simulation results is noteworthy. It should be noted that although the order of goodness of fit has not changed in two different applications, it is recommended to test both estimators by considering the left and right inflations and general distribution of the data to be investigated. In addition, the study draws attention to conceptual confusion for types of zeros and models for zeros. In particular, basic information in the literature is reviewed and presented to the researchers about types of zeros and the selection of models according to the types of zeros. The importance of the distinction between non-censored zeros and censored zeros is also highlighted in this study.

In future studies, the estimation will be applied to longitudinal data. Thus, the effect of censorship will be investigated in semi-continuous data for repeated measurements. Censored response variables occurrence with probability $q$ can be modeled by a random process, as can the zeros on the left side change randomly with probability $p$. Such an estimation, which would require a 3-part estimation, will be studied.

## REFERENCES


Acitas, S., Yenilmez, I., Senoglu, B., & Mert Kantar, Y. (2021). Modified Maximum Likelihood Estimator Under the Jones and Faddy's Skew t-error Distribution for Censored Regression Model. *Journal of Applied Statistics., 48*(12), 2136-2151. doi:https://doi.org/10.1080/02664763.2020.1786673

Aitkin, M. (1999). A General Maximum Likelihood Analysis of Overdispersion in Generalized Linear Models. *Stat. Comput., 6*, 127-130.

Alfo , M., Salvati, N., & Ranallli, M. G. (2017). Finite Mixtures of Quantile and M-Quantile Regression Models. *Statistics and Computing, 27*, 547-570.

Amemiya, T. (1984). Tobit Models: A Survey. *Journal of Econometrics, 24*(1–2), 3-61. doi:https://doi.org/10.1016/0304-4076(84)90074-5

Amore, M. D., & Murtinu, S. (2021). Tobit Models in Strategy Research: Critical Issues and Applications. *Global Strategy Journal, 11*, 331-335. doi:https://doi.org/10.1002/gsj.1363

Arabmazar, A., & Schmidt, P. (1982). An Investigation of the Robustness of the Tobit Estimator to Nonnormality. *Econometrica, 50*, 1055–1063. doi:https://doi.org/10.2307/1912776

Arellano-Valle, R. B., Castro, L. M., González-Farías, G., & et al. (2012). Student-t Censored Regression Model: Properties and Inference. *Stat Methods Appl., 21*, 453-473. doi:https://doi.org/10.1007/s10260-012-0199-y

Belasco, E. J., & Ghosh, S. K. (2012). Modelling Semi-Continuous Data Using Mixture Regression Models with an Application to Cattle Production Yields. *The Journal of Agricultural Science, 150*(1), 109-121. doi: https://doi.org/10.1017/S0021859611000608

Belotti, F., Deb, P., Manning, W. G., & Norton, E. C. (2015). Twopm: Two-Part Models. *The Stata Journal, 15*(1), 3–20. doi:https://doi.org/10.1177/1536867X1501500102

Biswas, J., Ghosh, P., & Das, K. (2020). A Semi-Parametric Quantile Regression Approach to Zero-Inflated and Incomplate Longitudinal Outcomes. *AStA*

*Advances in Statistical Analysis, 104*, 1-23. doi:https://doi.org/10.1007/s10182-020-00362-9

Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., & Castells, E. (2019). What Does A Zero Mean? Understanding False, Random and Structural Zeros in Ecology. *Methods Ecol Evol., 10*(7), 949– 959. doi:https://doi.org/10.1111/2041-210X.13185

Boulton, A. J., & Williford, A. (2018). Analyzing Skewed Continuous Outcomes with Many Zeros: A Tutorial for Social Work and Youth Prevention Science Researchers. *Journal of the Society for Social Work and Research, 9*(4). doi:https://doi.org/10.1086/701235

Buckley, J., & James, I. (1979). Linear Regression with Censored Data. *Biometrika, 66*(3), 429-436. doi:https://doi.org/10.1093/biomet/66.3.429

Burke, W. J. (2009). Fitting and Interpreting Cragg's Tobit Alternative using Stata. *The Stata Journal, 9*(4), 584–592. doi:https://doi.org/10.1177/1536867X0900900405

Buuren, S. V. (2018). *Flexible Imputation of Missing Data.* Boca Raton, FL: Chapman and Hall/CRC. doi:https://doi.org/10.1201/9780429492259

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data.* Cambridge: Cambridge University Press.

Caudill, S. B. (2012). A Partially Adaptive Estimator for The Censored Regression Model Based on A Mixture of Normal Distributions. *Stat. Methods Appl., 21*, 121-137. doi:https://doi.org/10.1007/s10260-011-0182-z

Chai, H. S., & Bailey, K. R. (2008). Use of Log-Skew-Normal Distribution in Analysis of Continuous Data with a Discrete Component at Zero. *Statistics in Medicine, 27*(18), 3643-3655. doi: https://doi.org/10.1002/sim.3210

Cordeiro, G. M., Cunha, D. C., & Ortega, E. M. (2013). The Exponentiated Generalized Class of Distributions. *J. Data Sci., 11*(1), 777–803.

Cox, D. R. (1972). Regression Models and Life-Tables. *J R Stat Soc [B], 34*(2), 187-202.

Cragg, J. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica, 39*(5), 829–844.

Davidson, R., & MacKinnon, J. G. (2003). *Econometric Theory*. New York: USA: Oxford University Press.

Dow, W. H., & Norton, E. C. (2003). Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions. *Health Services & Outcomes Research Methodology, 4*, 5-18. doi:https://doi.org/10.1023/A:1025827426320

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1983). A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business & Economic Statistics, 1*(2), 115–126.

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1984). Choosing between the sample-selection model and the multi-part model. *Journal of Business and Economic Statistics, 2*, 283–289.

Eggers, J. (2015). *On Statistical Methods For Zero-Inflated Models.* Department of Mathematics, Uppsala University, Sweden.

Fair, R., & Taylor, J. B. (1983). Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models. *Econometrica, 51*(4), 1169-1185. doi:10.2307/1912057

Feng, C. A. (2021). A Comparison of Zero-inflated and Hurdle Models for Modeling Zero-inflated Count Data. *J Stat Distrib App, 8*(8). doi:https://doi.org/10.1186/s40488-021-00121-4

Freitas Costa, E., Schneider, S., Carlotto, G. B., & al., e. (2021). Zero-inflated-censored Weibull and Gamma Regression Models to Estimate Wild Boar Population Dispersal Distance. *Jpn J Stat Data Sci*. doi:https://doi.org/10.1007/s42081-021-00124-0

Galvao, A. F., Lamarche, C., & Lima, L. R. (2013). Estimation of Censored Quantile Regression for Panel Data with Fixed Effects. *Journal of the American Statistical Association, 108*(503), 1075-1089. doi:https://doi.org/10.1080/01621459.2013.818002

Garay , A., Lachos, V., Bolfarine, H., & et al. (2017). Linear Censored Regression Models With Scale Mixtures of Normal Distributions. *Statistical Papers., 58*, 247-278. doi:https://doi.org/10.1007/s00362-015-0696-9

Goldberger, A. S. (1964). *Econometric Theory.* New York: Wiley.

Gong , Q., & Schaubel, D. E. (2018). Tobit Regression for Modeling Mean Survival Time Using Data Subject to Multiple Sources of Censoring. *Pharm Stat, 17*(2), 117–125. doi: https://doi.org/10.1002/pst.1844

Gottard, A., Stanghellini, E., & Capobianco, R. (2013). Semicontinuous Regression Models with Skew Distributions. *Complex Models and Computational Methods in Statistics*.

Greene, H. W. (2003). *Econometric Analysis.* New Jersey: Prentise Hall.

Greene, H. W. (2008). *Econometric Analysis.* New Jersey: Prentise Hall.

Greene, H. W. (2011). *Econometric Analysis.* New Jersey: Prentise Hall.

Greene, W. H. (1994). *Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models.* NYU.

Gupta, N., & Mehra, R. (1974). Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations. *IEEE Transactions on Automatic Control, 19*(6). doi:10.1109/TAC.1974.1100714

Guzmán, D. C., Ferreira, C. S., & Zeller, C. B. (2020). Linear Censored Regression Models with Skew Scale Mixtures of Normal Distributions. *Journal of Applied Statistics.* doi:https://doi.org/10.1080/02664763.2020.1795814

Hay, J. W., & Olsen, R. J. (1984). Let Them Eat Cake: A Note on Comparing Alternative Models of the Demand for Medical Care. *Journal of Business & Economic Statistics, 2*(3), 279-282.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153–161.

Humphreys, B. R. (2013). Dealing With Zeros in Economic Data.

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian Survival Analysis.* New York: Springer.

Johnston, J., & Dinardo, J. E. (1997). *Econometric Methods.* New York: McGraw Hill.

Jones, A. M. (2000). *Health Econometrics, in Handbook of Health Economics.* North Holland, Amsterdam: Elsevier.

Kantar, Y. M., Yenilmez, I., & Acitas, S. (2017). Estimation Based on Generalized Logistic Distribution for the Censored Regression Model. *Proceeding of the International Workshop on Mathematical Methods in Engineering.* Ankara, Turkey: Cankaya Universty Press.

Kaplan , E. L., & Meier, P. (1958). Nonparametric Estimation From Incomplete Observations. *Journal of the American Statistical Association, 53*(282), 457-481.

Karlsson, M., & Laitila , T. (2014). Finite Mixture Modeling of Censored Regression Models. *Stat. Papers 55*, 627–642.

King, C., & Song, J. J. (2019). A Bayesian Two-part Quantile Regression Model for Count Data with Excess Zeros. *Statistical Modelling, 19*(6), 653-673.

Klein, J. P., & Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data.* New York: Springer-Verlag.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press.

Koenker, R. (2008). Censored Quantile Regression Redux. *Journal of Statistical Software, 27*(6), 1-25. doi:https://doi.org/10.18637/jss.v027.i06

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica: Journal of the Econometric Society, 46*(1).

Koenker, R., & Machado, J. A. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association, 94*(448), 1296-1310.

Kozumi, H., & Kobayashi, G. (2011). Gibbs Sampling Methods for Bayesian Quantile Regression. *Journal of Statistical Computation and Simulation, 81*(11), 1565-1578.

Lai, T. L., & Ying, Z. (1994). A Missing Information Principle and M-estimator in Regression Analysis with Censored and Truncated Data. *Ann. Statist, 22*(3), 1222–1225. doi:https://doi.org/10.1214/aos/1176325627

Laird, N. (1978). Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. *Journal of the American Statistical Association, 73*(364), 805-811.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics, 34*(1), 1-14.

Lawrence, J. (2020). *Covid-19, Export Restrictions, and the WTO: Magnifying Global Divisions in a Time of Crisis.* University of Essex.

Leung, S. F., & Yu, S. (1996). On the Choice Between Sample Selection and Two-part Models. *Journal of Econometrics, 72*(1-2), 197-229.

Leurgans, S. (1987). Linear Models, Random Censoring and Synthetic Data. *Biometrika, 74*(2), 301-309. doi:https://doi.org/10.1093/biomet/74.2.301

Lima, M. C., Cordeiro, G. M., Ortega, E. M., & et al. (2019). A New Extended Normal Regression Model: Simulations and Applications. *J Stat Distrib App., 6*(7). doi:https://doi.org/10.1186/s40488-019-0098-y

Lin, T. I., Lachos, V. H., & Wang, W. L. (2018). Multivariate Longitudinal Data Analysis with Censored and Intermittent Missing Responses. *Statistics in Medicine, 37*(19), 2822-2835. doi:https://doi.org/10.1002/sim.7692

Liu, M., & Ying, Z. (2007). Joint Analysis of Longitudinal Data with Informative Right Censoring. *Biometrics, 63*(2), 363-371. doi: https://doi.org/10.1111/j.1541-0420.2007.00708.x

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks: London: Sage Publications, International Educational and Professional Publisher.

Lopez, O., Milhaud, X., & Thérond, P. E. (2016). Tree-based Censored Regression with Applications in Insurance. *Electronic Journal of Statistics, 10*(2), 2685-2716. doi:https://doi.org/10.1214/16-EJS1189

Lourens, S., Zhang, Y., Long, J. D., & Paulsen, J. S. (2017). Analysis of Longitudinal Censored Semicontinuous Data with Application to the Study of Executive Dysfunction: The Towers Task. *Statistical Methods in Medical Research, 26*(2), 865-879. doi:https://doi.org/10.1177/0962280214560187

Maddala, I. B. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge, UK: Cambridge University Press.

Manning, W. G. (1987). Monte carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics, 35*, 59–82.

Martínez-Flórez , G., Bolfarine, H., & Gómez, H. W. (2013). The Alpha-Power Tobit Model. *Communications in Statistics - Theory and Methods, 42*(4), 633-643. doi:https://doi.org/10.1080/03610926.2011.630770

Martín-Fernández, J. A., Barceló-Vidal, C., & Pawlowsky-Glahn, V. (2003). Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Mathematical Geology, 35*, 253–278. doi:https://doi.org/10.1023/A:1023866030544

Mattos, T. B., Garay, A. M., & Lachos, V. H. (2018). Likelihood Based Inference for Censored Linear Regression Models with Scale Mixtures of Skew-Normal Distributions. *Journal of Applied Statistics, 45*(11), 2039-2066. doi:https://doi.org/10.1080/02664763.2017.1408788

McDonald, J. B., & Nguyen, H. (2015). Heteroscedasticity and Distributional Assumptions in the Censored Regression Model. *Communications in Statistics - Simulation and Computation, 44*(8), 2151-2168. doi:https://doi.org/10.1080/03610918.2013.851217

Merlo, L., Maruotti, A., & Petrella, L. (2021). Two-Part Quantile Regression Models for Semi-Continuous Longitudinal Data: A Finite Mixture Approach. *Statistical Modelling*. doi:https://doi.org/10.1177/1471082X21993603

Miller, R. G. (1976). Least Squares Regression with Censored Data. *Biometrika, 63*(3), 449-464. doi:https://doi.org/10.1093/biomet/63.3.449

Min, Y., & Agresti, A. (2002). Modeling Nonnegative Data with Clumping at Zero: A Survey. *Journal of the Iranian Statistical Society, 1*(1-2), 7-33.

Moon, C. G. (1989). A Monte Carlo Comparison of Semiparametric Tobit Estimators. *J. Appl. Econ., 4*(4), 361-382. doi: https://doi.org/10.1002/jae.3950040405

Moulton, L. H., & Halsey, N. A. (1995). A Mixture Model with Detection Limits for Regression Analyses of Antibody Response to Vaccine. *Biometrics, 51*(4), 1570-1578. doi:https://doi.org/10.2307/2533289

Mroz, T. (1987). The Sensitivity of An Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica, 55*(4), 765-799.

Mullahy, J. (1986). Specification and Testing of Some Modified Count Data Models. *Journal of Econometrics, 33*(3), 341 – 365. doi:https://doi.org/10.1016/0304-4076(86)90002-3

Neelon, B., O'Malley, A. J., & Smith, V. A. (2016a). Modeling Zero-Modified Count and Semicontinuous Data in Health Services Research Part 1: Background and Overview. *Statist. Med., 35*(27), 5070-5093. doi: https://doi.org/10.1002/sim.7050

Neelon, B., O'Malley, A. J., & Smith, V. A. (2016b). Modeling Zero-Modified Count and Semicontinuous Data in Health Services Research Part 2: Case Studies. *Statist. Med., 35*(27), 5094– 5112. doi:https://doi.org/10.1002/sim.7063

Nguyen, V. T., & Dupuy, J. F. (2019). Asymptotic Results in Censored Zero-inflated Poisson Regression. *Communications in Statistics - Theory and Methods, 50*(12), 1-21. doi:https://doi.org/10.1080/03610926.2019.1676442

Olsen, R. J. (1978). Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model. *Econometrica, 46*(5), 1211-1215. doi:https://doi.org/10.2307/1911445

Paarsch, H. (1984). A Monte Carlo Comparison of Estimators for Censored Regression Models. *Journal of Econometrics, 24*, 197-213.

Pelc, K. (2020). Can Covid-era Export Restrictions be Deterred? *Canadian Journal of Political Science, 53*(2), 349-356. doi:https://10.2139/ssrn.3624086

Peng, L., & Huang, Y. (2008). Survival Analysis With Quantile Regression Models. *Journal of the American Statistical Association, 103*(482), 637-649. doi:https://doi.org/10.1198/016214508000000355

Powell, J. L. (1984). Least Absolute Deviations Estimation of the Censored Regression Model. *Journal of Econometrics, 25*(3), 303-325. doi:https://doi.org/10.1016/0304-4076(84)90004-6

Powell, J. L. (1986). Symmetrically Trimmed Least Squares Estimation for Tobit Models. *Econometrica, 54*(6), 1435-1460. doi:https://doi.org/10.2307/1914308

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects. *Journal of Econometrics, 128*(2), 301-323. doi:https://doi.org/10.1016/j.jeconom.2004.08.017

Rufino, C. C. (2016). Microeconometric Analysis of the Eating-out Behavior of Modern Filipino Households: The Tobit, the Craggit and the Heckit Models. *DLSU Business & Economics Review, 26*(1), 50-69.

Saffari, S. E., & Adnan, R. (2011). Zero-inflated Poisson Regression Models with Right Censored Count Data. *Matematika, 27*(7), 21-29. doi: https://doi.org/10.11113/matematika.v27.n.293

Salibian-Barrera, M., & Yohai, V. J. (2008). High Breakdown Point Robust Regression with Censored Data. *The Annals of Statistics, 36*(1), 118-146. doi:https://doi.org/10.1214/009053607000000794

Santos, B., & Bolfarine, H. (2015). Bayesian Analysis for Zero-or-One Inflated Proportion Data Using Quantile Regression. *Journal of Statistical Computation and Simulation, 85*(17), 3579-3593. doi:https://doi.org/10.1080/00949655.2014.986733

Saulo, H., Leão , J., Nobre, J., & Balakrishnan, N. (2021). A Class of Asymmetric Regression Models for Left-Censored Data. *Braz. J. Probab. Stat., 35*(1), 62-84. doi:https://doi.org/10.1214/20-BJPS494

Sauzet, O., Razum, O., Widera, T., & Brzoska, P. (2019). Two-part Models and Quantile Regression for the Analysis of Survey Data with a Spike. The Example of Satisfaction with Health Care. *Frontiers in Public Health, 7*, 146. doi:https://doi.org/10.3389/fpubh.2019.00146

Shmueli, G., Jank, W., & Hyde, V. (2008). Transformations for Semi-Continuous Data. *Computational Statistics & Data Analysis, 52*(8), 4000-4020. doi:https://doi.org/10.1016/j.csda.2008.01.025

Stapleton, D. C., & Young, D. J. (1984). Censored Normal Regression with Measurement Error on the Dependent Variable. *Econometrica, Econometric Society, 52*(3), 737-760. doi:https://doi.org/10.2307/1913474

Stavseth, M. R., Clausen, T., & Roislien, J. (2019). How Handling Missing Data May Impact Conclusions: A Comparison of Six Different Imputation Methods for Categorical Questionnaire Data. *SAGE Open Medicine, 7*. doi:https://doi.org/10.1177/2050312118822912

Stute, W. (1999). Nonlinear Censored Regression. *Statistica Sinica, 9*(4), 1089-1102.

Su, X., & Luo, S. (2017). Analysis of Censored Longitudinal Data with Skewness and a Terminal Event. *Communications in Statistics - Simulation and Computation, 46*(7), 5378-5391. doi:https://doi.org/10.1080/03610918.2016.1157181

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica: Journal of the Econometric Society, 26*(1), 24-36. doi:https://doi.org/10.2307/1907382

Usta, İ., Kantar, Y. M., & Yenilmez, İ. (2018). Estimation of Censored Regression Model with Maximum Entropy Distributions. *Proceeding of the 4th IRSYSC.* Izmir, Turkey: IRSYSC.

Vock, D. M., Davidian, M., Tsiatis, A. A., & Muir, A. J. (2011). Mixed Model Analysis of Censored Longitudinal Data with Flexible Random-effects Density. *Biostatistics, 13*(1), 61-73. doi:https://doi.org/10.1093/biostatistics/kxr026

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine Learning for Survival Analysis: A Survey. *ACM Computing Surveys, 51*(6), 1-36. doi:https://doi.org/10.1145/3214306

Weinblatt, J. (1986). *The Economics Of Export Restrictions: Free Access To Commodity Markets As An Element Of The New International Economic Order.* Routledge.

Wilson, T., Loughran, T., & Brame, R. (2020). Substantial Bias in the Tobit Estimator: Making a Case for Alternatives. *Justice Quarterly, 37*(2), 231-257. doi:https://10.1080/07418825.2018.1517220

Wooldridge, J. M. (2012). *Introductory Econometrics : A Modern Approach.* Mason: Ohio :South-Western Cengage Learning.

Yang, Y., & Simpson, D. G. (2012). Conditional Decomposition Diagnostics for Regression Analysis of Zero-inflated and Left-censored Data. *Statistical Methods in Medical Research, 21*(4), 393-408. doi:https://doi.org/10.1177/0962280210387525

Yenilmez, İ. (2017a). Limited Dependent Variable Models and Estimation Methods. *MS Thesis*. Eskisehir, Turkey: Anadolu University.

Yenilmez, İ., & Mert Kantar, Y. (2017b). A Partially Adaptive Estimator for the Censored Regression Model Based on Generalized Normal Distribution. *Proceedings of the 3rd IRSYSC-2017.* Konya, Turkey.

Yenilmez, İ., & Mert Kantar, Y. (2017c). Estimators of the Censored Regression in the Cases of Heteroscedasticity and Non-Normality. *Proceedings of the 10th International Statistics Congress.* Ankara, Turkey.

Yenilmez, İ., & Mert Kantar, Y. (2019b). An Alternative Estimation Method Based on Alpha Skew Logistic Distribution for Parameters of Censored Regression Model. *Data Science and Applications, 2*(2), 16-20.

Yenilmez, İ., Mert Kantar, Y., & Acıtaş, Ş. (2018). Estimation of Censored Regression Model in the Case of Non-Normal Error. *Sigma Journal of Engineering and Natural Sciences, 36*(2), 513-521.

Yenilmez, İ., Usta, İ., & Mert Kantar, Y. (2019a). Quasi-Maximum Likelihood Estimator based on Moyal Distribution for Censored Data,. *5th International Researchers, Statisticians and Young Statisticians Congress.* Aydın, Turkey.

Yu, K., & Moyeed, R. A. (2001). Bayesian Quantile Regression. *Statistics & Probability Letters, 54*(4), 437-447. doi:https://doi.org/10.1016/S0167-7152(01)00124-9

Yu, K., & Stander, J. (2007). Bayesian Analysis of a Tobit Quantile Regression Model. *Journal of Econometrics, 137*(1), 260-276. doi:https://doi.org/10.1016/j.jeconom.2005.10.002

Zhao, Z., & Severini, T. A. (2017). Integrated Likelihood Computation Methods. *Comput Stat, 32*. doi:https://doi.org/10.1007/s00180-016-0677-z

**APPENDIX**

Estimates of parameters are obtained by solving equations $u(\Theta) = 0$ for all components of parameter space $\Theta = (\mu, \sigma, a, b)$. In addition, the Hessian matrix is obtained as a $4 \times 4$ matrix. For $F_{LEN_0} = F_{LEN}\left[\frac{\mu}{\sigma}\right]$, $F_{LEN_i} = F_{LEN_i}\left[\frac{x_i-\mu}{\sigma}\right]$, $F_{LEN_U} = F_{LEN_U}\left[\frac{c_U-\mu}{\sigma}\right]$, $f_{LEN_0} = f_{LEN}\left[\frac{\mu}{\sigma}\right]$, $f_{LEN_i} = f_{LEN_i}\left[\frac{x_i-\mu}{\sigma}\right]$, and $f_{LEN_U} = f_{LEN_U}\left[\frac{c_U-\mu}{\sigma}\right]$, Hessian matrix is obtained.

In addition, finite-sample properties of an estimator $\widehat{\boldsymbol{\beta}}$, for any finite sample size $n < \infty$ involve expected value $E(\widehat{\boldsymbol{\beta}})$ and variance denoted as $Var(\widehat{\boldsymbol{\beta}})$.

To find, $E(\widehat{\boldsymbol{\beta}})$, $\widehat{\boldsymbol{\beta}}$ found from the solution-root of the score function:

$$S(\beta) = \frac{\partial logL}{\partial \beta} = 0$$

Following Gong and Schaubel (2018), variance of $\widehat{\boldsymbol{\beta}}$ is estimated as:

$$n\left[-\frac{\partial^2 logL}{\partial \beta^2}\right]^{-1} = n\left[-\frac{\partial S(\beta)}{\partial \beta}\right]^{-1}$$

## CURRICULUM VITAE

**Personal Information**

Surname-Name : YENİLMEZ, İsmail

Language Information : Turkish, English

Web Page : https://akademik.eskisehir.edu.tr/ismailyenilmez

: https://www.researchgate.net/profile/Ismail-Yenilmez

**Education Information**

***Doctorate***

University: Eskisehir Technical University

Institute: Institute of Graduate Programs

Program: Program in Statistical Theory

Start - Finish Year: 2017 / 2021

***Master's Degree-1***

University: Anadolu University

Institute: Graduate School of Sciences

Program: Program in Statistical Theory

Start - Finish Year: 2014 / 2017 Master with Thesis

***Master's Degree-2***

University: Marmara University

Institute: Institute of Educational Sciences

Program: Mathematics Teaching

Start - Finish Year: 2012 / 2016 Master with Thesis

*Master's Degree3*

University: Marmara University

Faculty: Atatürk Faculty of Education

Department: Department of Mathematics and Science Education

Program: Mathematics Education

Start - Finish Year: 2007 / 2012 Master without Thesis

*Undergraduate*

University: Marmara University

Faculty: Atatürk Faculty of Education

DEPARTMENT: Department of Mathematics and Science Education

Program: Mathematics Teaching Pr. (5 Years)

Start - Finish Year: 2007 / 2012

**Articles**

**Yenilmez, I.**, Yılmaz, E., Mert Kantar, Y., Aydın, D. (2021) Comparison of Parametric and Semi-Parametric Models with Randomly Right-Censored Data by Weighted Estimators: Two Applications in Colon Cancer and Hepatocellular Carcinoma Datasets, Statistical Methods in Medical Research, (Accepted Paper).

Toker, S., Özbay, N., Üstündağ Şiray, G., **Yenilmez, I.** (2021) Tobit Liu estimation of censored regression model: an application to Mroz data and a Monte Carlo simulation study, Journal of Statistical Computation and Simulation, 91:6, 1061-1091.

Acıtaş, Ş., **Yenilmez, I.**, Şenoglu, B., Mert Kantar, Y. (2021) Modified maximum likelihood estimator under the Jones and Faddy's skew t-error distribution for censored regression model, Journal of Applied Statistics, 48:12, 2136-2151.

Can, F., **Yenilmez, I.**, Çelik, S. (2020). Asma Anacı Çeliklerinde Kök Oluşum Yoğunluğu Yerlerinin Saptanması . Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi, 23 (3), 561-56.

Mert Kantar, Y., Yıldırım, V., **Yenilmez, I.** (2019). Spatial Distribution of Poverty Threshold In Turkey. Nicel Bilimler Dergisi, 1(2): 52-61.

**Yenilmez, I.**, Mert Kantar, Y. (2019). An alternative estimation method based on alpha skew logistic distribution for parameters of censored regression model. Data Sci Appl. 2(2):16–20.

**Yenilmez, I.**, Mert Kantar Y, Acıtaş Ş. (2018). Estimation of censored regression model in the case of non-normal error. Sigma J Eng Nat Sci. 36(2):513–521.

Usta, İ., Arik, İ., **Yenilmez, I.,** Mert Kantar, Y. (2018). A new estimation approach based on moments for estimating Weibull parameters in wind power applications, Energy Conversion and Management, 2018, 0196-8904, 164, 570-578.

Usta, I., Arik, İ., Mert Kantar, Y., **Yenilmez, I.** (2018). Wind speed analysis using the Extended Generalized Lindley Distribution, Renewable Energy, 0960- 1481, 118, 1024-1030.

Usta, İ., Arik, İ., **Yenilmez, I.**, Mert Kantar, Y. (2016). A Study on Estimation of Wind Speed Distribution by Using the Modified Weibull Distribution, International Journal of Informatics Technologies, 1307-9697, 9, 2, 63-70.

Arik, İ. Usta, İ., Mert Kantar, Y., **Yenilmez, I.** (2016). Comparison of Some Estimation Methods of the two parameter Weibull Distribution for Unusual Wind Speed Data Cases, International Journal of Informatics Technologies, 1307-9697, 9, 2, 81-89.

Arik, İ., Mert Kantar, Y., **Yenilmez, I.** (2015). The Evaluation of Robust and Efficient Estimators for Log Logistic Distribution for Censored Data with without Outliers, Journal of Scientific Research and Development, 1115-7569, 2, 12, 24-32.

**Conference Papers**

Yalınbaş Kaya, B., Dizen, H., **Yenilmez, I.** Acute Pancreatitis Severity and Clinicodemographic Features in Elderly Patients; A 5-Year Retrospective Analysis. In Proceedings of the, 3rd International and 14th Academic Geriatrics Congress, 23 – 24 October 2021.

Toker, S., Ozbay, N., Ustundag Siray, G., **Yenilmez, I.** Applying A New Tobit Estimation Method to a Labor Supply Dataset. In Proceedings of the 10th International Eurasian Conference on Mathematical Sciences and Applications (IECMSA-2021).

Yalınbaş Kaya, B., Canakcı, N., Dizen, H., **Yenilmez, I.** Akut Pankreatitte Splanknik Ven Trombozu: 2 Yıllık, Tek Merkez Sonuçları. Bildiri özetleri kitabı, Geleneksel İç Hastalıkları Günleri, İnteraktif Güncelleştirme 2021 (In Turkish).

Toker, S., Ozbay, N., Ustundag Siray, G., **Yenilmez, I.** Liu Estimation for Tobit Regression Model. In Proceedings of the 20th International Symposium on Econometrics, Operations Research and Statistics, 12 February 2020, 14 February 2020.

**Yenilmez, I.**, Mert Kantar, Y. An Estimation with Flexible Error Distributions for Stochastic Frontier Model. In Proceedings of the ICOAEF'19, VI International Conference on Applied Economics and Finance EXTENDED WITH SOCIAL SCIENCES, 16 November 2019, 17 November 2019.

**Yenilmez, I.**, Mert Kantar, Y., Usta, I. Explaining the Female Labor Force Participation in Turkey by Means of the Models With Limited Dependent Variable. In Proceedings of the ICOAEF'19, VI International Conference on Applied Economics and Finance EXTENDED WITH SOCIAL SCIENCES, 16 November 2019, 17 November 2019.

Yildirim, V., Mert Kantar, Y., **Yenilmez, I.** Spatial Distribution of Happiness in Turkey. In Proceedings of the ICOAEF'19, VI International Conference on Applied Economics and Finance EXTENDED WITH SOCIAL SCIENCES, 16 November 2019, 17 November 2019.

**Yenilmez, I.**, Mert Kantar, Y. Analysis of Turkey Household Budget Survey Data with Quantile Regression. In Proceedings of the 5th International Researchers, Statisticians and Young Statisticians Congress (IRSYSC19), 18 October 2019, 20 October 2019.

**Yenilmez, I.**, Usta, I., Mert Kantar, Y. Quasi-Maximum Likelihood Estimator based on Moyal Distribution for Censored Data. In Proceedings of the 5th International Researchers, Statisticians and Young Statisticians Congress (IRSYSC19), 18 October 2019, 20 October 2019.

**Yenilmez, I.**, Mert Kantar, Y. An Analysis of Export Data with Panel Tobit Model. In Proceedings of the 2nd International Conference on Data Science and Applications (ICONDATA'19), 03 October 2019, 06 October 2019.

**Yenilmez, I.**, Mert Kantar, Y. Estimation of Censored Regression in case of Skew and Bimodal Errors. In Proceedings of the 2nd International Conference on Data Science and Applications (ICONDATA'19), 03 October 2019, 06 October 2019, 978-605-031-662-9.

Usta, I. Mert Kantar, Y., **Yenilmez, I.** Estimation of Censored Regression Model with Maximum Entropy Distributions. In Proceedings of the 4th International Researchers, Statisticians and Young Statisticians Congress, 28 April 2018, 30 April 2018.

Acitas, S., Senoglu, B., Mert Kantar, Y., **Yenilmez, I.**, Estimation for the Censored Regression Model with the Jones and Faddy's Skew-*t* Distribution: Maximum Likelihood and Modified Maximum Likelihood Estimation Methods. In Proceedings of the 10th International Statistics Congress, 06 December 2017, 08 December 2017.

**Yenilmez, I.**, Mert Kantar, Y. Estimators of the Censored Regression in the Cases of Heteroscedasticity and Non-Normality. In Proceedings of the 10$^{th}$ International Statistics Congress, 06 December 2017, 08 December 2017.

Acitas, S., **Yenilmez, I.**, Senoglu, B., Mert Kantar, Y. Modified Maximum Likelihood Estimation for the Censored Regression Model. In Proceedings of the 13th IMT-GT International Conference On Mathematics, Statistics and Their Applications, 04 December 2017, 07 December 2017.

Mert Kantar, Y., **Yenilmez, I.** Technical Efficiency Assessment of Turkish Banks Using Stochastic Frontier Method. In Proceedings of the I. International Symposium on Economics, Finance, and Econometrics, 21 September 2017, 23 September 2017.

**Yenilmez, I.**, Mert Kantar, Y. A Partially Adaptive Estimator for the Censored Regression Model Based on Generalized Normal Distribution. In Proceedings of the, 3$^{rd}$ International Researchers, Statisticians and Young Statisticians Congress, 24 May 2017, 26 May 2017.

**Yenilmez, I.**, Acitas S., Mert Kantar, Y. Estimation Based on Generalized Logistic Distribution for the Censored Regression Model. In Proceedings of the International Workshop on Mathematical Methods in Engineering, 27 April 2017, 29 April 2017.

Usta, I., Mert Kantar, Y., Arik, I., **Yenilmez, I.** A Statistical Investigation on Wind Energy Potential of Northwest of Turkey. 2016 International Conference on Engineering & MIS (ICEMIS), 2016, pp. 1-6, doi: 10.1109/ICEMIS.2016.7745315.

Mert Kantar, Y., Usta, I., Arik, I., **Yenilmez, I.** Distributions of Wind Speed at Different Heights. 2016 International Conference on Engineering & MIS (ICEMIS), 2016, pp. 1-6, doi: 10.1109/ICEMIS.2016.7745320.

Usta, I., Mert Kantar, Y., Arik, I., **Yenilmez, I.** A New Estimator Based on Median and Mode for Weibull Distribution in Wind Energy. In Proceedings of the 4th European Conference on Renewable Energy Systems 4. European Conference on Renewable Energy Systems (ECRES'16), 28 August 2016, 31 August 2016..

**Yenilmez, I.**, Arik, I., Usta, I., Mert Kantar, Y. The Generalized Lindley Distribution to Model Wind Speed. In Proceedings of the 4th European Conference on Renewable Energy Systems 4. European Conference on Renewable Energy Systems (ECRES'16), 28 August 2016, 31 August 2016.

**Yenilmez, I.**, Mert Kantar, Y. Tobit Regression for Modeling Gastric Cancer. In Proceedings of the International Congress on Fundamental and Applied Sciences (ICFAS'16), 22 August 2016, 26 August 2016.

**Yenilmez, I.**, Yavuz, I. Using Technology in Statistical Education at High School Level VUstat Software and a Case of Eskisehir Tepebaşı. In Proceedings of the 7th International Congress on New Trends in Education, 13 May 2016, 15 May 2016.

**Yenilmez, I.**, Mert Kantar, Y., Arik, I., Usta, I. Analysis of the Modified Weibull Distribution for Estimation of Wind Speed Distribution. In Proceedings of the International Conference on Engineering & MIS 2015 (ICEMIS'15). Association for Computing Machinery, New York, NY, USA, Article 49, 1–5. DOI:https://doi.org/10.1145/2832987.2833059.

Arik, I., Mert Kantar, Y., Usta, I., **Yenilmez, I.** Evaluation of Robust Estimation Methods in Estimating Weibull Parameters for Wind Energy Application. In Proceedings

of the International Conference on Engineering & MIS 2015 (ICEMIS'15). Association for Computing Machinery, New York, NY, USA, Article 33, 1–6. DOI:https://doi.org/10.1145/2832987.2833041.

**Projects**

Quantile Regression Methods and Models for Censored Data. **Researcher**. Start - Finish Date: 2021 / -

Censored Regression Models and Robust Estimation. Scientific research projectsupported by Higher Education Institutions. **Researcher**. Start - Finish Date: 2020 / -

Robust Estimation and Applications of Econometric Models with Limited Dependent Variables. Scientific research project supported by Higher Education Institutions. **Researcher**. Start - Finish Date: 2020 / -

Tobit, Censored and Quantile Regression Models and Estimation. Scientific research project supported by Higher Education Institutions. **Researcher**. Start - Finish Date: 2019 / 2020

Censored Regression Models: Non-normalty and Heteroscedasticity problem. Project Type: Scientific research project supported by Higher Education Institutions. **Researcher**. Start - Finish Date: 2017 / 2018

Distributions in the Case of Censored Data, Regression Models and Applications. Project Type: Scientific research project supported by Higher Education Institutions. **Researcher**. Start - Finish Date: 2017 / 2018

Maximum Likelihood and Modified Maximum Likelihood Estimation Methods for Censored Regression Models. **Researcher**. Scientific research project supported by Higher Education Institutions. Start - Finish Date: 2016 / 2018

Limited Dependent Variable Models and Estimation Methods. **Researcher.** Scientific research project supported by Higher Education Institutions. Start - Finish Date: 2016 / 2017

Survival Models and Applications. **Researcher.** Scientific research project supported by Higher Education Institutions. Start - Finish Date: 2016 / 2017

Wind Speed Modeling the proposed Statistical Distributions Families: An Investigation of the Proposed Distribution of Wind Power Potential in Different Regions of Turkey. **Researcher.** Scientific research project supported by Higher Education Institutions. Start - Finish Date: 2015 / 2016

**Awards**

Turkey Student Selection Examination Scholarship given to students rank among the top 1000, HÜSNÜ M. ÖZYEĞİN FOUNDATION, 2007-2012 (for a period of five years); Yenilmez ranks 742nd out of 1615534 students

BİDEB Support: Yenilmez is supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) within the scope of the 2214 / A International Doctoral Research Fellowship Program run by Science Fellowships and Grant Programme Directorate; Yenilmez continued his studies for a while at Sapienza University of Rome.