University of Brasília
Institute of Exact Sciences
Department of Statistics

Master's Dissertation

Parametric quantile regression for income data

by

Giovanna Valadares Borges

Brasília, April, 2022

Parametric quantile regression for income data

by

Giovanna Valadares Borges

A dissertation submitted to the Departament of Statistics at the University of Brasília in partial fulfilment of the requirements for the degree Master in Statistics.

Supervisor: Prof. Dr. Helton Saulo Bezerra dos Santos

Brasília, April, 2022

A dissertation submitted to the Departament of Statistics at the University of Brasília in partial fulfilment of the requirements for the degree Master in Statistics.

Approved by:

Prof. Dr. Helton Saulo Bezerra dos Santos
Supervisor, EST/UnB

Prof. Dr. Roberto Vila Gabriel
EST/UnB

Prof. Dr. Jeremias da Silva Leão
DE/UFAM

Prof. Dr. José Augusto Fiorucci
EST/UnB

I hope you know your limits well, but don't stay within those limits.

Overcome the limits each day.

(Nam Joon Kim)

To my family, Cláudia, Eduardo, Raquel, Clarice e Antônio (*in memoriam*).

I am who I am today because all and each of you believe in me.

Acknowledgments

Thanks to my advisor, Dr. Helton Saulo, for all the knowledge given and immense understanding and support he gave throughout this work. Also thanks to professors Dr. Roberto Vila, and Dr. Jeremias Leão, for accepting the invitation to be part of this work's evaluation committee.

Thanks to my friends and colleagues from our masters program, for the good moments and all the knowledge we shared together.

Lastly, thanks to my family, for all the support and comforting words during the last 2 years. I wouldn't have done it without you all.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo

REGRESSÃO QUANTÍLICA PARAMÉTRICA PARA DADOS DE RENDA

Modelos de regressão normais univariados são ferramentas estatísticas amplamente aplicadas em muitas áreas da economia. No entanto, dados de renda têm comportamento assimétrico e tem melhor performance de modelagem com distribuições não-normais. A modelagem da renda desempenha um papel importante na determinação dos ganhos dos trabalhadores, além de ser um importante tópico de pesquisa em economia do trabalho. Os modelos de regressão quantílica são alternativas robustas aos modelos tradicionais baseados em média. Isso porque, ao invés de focar na média condicional, esses modelos são baseados no quantil condicional, como a mediana. A abordagem quantílica tem a vantagem de proporcionar flexibilidade na modelagem, pois permite considerar os efeitos das variáveis explicativas em todo o espectro da variável dependente, incluindo assim também o efeito sobre a mediana, que é uma medida de tendência central melhor que a média no contexto assimétrico. A modelagem de renda começou com proposições de Vilfredo Pareto, estabelecendo uma lei sobre como funciona a distribuição de renda. Mais tarde, essas proposições sugeriram uma distribuição - conhecida como distribuição de Pareto - que estabeleceu uma referência para outras distribuições, como log-normal e gama, para mostrar seu potencial na descrição da distribuição de renda. No entanto, apesar de Pareto, Log-Normal e Gama serem as distribuições mais frequentemente aplicadas aos dados de renda devido à sua capacidade de descrever esse tipo de dado, elas têm limitações. O modelo de Pareto é mais apropriado apenas para descrever a cauda superior da distribuição. Por outro lado, as distribuições

Log-Normal e Gama têm um desempenho ruim na descrição das caudas superior e inferior em distribuições reais. Distribuições de renda como Dagum e Singh-Maddala mostraram, na literatura, terem capacidade de superar as distribuições de Pareto, log-normal e gama em termos de ajuste de modelo. A distribuição Dagum tem flexibilidade para lidar com mudanças de distribuição, renda nula e negativa; faixa de renda com início de renda mínima positiva não predeterminada e funções de densidade estritamente decrescentes e unimodais. Ela também demonstra ter uma boa qualidade de ajuste, acomodando bem caudas pesadas, característica comumente encontrada em dados de renda. Já a distribuição de Singh-Maddala deriva do conceito de taxa de risco, uma abordagem amplamente utilizada na literatura de confiabilidade. Uma de suas vantagens é ser mais flexível que outras distribuições de renda. As distribuições Dagum e Singh-Maddala são casos especiais da distribuição beta generalizada do segundo tipo (GB2) e ambas obedecem à lei fraca lei de Pareto, isto é, convergem assintoticamente à distribuição de Pareto. Desta forma, o objetivo deste trabalho é propor modelos de regressão quantílicos paramétricos baseados em duas importantes distribuições de renda assimétricas, a saber, as distribuições Dagum e Singh-Maddala. Os modelos quantílicos propostos são baseados em reparametrizações das distribuições originais, através da inserção de um parâmetro de quantil. Essa abordagem leva em consideração o ganho em flexibilidade de modelagem, tornando possível analisar os modelos reparametrizados propostos a partir dos efeitos das variáveis explicativas ao longo do espectro da variável dependente, com especial foco no efeito da mediana. São apresentadas as reparametrizações, algumas importantes propriedades das distribuições, e os modelos de regressão quantílica em conjunto com o processo inferencial, levando em consideração também a inclusão do parâmetro quantílico em métricas econométricas como Índice de Gini e as Curvas de Lorenz e Bonferroni – utilizadas para avaliar o grau de desigualdade entre indivíduos, principalmente no âmbito de renda. Em seguida, apresentamos estudos de simulação de Monte Carlo, considerando a avaliação do desempenho das estimativas de máxima verossimilhança e uma análise da distribuição empírica dos resíduos para avaliar a performance geral dos modelos. Para estudar os estimadores de máxima verossimilhança, usamos o cálculo

do viés, erro quadrático médio (MSE) e probabilidade de cobertura (CP). Espera-se que, à medida que o tamanho da amostra aumente, o viés e o MSE diminuam, à medida que o CP se aproxima do nível nominal de 95%. Já para o estudo dos resíduos, analisamos os resultados dos resíduos Cox-Snell Generalizado (GCS) e Quantil Randomizado (RQ) com as respectivas estatísticas descritivas (média, mediana, desvio padrão, coeficiente de assimetria e coeficiente de curtose). Foi possível notar que ambos os modelos atendem aos resultados esperados para esses estudos, com as estimativas apresentando bom desempenho e os resíduos com boa concordância em relação às suas distribuições de referência. Para a aplicação a dados reais, foram usados dados de renda familiar chilena em 2016, fornecido pelo Instituto Nacional de Estatística do Chile. Os resultados dos nossos modelos com esse conjunto de dados foram comparados ao estudo anterior que propôs um modelo de regressão quantílica Birnbaum-Saunders (Sánchez et al., 2021b). Os resultados mostram que ambos os modelos propostos tiveram um desempenho melhor que o Birnbaum-Saunders, com ajustes melhores e melhor adaptabilidade em relação aos resíduos. Assim, conclui-se que os resultados foram favoráveis ao uso dos modelos de regressão quantílica de Singh-Maddala e Dagum para dados positivos com assimetria, especialmente para dados de renda.

Palavras-chave: Distribuições de renda, regressão quantílica, dados de renda, reparametrização.

Abstract

Univariate normal regression models are statistical tools widely applied in many areas of economics. Nevertheless, income data have asymmetric behavior and are best modeled by non-normal distributions. The modeling of income plays an important role in determining workers' earnings, as well as being an important research topic in labor economics. Thus, the objective of this work is to propose parametric quantile regression models based on two important asymmetric income distributions, namely, Dagum and Singh-Maddala distributions. The proposed quantile models are based on reparameterizations of the original distributions by inserting a quantile parameter. The quantile approach has the advantage of providing flexibility in modeling, as it allows considering the effects of explanatory variables throughout the spectrum of the dependent variable, thus also including the effect on the median, which is a measure of central tendency better than the mean in the asymmetric context. We present the reparameterizations, some important properties of the distributions, and the quantile regression models with their inferential aspects. We proceed with Monte Carlo simulation studies, considering the maximum likelihood estimation performance evaluation and an analysis of the empirical distribution of two residuals. The Monte Carlo results show that both models meet the expected outcomes. We apply the proposed quantile regression models to a household income data set provided by the National Institute of Statistics of Chile. The results to our models with this data set were compared to a previous study by Sánchez et al. (2021b) that introduced a Birnbaum-Saunders quantile regression model. We showed that both proposed models had a better performance than the Birnbaum-Saunders model. Thus, we conclude that results were favorable to the use of Singh-Maddala and Dagum quantile regression models for positive asymmetric data, such as income data.

Keywords: Income distributions, quantile regression, income data, reparameterization.

Contents

List of Tables

List of Figures

Chapter 1

Parametric quantile regression

1.1 Introduction

Income modeling plays an important role in determining workers' earnings, as well as being an important research topic in labor economics. In general, income data are modeled using mean-based regression models based on the normality assumption. Nevertheless, income is often unequally distributed, hence why this type of data usually has an asymmetric behavior and then the mean is not an appropriate central tendency measure. Therefore, quantile regression models are usually more useful in this context; see Galarza, Zhang, and Lachos (2020), Sánchez et al. (2021a) and Saulo et al. (2021).

Quantile regression models are robust alternatives to traditional mean-based models. That is because instead of focusing on the conditional mean, these models are based on the conditional quantile, such as median; see Koenker (2005). The quantile approach has the advantage of providing flexibility in modeling, as it allows considering the effects of explanatory variables throughout the spectrum of the dependent variable, thus also including the effect on the median, which is a measure of central tendency better than the mean in the asymmetric context.

Income modelling begins with Pareto (1897) propositions, establishing a law on how income distribution works. Later on, this suggested a distribution – known as Pareto distribution – and it has set a reference for other distributions, such as log-normal and gamma, to show their potential as for describing income distribution; see Shirras (1935) and Reed (2003). Even though the Pareto, log-normal and gamma are the most frequently distributions applied to income data because of their abilities to describe this type of data, they have limitations.

The Pareto model is appropriate to describe only the upper tail of the distribution. On the other hand, the log-normal and gamma distributions perform poor in describing both the upper and lower tails of the actual distributions. Income distributions such as Dagum and Singh-Maddala have outperformed the Pareto, log-normal and gamma distributions in terms of model fitting; see Cramer (1971) and Dagum (2008).

Originally proposed by Dagum (1973) and Dagum (1975), the Dagum distribution has flexibility to deal with distribution changes, nil and negative income; income range with non-predetermined positive minimum income start, and strictly decreasing and unimodal density functions. This distribution also shows good goodness of fit to income data and obeys a weak version of the Pareto law, i.e. it asymptotically approaches the Pareto distribution. The Dagum model accommodates both heavy tails and an interior mode, characteristics commonly found in income data, and not found singly in well-known distributions – such as log-normal and Pareto; see Krämer and Ziebach (2002), Dagum (2008), and Kleiber (2008).

The Singh-Maddala distribution was derived from the concept of hazard rate, an approach widely used in the reliability literature; see Singh and Maddala (1976). This model also obeys the weak Pareto law, and one of its advantages is to be more flexible than other income distributions. The Dagum and Singh-Maddala distributions are special cases of the generalized beta distribution of the second kind (GB2); for more details on these models, one may refer to the works by Kleiber (1996), Kleiber (2008), Kumar (2017a), and Hajargasht et al. (2012).

This work explores a parametric quantile regression approach for the Dagum and Singh-Maddala distributions. We first introduce reparameterizations of the Dagum and Singh-Maddala model by inserting quantile parameters, and then develop the new regression models. We then demonstrate that the proposed models outperform the recently proposed Birnbaum-Saunders quantile regression model (Sánchez et al., 2021b) in terms of model fitting.

The rest of this paper proceeds as follows. In Section 1.2, we describe the usual Dagum and Singh-Maddala distributions and propose reparameterizations of these distributions in terms of a quantile parameter. In this section, we also derive some properties including mode, moments, Bonferroni curve, Lorenz curve, and Gini coefficient. In Section 1.3, we introduce the quantile regression models and also describe the parameter estimation by the maximum likelihood (ML) method.In Section 1.4, we carry out a Monte Carlo simulation study to evaluate the performance of the estimators and generalized Cox-Snell (GCS) and random quantile (RQ) residuals. In Section 1.5, we apply the Dagum and Singh-Maddala quantile regression models to a household income data set provided by the National Institute of Statistics of Chile, and finally in Section 1.6, we provide some concluding remarks.

1.2 Classical and quantile-based income distributions

In this section, we describe the classical Singh-Maddala and Dagum distributions along with the proposed quantile-based reparameterizations of these distributions, which will be useful subsequently for developing the parametric quantile regression models. We also derive some important properties for each model, including mode, moments, Bonferroni curve, Lorenz curve, and Gini coefficient.

1.2.1 Classical Singh-Maddala distribution

If a random variable Y follows a Singh-Maddala distribution with shape parameters $a, q > 0$ and scale parameter $b > 0$, denoted by $Y \sim \text{SM}(a, b, q)$, then the corresponding probability density function (PDF) and cumulative distribution function (CDF) are given by

$$f_{\text{SM}}(y; a, b, q) = \frac{aq(y/b)^{a-1}}{b[1 + (y/b)^a]^{1+q}}, \quad y > 0, \tag{1.1}$$

and

$$F_{\mathrm{SM}}(y; a, b, q) = 1 - [1 + (y/b)^a]^{-q}, \quad y > 0, \tag{1.2}$$

respectively. The Singh-Maddala distribution includes as special cases the Lomax distribution when $a = 1$, and the log-logistic when $q = 1$. If Y follows a Singh-Maddala distribution, then $1/Y$ follows a Dagum distribution, and vice-versa.

The 100τ-th quantile of $Y \sim \mathrm{SM}(a, b, q)$ is obtained by inverting Equation (1.2), which yields

$$q(\tau; a, b, q) = b{c_q}^{1/a} \quad \text{for } c_q = (1-\tau)^{-1/q} - 1 \text{ and } 0 < \tau < 1. \tag{1.3}$$

1.2.2 Quantile-based Singh-Maddala distribution

From the quantile function (1.3), we find that the most parsimonious way of conducting the reparametrization is using the scale parameter b, where we can then write

$$b = \gamma {c_q}^{-1/a}, \quad \gamma > 0,$$

where $\gamma = q(\tau; a, b, q)$. Then, the quantile-based Singh-Maddala PDF is given by

$$f_{\mathrm{QSM}}(y; a, \gamma, q) = \frac{aqc_q(y/\gamma)^{a-1}}{\gamma[1 + c_q(y/\gamma)^a]^{1+q}}, \quad y > 0,$$

with notation $Y \sim \mathrm{QSM}(a, \gamma, q)$.

If $Y \sim \mathrm{QSM}(a, \gamma, q)$, then the following properties hold:

(QSM1) Mode (Kumar, 2017b; Klugman, Panjer, and Willmot, 2019):

$$\left(\frac{a-1}{aq+1}\right)^{1/a}, \quad a > 1, \text{ else } 0.$$

(QSM2) Real moments (Kumar, 2017b; Klugman, Panjer, and Willmot, 2019):

$$\mathbb{E}(Y^r) = \frac{q\gamma^r}{c_q^{r/a}} \mathrm{B}\left(1 + \frac{r}{a}, q - \frac{r}{a}\right), \quad -a < r < aq,$$

where $\mathrm{B}(x, y)$ denotes the beta function.

(QSM3) Renyi entropy (Kumar, 2017b):

$$I_R(\rho) = \frac{\rho \log(q)}{1-\rho} - \log(a) + \log(\gamma) + \frac{1}{a}\log(c_q) \\ - \frac{1}{1-\rho} \log \mathrm{B}\left(\frac{\rho(a-1)+1}{a}, \rho(q+1) - \frac{\rho(a-1)+1}{a}\right), \quad \rho > 0,\ \rho \neq 1.$$

When $\rho \to 1$, the Renyi entropy converges to the Shannon entropy.

(QSM4) Truncated moments (Kumar, 2017b):

$$\mathbb{E}(Y^r \mathbb{1}_{\{Y>x\}}) = \frac{aq\gamma^r(\gamma/x)^{aq-r}}{(aq-r)c_q^q} {}_2F_1\left(1+q, q-\frac{r}{a}; q-\frac{r}{a}+1; -\frac{(\gamma/x)^a}{c_q}\right), \quad aq > r,$$

where ${}_2F_1(a, b; c; x)$ denotes the Gauss hypergeometric function.

(QSM5) Bonferroni curve (Kumar, 2017b):

$$B(\tau) = \frac{ac_q^{1+1/a}}{\tau(1+a)\mathrm{B}(1+\frac{1}{a}, q-\frac{1}{a})} {}_2F_1\left(1+q, 1+\frac{1}{a}; 2+\frac{1}{a}; -c_q\right), \quad aq > 1.$$

(QSM6) Lorenz curve: $L(\tau) = \tau B(\tau)$.

(QSM7) Gini coefficient (see Theorem 1):

$$G = \frac{\mathrm{B}(1+\frac{1}{a}, 2q-\frac{1}{a})}{\mathrm{B}(1+\frac{1}{a}, q-\frac{1}{a})} \left\{ \frac{2q(1+\frac{1}{a})}{(2+\frac{1}{a})(1+2q)} \left[\frac{\Gamma(1+2q)\Gamma(q-\frac{1}{a})}{\Gamma(1+q)\Gamma(2q-\frac{1}{a})} - 1\right] - 1 \right\}, \quad aq > 1, \tag{1.4}$$

where $\Gamma(z)$ is the gamma function.

Lemma 1. The Gini coefficient corresponding to a continuous random variable X with density $f(x)$, admits the representation: $G = (I - J)/\mu$, where μ is the mean of X and

$$I = \int_0^\infty \mathbb{E}(X\mathbb{1}_{\{X>y\}})f(y)\,\mathrm{d}y, \quad J = \int_0^\infty y\mathbb{P}(X > y)f(y)\,\mathrm{d}y.$$

Theorem 1. The Gini coefficient corresponding to Singh-Maddala distribution is given by formula (1.4).

Proof. See proof in Appendix A. □

Note that the Gini coefficient is presented as a property here since it is a well-known economic index that, considering a certain population and a variable of interest, measures inequality. Since it is an abstract concept widely discussed, creating metrics to measure inequality can be difficult. Gini coefficient can be considered a pioneer when it comes to the adoption of more precise measures to evaluate inequality, as much as it has good properties such as dealing with negative incomes; see Allison (1978), Berrebi and Silber (1985), Chen, Tsaur, and Rhai (1982), and Stich (1996). Lorenz and Bonferroni curves are also considered here since both hold some relationship with Gini coefficient; see Fellman et al. (2012) and Pundir, Arora, and Jain (2005).

1.2.3 Classical Dagum distribution

The PDF and CDF of a random variable Y following a classical Dagum distribution with shape parameters $a, p > 0$ and scale parameter $b > 0$, denoted by $Y \sim \text{DA}(a, b, p)$, are given by

$$f_{\text{DA}}(y; a, b, p) = \frac{ap(y/b)^{ap-1}}{b[1 + (y/b)^a]^{1+p}}, \quad y > 0,$$

and

$$F_{\text{DA}}(y; a, b, p) = \left[1 + (y/b)^{-a}\right]^{-p}, \quad y > 0, \tag{1.5}$$

respectively. It is simple to observe that $f_{\text{DA}}(y; a, b, p) = (y/b)^{a(p-1)} f_{\text{SM}}(y; a, b, p)$ and that when $p = 1$ both densities coincide with the log-logistic distribution.

The 100τ-th quantile of $Y \sim \text{DA}(a, b, p)$ is given by

$$q(\tau; a, b, p) = b{e_p}^{-1/a} \quad \text{for } e_p = \tau^{-1/p} - 1 \text{ and } 0 < \tau < 1.$$

1.2.4 Quantile-based Dagum distribution

By observing the three parameters of the classical Dagum distribution, the isolation of the scale according to the quantile would produce the simplest form of the new quantile-based Dagum distribution; it is represented as follows:

$$b = \gamma {e_p}^{1/a}, \quad \gamma > 0,$$

where $\gamma = q(\tau; a, b, p)$. Then, the quantile-based Dagum PDF can be written as

$$f_{\text{QDA}}(y; a, \gamma, p) = \frac{ap(y/\gamma)^{ap-1}}{\gamma {e_p}^p[1 + e_p^{-1}(y/\gamma)^a]^{1+p}}, \quad y > 0.$$

If $Y \sim \text{QDA}(a, \gamma, p)$, then the following properties hold:

(QDA1) Mode (Klugman, Panjer, and Willmot, 2019):

$$\gamma e_p^{1/a}\left(\frac{ap-1}{p+1}\right)^{1/p}, \quad ap > 1, \text{ else } 0.$$

(QDA2) Real moments (Klugman, Panjer, and Willmot, 2019): If $Y \sim \text{DA}(a, \gamma, p)$ then

$$\mathbb{E}(Y^r) = \gamma^r e_p^{r/a}\text{B}\left(a + \frac{r}{p}, 1 - \frac{r}{p}\right), \quad -ap < r < p.$$

(QDA3) Renyi entropy (Dey, Al-Zahrani, and Basloom, 2017):

$$I_R(\rho) = \frac{1}{1-\rho}\log\left[a^{\rho-1}p^{\rho}e_p^{\rho}\gamma^{1-\rho}e_p^{(1-\rho(1+a))/a}\,\frac{\Gamma(p\rho+\frac{1-\rho}{a})\Gamma(\rho-\frac{1-\rho}{a})}{\Gamma(\rho p+\rho)}\right], \quad \rho>0,\ \rho\neq 1.$$

(QDA4) Truncated moments (see Appendix A):

$$\mathbb{E}(Y^r \mathbb{1}_{\{Y>x\}}) = \frac{p\gamma^r(\gamma/x)^{a(1-ar)}}{(1-ar)e_p^{-p}}\,{}_2F_1\left(1+p, 1-ar; 2-ar; -\frac{(\gamma/x)^a}{e_p^{-1}}\right), \quad ar<1.$$

(QDA5) Bonferroni curve (Dey, Al-Zahrani, and Basloom, 2017):

$$B(\tau) = \frac{p}{\tau e_p^{1/a}\mathrm{B}(a+\frac{1}{p}, 1-\frac{1}{p})}\sum_{n=0}^{\infty}(-1)^n\binom{p+n+1}{n}\frac{e_p^{-(n+1)/a}}{n-\frac{1}{a}+1}, \quad p>1.$$

(QDA6) Lorenz curve: $L(\tau)=\tau B(\tau)$.

(QDA7) Gini coefficient (see Theorem 2):

$$G = \frac{p\mathrm{B}(p-a(1-a), 2+p+a(1-a))}{\mathrm{B}(a+\frac{1}{p}, 1-\frac{1}{p})e_p^{a(1-a)+(1/a)-p}}\,I(a,p) - \frac{p\mathrm{B}(2p+\frac{1}{a}, 1-\frac{1}{a})}{\mathrm{B}(a+\frac{1}{p}, 1-\frac{1}{p})} - 1, \quad a<1,\ p>1 \tag{1.6}$$

where $I(a,p)=\int_0^1 t^{-a}\,{}_2F_1(1+p, p-a(1-a); 2(1+p); 1+t)\mathrm{d}t$.

Theorem 2. The Gini coefficient corresponding to quantile-based Dagum model is given by formula (1.6).

Proof. See proof in Appendix A. □

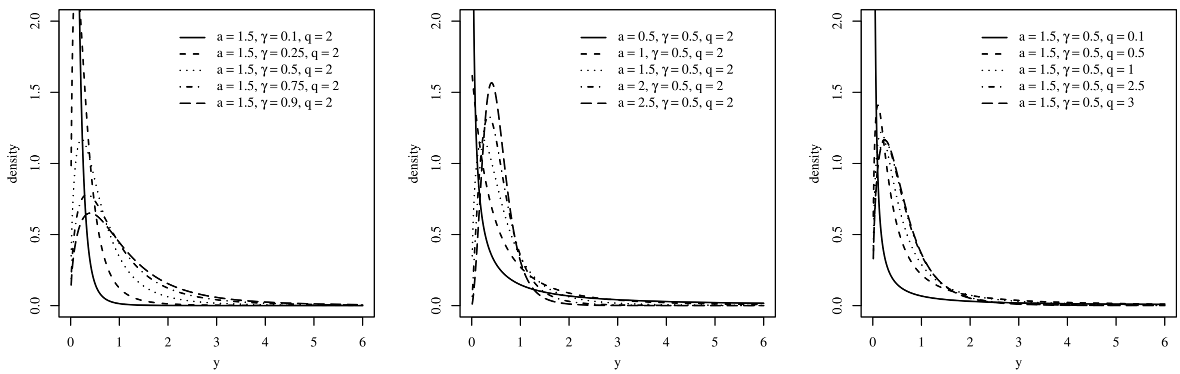
1.2.5 Summary table and density plots

Table 1.1 presents the Singh-Maddala and Dagum distributions in their original and quantile-based versions. Figures 1.1 and 1.2 display different shapes of the quantile-based income distributions for different combinations of parameters, considering scenarios where a, p, q and γ

are fixed. For Singh-Maddala, we can see that a influences the kurtosis and skewness, while q changes the kurtosis, as it decreases when q increases. For Dagum, we see a similar pattern for a, changing both kurtosis and skewness, while p affects the kurtosis.

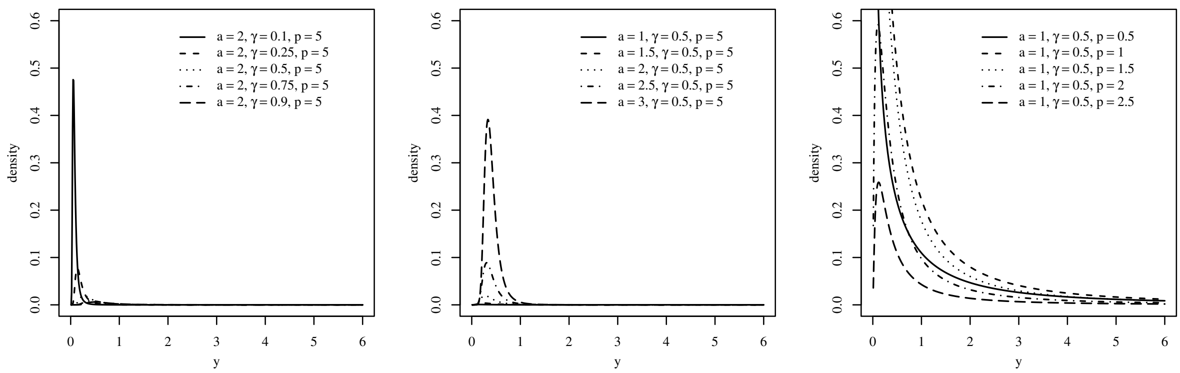
Table 1.1: Income distributions for the original and quantile parameterizations.

Distribution	Classical density	γ: τ-th quantile	Substitution	Quantile-based density
Singh-Maddala	$\frac{aq(y/b)^{a-1}}{b[1+(y/b)^a]^{1+q}}$	$\gamma = b{c_q}^{1/a}$	$b = \frac{\gamma}{{c_q}^{1/a}}$	$\frac{aqc_q(y/\gamma)^{a-1}}{\gamma[1+c_q(y/\gamma)^a]^{1+q}}$
Dagum	$\frac{ap(y/b)^{ap-1}}{b[1+(y/b)^a]^{1+p}}$	$\gamma = b{e_p}^{-1/a}$	$b = \gamma {e_p}^{1/a}$	$\frac{ap(y/\gamma)^{ap-1}}{\gamma {e_p}^p[1+e_p^{-1}(y/\gamma)^a]^{1+p}}$



(a) a and q fixed (b) γ and q fixed (c) γ and a fixed

Figure 1.1: Quantile-based Singh-Maddala PDFs for some choices of parameters.



(a) a and p fixed (b) γ and p fixed (c) γ and a fixed

Figure 1.2: Quantile-based Dagum PDFs for some choices of parameters.

1.3 Income quantile regression models

Let $Y_1, \ldots, Y_n$ be independent random variables such that each Y_i, for $i = 1, \ldots, n$, has PDF given by some reparameterized income distribution defined in Table 1.1, for a fixed (known) probability $\tau \in (0, 1)$ associated with the quantile of interest. Then, in the formulation of the Singh-Maddala and Dagum quantile regression models, the parameter γ of Y_i assumes the following functional relation:

$$g(\gamma_i) = \mathbf{x}_i^\top \boldsymbol{\beta}(\tau), \tag{1.7}$$

where $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \ldots, \beta_k(\tau))^\top$ is the vector of the unknown regression coefficients, which are assumed to be functionally independent; $\boldsymbol{\beta}(\tau) \in \mathbb{R}^{(k+1)}$, with $k + 1 < n$; and $\mathbf{x}_i = (x_{i1}, \ldots, x_{il})^\top$ is the observations of the l known regressors, for $i = 1, \ldots, n$. In addition, we assume that the covariate matrices $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ has rank l. The link function $g : \mathbb{R}^+ \to \mathbb{R}$ in (1.7) must be strictly monotone, positive, and at least twice differentiable, with $g^{-1}(\cdot)$ being the inverse function of $g(\cdot)$. Here, we chose to work with $\log$ as link since it is widely used and more flexible when it comes to simulation studies.

Consider a sample of size n, $Y_1, \ldots, Y_n$ say, such that $Y_i \sim \text{QSM}(a, \gamma_i, q)$. Then, the corresponding likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}(\tau)^\top, a, q)^\top$, is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{aqc_q(y/\gamma_i)^{a-1}}{\gamma_i[1 + c_q(y/\gamma_i)^a]^{1+q}}\,, \tag{1.8}$$

where γ_i is as in (1.7). By applying the logarithm in (1.8), we obtain the log-likelihood function

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \left[(a-1)\log(aqc_q(y/\gamma_i))\right] - \left[\log(\gamma_i) + (1+q)\log(1 + c_q(y/\gamma_i)^a)\right] \right\}. \tag{1.9}$$

Now, consider a sample of size n, $Y_1, \ldots, Y_n$ say, such that $Y_i \sim \text{QDA}(a, \gamma_i, p)$. Then, the corresponding likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}(\tau)^{\top}, a, p)^{\top}$, is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{ap(y/\gamma_i)^{ap-1}}{\gamma_i {e_p}^{p}[1 + e_p^{-1}(y/\gamma_i)^{a}]^{1+p}} \tag{1.10}$$

where γ_i is as in (1.7). By applying the logarithm in (1.10), we obtain the log-likelihood function

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \left[(ap-1)\log(ap(y/\gamma_i))\right] - \left[p\log(e_p\gamma_i) + (1+p)\log(1 + e_p^{-1}(y/\gamma_i)^{a})\right] \right\}. \tag{1.11}$$

To obtain the ML estimate of $\boldsymbol{\theta}$, it is necessary to maximize the log-likelihood functions in (1.9) and (1.11). Therefore, we need to differentiate the log-likelihood functions to find the score vector $\dot{\ell}(\boldsymbol{\theta})$ and then equate it to zero, providing the likelihood equations. They are solved using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method, see Mittelhammer, Judge, and Miller (2000). The method is implemented and applied using the `R` software.

Departures from regression models assumptions and goodness of fit are assessed by means of a residual analysis. Particularly, we use the generalized Cox-Snell (GCS) and randomized quantile (RQ) residuals:

$$\hat{r}_i^{\text{GCS}} = -\log(1 - F_Y(y_i;\widehat{\boldsymbol{\theta}})) \text{ and } \hat{r}_i^{\text{RQ}} = \Phi^{-1}(F_Y(y_i;\widehat{\boldsymbol{\theta}})), \ \ i = 1, \ldots, n, \tag{1.12}$$

where F_Y is quantile-based Singh-Maddala or Dagum CDF, and $\widehat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$. If the model is correctly specified, the GCS is asymptotically standard exponential distributed, while the RQ is asymptotically standard normal distributed. With both residuals, graphical techquines, such as quantile-quantile (QQ) plots with simulated envelope, can be used to assess distributions assumptions.

1.4 Monte Carlo simulation

In this section, we present Monte Carlo simulation studies for each reparameterized quantile model, considering different scenarios of parameters and sample sizes. The first part of the study consists in evaluating ML estimation performance, while the second evaluates the empirical distribution of the GCS and RQ residuals. Both studies consider simulated data generated from each one of the Singh-Maddala and Dagum quantile regression models according to

$$\gamma_i = \exp\left(\beta_0(\tau) + \beta_1(\tau)x_{1i} + \beta_2(\tau)x_{2i}\right).$$

1.4.1 ML estimation

The simulation scenario considers the following settings: sample sizes $n \in 50, 100, 200, 400$, vector of betas $\boldsymbol{\beta}(\tau) = (1, 0.5, 1.5)^\top$ and quantiles $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$, with 500 Monte Carlo replications for each sample size. For Singh-Maddala, three sets of true parameters for (a, q) were considered, given as $(1, 0.5), (2, 1)$ and $(5, 1)$. For Dagum, the following sets of true parameters for (a, p) were considered: $(1, 0.5), (1, 2)$ and $(2, 4)$. Values were chosen considering typical values found in real data. Covariate values x_{1i}, x_{2i} are obtained from a uniform distribution in the interval (0,1). To study the ML estimators, we use compute the bias, mean square error (MSE) and the coverage probability (CP). We expect that, as sample size increases, bias and MSE reduces, as CP approaches the 95% nominal level.

The results for Singh-Maddala models are shown in Tables 1.2, 1.3 and 1.4. It is possible to see that the simulations produced the expected outcomes. As the sample size increases, the bias and MSE both decrease, and the CP tends to 95%. It is also possible to observe that the MSE is smaller for $\tau = 0.50$ than for other quantiles. The results for Dagum models are shown in Tables 1.5, 1.6 and 1.7. These tables present results similar to those found for the Singh-Maddala models.

Table 1.2: Monte Carlo simulation results for the Singh-Maddala with $a = 1$ and $q = 0.5$.

γ		$n = 50$			$n = 100$			$n = 200$			$n = 400$		
		Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
Scenario 1: $(a, q) = (0.5, 1)$													
0.10	β_0	0.1121	0.8706	0.8978	0.0402	0.3955	0.930	0.0195	0.2115	0.934	0.0143	0.1026	0.946
	β_1	0.0512	1.3289	0.8958	0.0228	0.6821	0.930	0.0170	0.3367	0.932	0.0184	0.1499	0.942
	β_2	0.0688	1.2301	0.9100	0.0007	0.6696	0.936	0.0109	0.3175	0.942	0.0040	0.1601	0.942
	a	1.4695	60.8383	0.9432	0.0774	0.0642	0.954	0.0334	0.0239	0.938	0.0205	0.0087	0.958
	q	0.1427	2.0081	0.8283	0.0285	0.0426	0.910	0.0085	0.0152	0.922	0.0004	0.0060	0.944
0.25	β_0	0.0530	0.8841	0.9000	0.0370	0.3382	0.958	0.0138	0.1802	0.948	0.0139	0.0799	0.974
	β_1	0.0958	1.4415	0.8920	0.0400	0.6032	0.948	0.0345	0.3228	0.940	0.0057	0.1328	0.964
	β_2	0.0393	1.4047	0.8980	−0.0550	0.6814	0.926	0.0002	0.2727	0.964	0.0346	0.1514	0.948
	a	1.0751	53.6624	0.9399	0.0803	0.0739	0.954	0.0208	0.0221	0.932	0.0102	0.0085	0.960
	q	0.2120	4.0650	0.8465	0.0227	0.0413	0.928	0.0180	0.0158	0.930	0.0071	0.0065	0.950
0.5	β_0	0.0249	0.8318	0.9098	−0.0098	0.3541	0.964	0.0006	0.1747	0.948	0.0069	0.0947	0.938
	β_1	0.0032	1.3308	0.9178	0.0127	0.5385	0.978	0.0025	0.3168	0.942	0.0024	0.1439	0.966
	β_2	0.0365	1.3408	0.9180	−0.0129	0.6176	0.934	0.0103	0.3044	0.934	0.0234	0.1557	0.952
	a	0.9339	42.4393	0.9394	0.0883	0.0900	0.952	0.0317	0.0208	0.950	0.0096	0.0094	0.940
	q	0.1652	2.1179	0.8806	0.0196	0.0457	0.888	0.0146	0.0150	0.936	0.0131	0.0075	0.942
0.75	β_0	0.0045	0.8646	0.9260	−0.0075	0.4219	0.952	0.0184	0.1989	0.950	0.0146	0.1022	0.950
	β_1	0.0389	1.5429	0.9120	−0.0408	0.6431	0.930	0.0109	0.2746	0.954	0.0019	0.1536	0.942
	β_2	0.0133	1.3177	0.9440	−0.0116	0.5971	0.944	0.0370	0.3160	0.946	0.0066	0.1585	0.950
	a	0.5529	19.1928	0.9498	0.0766	0.0657	0.946	0.0306	0.0207	0.946	0.0075	0.0092	0.936
	q	0.1637	2.6273	0.8750	0.0269	0.0417	0.926	0.0143	0.0157	0.938	0.0083	0.0069	0.956
0.90	β_0	0.1410	1.2745	0.9180	−0.0491	0.5126	0.942	0.0260	0.2522	0.962	0.0220	0.1428	0.952
	β_1	0.0731	1.4873	0.9120	0.0242	0.6634	0.924	0.0057	0.2990	0.954	0.0164	0.1418	0.944
	β_2	0.0112	1.3225	0.9160	−0.0104	0.6772	0.932	0.0093	0.2954	0.950	0.0179	0.1635	0.952
	a	0.6327	18.5332	0.9372	0.0759	0.0627	0.944	0.0359	0.0191	0.964	0.0112	0.0084	0.938
	q	0.4203	12.5594	0.8630	0.0297	0.0403	0.918	0.0069	0.0133	0.936	0.0071	0.0067	0.948

Table 1.3: Monte Carlo simulation results for the Singh-Maddala with $a = 2$ and $q = 1$.

γ		$n = 50$			$n = 100$			$n = 200$			$n = 400$		
		Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
					Scenario 2: $(a, q) = (2, 1)$								
	β_0	0.0059	0.1275	0.9459	0.0154	0.0773	0.9080	0.0019	0.0284	0.964	−0.0101	0.0173	0.946
	β_1	0.0443	0.2001	0.9180	0.0118	0.1007	0.9240	0.0175	0.0506	0.944	0.0038	0.0248	0.932
0.10	β_2	0.0343	0.1962	0.9340	0.0008	0.1073	0.9240	0.0077	0.0467	0.948	0.0177	0.0253	0.936
	a	0.2290	0.9315	0.9194	0.1124	0.1858	0.9300	0.0478	0.0561	0.948	0.0142	0.0246	0.958
	q	2.0100	54.0823	0.8644	0.1998	1.0213	0.8976	0.0633	0.1346	0.932	0.0362	0.0415	0.962
	β_0	0.0228	0.1462	0.9178	0.0193	0.0602	0.9360	−0.0080	0.0253	0.956	−0.0071	0.0144	0.936
	β_1	0.0019	0.2425	0.9140	0.0146	0.0996	0.9500	0.0193	0.0431	0.964	0.0047	0.0226	0.948
0.25	β_2	−0.0079	0.2177	0.9218	−0.0215	0.0935	0.9519	−0.0035	0.0442	0.958	0.0082	0.0285	0.926
	a	0.1543	0.4407	0.9052	0.0654	0.1295	0.9618	0.0275	0.0492	0.954	0.0172	0.0236	0.958
	q	2.2468	63.2706	0.8584	0.3627	4.8904	0.9374	0.0752	0.1109	0.938	0.0275	0.0423	0.954
	β_0	0.0048	0.1302	0.9118	0.0035	0.0664	0.9180	0.0096	0.0254	0.952	−0.0033	0.0124	0.952
	β_1	0.0079	0.2139	0.9160	0.0037	0.1050	0.9280	−0.0057	0.0449	0.944	−0.0089	0.0215	0.948
0.5	β_2	−0.0092	0.2143	0.9078	0.0006	0.1014	0.9400	−0.0066	0.0446	0.930	0.0114	0.0227	0.952
	a	0.2757	0.7384	0.9133	0.0819	0.1608	0.9540	0.0459	0.0639	0.944	0.0329	0.0251	0.962
	q	1.6195	37.9874	0.8499	0.2646	1.2925	0.9158	0.0869	0.1403	0.938	0.0116	0.0381	0.944
	β_0	−0.0250	0.1202	0.9319	−0.0067	0.0605	0.9340	−0.0051	0.0297	0.944	0.0061	0.0121	0.966
	β_1	0.0131	0.1980	0.9198	−0.0063	0.0947	0.9460	−0.0051	0.0472	0.948	−0.0079	0.0214	0.950
0.75	β_2	−0.0033	0.2132	0.9160	0.0024	0.0858	0.9520	0.0047	0.0473	0.948	−0.0090	0.0235	0.952
	a	0.2563	0.7865	0.9257	0.0765	0.1367	0.9440	0.0419	0.0587	0.954	0.0124	0.0260	0.938
	q	1.7187	50.0852	0.8475	0.5295	13.6786	0.9217	0.0739	0.1479	0.902	0.0442	0.0470	0.960
	β_0	−0.0058	0.1585	0.9259	−0.0152	0.0713	0.9240	−0.0022	0.0302	0.952	−0.0036	0.0178	0.936
	β_1	−0.0198	0.2192	0.9140	0.0034	0.0938	0.9460	−0.0045	0.0443	0.952	0.0032	0.0241	0.938
0.90	β_2	−0.0116	0.2285	0.9157	0.0011	0.1018	0.9340	0.0014	0.0504	0.946	−0.0068	0.0255	0.940
	a	0.3917	22.5662	0.9091	0.0650	0.1464	0.9580	0.0303	0.0563	0.956	0.0361	0.0263	0.938
	q	2.2773	66.3719	0.8617	0.3925	10.8488	0.9317	0.0669	0.1140	0.958	0.0199	0.0389	0.946

Table 1.4: Monte Carlo simulation results for the Singh-Maddala with $a = 5$ and $q = 1$.

γ		$n = 50$			$n = 100$			$n = 200$			$n = 400$		
		Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
					Scenario 3: $(a, q) = (5, 1)$								
	β_0	0.0220	0.0244	0.9140	0.0150	0.0113	0.9260	0.0127	0.0058	0.918	0.0008	0.0026	0.952
	β_1	0.0031	0.0337	0.9200	−0.0075	0.0151	0.9360	0.0067	0.0077	0.940	0.0038	0.0039	0.940
0.10	β_2	0.0079	0.0305	0.9080	−0.0069	0.0147	0.9540	0.0044	0.0074	0.944	−0.0008	0.0036	0.944
	a	1.0053	80.0172	0.8992	0.2692	1.3922	0.9518	0.1137	0.3327	0.960	0.0541	0.1517	0.952
	q	2.3242	92.3992	0.8556	0.4202	17.9817	0.9054	0.0870	0.1941	0.930	0.0250	0.0399	0.936
	β_0	0.0097	0.0212	0.9038	0.0121	0.0095	0.9299	−0.0005	0.0047	0.952	−0.0026	0.0023	0.954
	β_1	−0.0001	0.0343	0.9078	−0.0126	0.0153	0.9399	0.0041	0.0085	0.920	0.0074	0.0038	0.958
0.25	β_2	−0.0048	0.0371	0.9040	−0.0083	0.0144	0.9380	0.0026	0.0076	0.942	−0.0002	0.0037	0.948
	a	0.9638	36.7433	0.9152	0.2751	1.0971	0.9558	0.0842	0.3294	0.954	0.0427	0.1768	0.944
	q	2.5581	95.2154	0.8386	0.3307	5.4881	0.9091	0.0921	0.1560	0.940	0.0358	0.0452	0.936
	β_0	0.0046	0.0200	0.9200	−0.0029	0.0079	0.9580	0.0011	0.0048	0.934	−0.0007	0.0023	0.946
	β_1	−0.0060	0.0342	0.9320	0.0089	0.0147	0.9480	0.0033	0.0074	0.950	0.0018	0.0037	0.952
0.5	β_2	−0.0048	0.0322	0.9280	−0.0019	0.0142	0.9440	−0.0044	0.0077	0.930	0.0010	0.0038	0.954
	a	0.4603	2.8850	0.9271	0.2101	1.0549	0.9420	0.1200	0.3555	0.962	0.0449	0.1463	0.956
	q	2.2190	87.5124	0.8894	0.2957	2.6461	0.9058	0.0534	0.0980	0.944	0.0312	0.0406	0.952
	β_0	−0.0008	0.0194	0.9237	0.0036	0.0089	0.9540	−0.0012	0.0043	0.952	−0.0028	0.0022	0.958
	β_1	−0.0191	0.0385	0.9096	−0.0112	0.0146	0.9540	0.0010	0.0071	0.956	0.0040	0.0036	0.956
0.75	β_2	0.0086	0.0300	0.9376	0.0045	0.0150	0.9420	−0.0028	0.0075	0.948	0.0009	0.0033	0.960
	a	0.8082	65.9950	0.9189	0.1383	0.8212	0.9359	0.1601	0.3927	0.952	0.0366	0.1661	0.946
	q	2.5054	85.8722	0.8550	0.2856	1.5773	0.9214	0.0306	0.0877	0.912	0.0356	0.0475	0.938
	β_0	−0.0234	0.0266	0.8940	−0.0062	0.0118	0.9400	−0.0084	0.0058	0.938	0.0004	0.0026	0.938
	β_1	−0.0052	0.0323	0.9360	0.0030	0.0181	0.9340	0.0021	0.0075	0.946	−0.0034	0.0032	0.948
0.90	β_2	0.0166	0.0354	0.9220	0.0016	0.0138	0.9620	0.0026	0.0080	0.948	−0.0049	0.0038	0.950
	a	1.0366	106.5326	0.9277	0.1346	0.8340	0.9560	0.1077	0.3554	0.936	0.0090	0.1434	0.940
	q	2.4482	95.2088	0.8604	0.3007	3.4418	0.9198	0.0779	0.1319	0.930	0.0568	0.0498	0.972

Table 1.5: Monte Carlo simulation results for the Dagum with $a = 1$ and $p = 0.5$.

γ		$n = 50$			$n = 100$			$n = 200$			$n = 400$		
		Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
Scenario 1: $(a, p) = (0.5, 1)$													
0.10	β_0	0.0294	1.1980	0.9251	−0.0025	0.5860	0.9380	0.0123	0.3049	0.938	0.0184	0.1454	0.948
	β_1	0.0071	1.5734	0.9038	0.0305	0.6246	0.9360	0.0230	0.3270	0.940	0.0080	0.1524	0.950
	β_2	0.0590	1.3702	0.9158	0.0748	0.6173	0.9400	0.0015	0.3126	0.944	0.0481	0.1484	0.958
	a	1.4569	87.2549	0.9293	0.0872	0.0877	0.9439	0.0290	0.0196	0.958	0.0129	0.0088	0.954
	q	0.3058	7.1869	0.8811	0.0241	0.0454	0.9038	0.0131	0.0153	0.934	0.0058	0.0067	0.940
0.25	β_0	0.0417	0.8489	0.9240	0.0677	0.3846	0.932	0.0555	0.1916	0.958	0.0109	0.0930	0.956
	β_1	0.0317	1.4268	0.9080	−0.0199	0.6839	0.942	0.0231	0.3119	0.940	0.0314	0.1454	0.962
	β_2	0.0115	1.3594	0.9220	−0.0349	0.5863	0.930	0.0451	0.3232	0.946	0.0021	0.1454	0.958
	a	0.5899	35.5220	0.9659	0.0796	0.0735	0.954	0.0324	0.0197	0.958	0.0224	0.0100	0.942
	q	0.1195	2.3333	0.8669	0.0306	0.0433	0.906	0.0144	0.0158	0.936	0.0022	0.0068	0.930
0.5	β_0	0.0568	0.8688	0.9060	0.0141	0.3643	0.940	0.0460	0.1655	0.954	0.0055	0.0839	0.954
	β_1	0.0298	1.4548	0.9140	−0.0329	0.5889	0.954	0.0175	0.3224	0.934	0.0016	0.1543	0.936
	β_2	0.0925	1.4459	0.9020	0.0444	0.6181	0.940	0.0962	0.3119	0.942	0.0108	0.1572	0.948
	a	1.9530	201.6857	0.9274	0.0945	0.0735	0.962	0.0378	0.0186	0.966	0.0145	0.0099	0.938
	q	0.3422	8.6680	0.8401	0.0147	0.0367	0.886	0.0030	0.0135	0.946	0.0078	0.0070	0.914
0.75	β_0	0.0794	0.7924	0.9060	−0.0109	0.3607	0.952	0.0189	0.1838	0.944	0.0060	0.0861	0.946
	β_1	0.0408	1.2941	0.9180	0.0096	0.6878	0.944	0.0136	0.3480	0.940	0.0037	0.1353	0.960
	β_2	0.0487	1.3624	0.9120	−0.0298	0.6079	0.952	0.0048	0.2777	0.958	0.0249	0.1455	0.950
	a	1.6779	155.4084	0.9475	0.0519	0.0524	0.952	0.0203	0.0205	0.932	0.0173	0.0090	0.946
	q	0.1303	1.1552	0.8508	0.0431	0.0685	0.928	0.0240	0.0187	0.944	0.0039	0.0062	0.946
0.90	β_0	0.0513	0.8453	0.9220	−0.0713	0.4561	0.9260	0.0541	0.2000	0.942	0.0232	0.1095	0.924
	β_1	0.0469	1.4430	0.9060	−0.0155	0.6397	0.9380	0.0719	0.2932	0.952	0.0028	0.1646	0.938
	β_2	0.0143	1.2423	0.9300	0.0351	0.6794	0.9240	0.0255	0.3240	0.942	0.0218	0.1500	0.950
	a	1.6653	297.6960	0.9439	0.1079	0.3700	0.9360	0.0325	0.0222	0.958	0.0152	0.0100	0.946
	q	0.2705	7.2677	0.8672	0.0454	0.1074	0.8878	0.0139	0.0150	0.942	0.0077	0.0076	0.948

Table 1.6: Monte Carlo simulation results for the Dagum with $a = 1$ and $p = 2$.

γ		$n = 50$			$n = 100$			$n = 200$			$n = 400$		
		Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
Scenario 2: $(a, p) = (0.5, 1)$													
0.10	β_0	0.0796	0.3382	0.9178	−0.0022	0.1689	0.9317	0.0339	0.0760	0.9500	0.0079	0.0371	0.9520
	β_1	−0.0183	0.5543	0.9198	0.0164	0.2568	0.9376	−0.0251	0.1331	0.9360	0.0006	0.0594	0.9520
	β_2	0.0261	0.5697	0.9100	0.0276	0.2955	0.9096	0.0042	0.1219	0.9420	0.0091	0.0539	0.9680
	a	0.0945	0.1451	0.8696	0.0109	0.0278	0.8224	0.0147	0.0115	0.9122	0.0061	0.0047	0.9571
	q	7.4914	264.7962	0.7084	3.4338	97.3523	0.8175	0.7256	8.1847	0.9006	0.1768	0.5421	0.9261
0.25	β_0	0.0651	0.3188	0.9240	0.0042	0.1533	0.9418	0.0223	0.0801	0.9400	0.0116	0.0341	0.9560
	β_1	−0.0688	0.5122	0.9339	0.0208	0.2358	0.9459	−0.0108	0.1263	0.9380	0.0084	0.0551	0.9500
	β_2	−0.0020	0.5229	0.9118	−0.0387	0.2387	0.9598	−0.0155	0.1381	0.9240	−0.0218	0.0628	0.9420
	a	0.0683	0.0658	0.8791	0.0221	0.0236	0.9047	0.0133	0.0110	0.9234	0.0053	0.0048	0.9547
	q	7.4426	263.2996	0.7494	1.8410	39.7247	0.8661	0.5824	6.0717	0.8948	0.1637	0.4527	0.9440
0.5	β_0	−0.0120	0.3052	0.9337	0.0042	0.1533	0.9418	0.0005	0.0770	0.9339	0.0036	0.0377	0.9420
	β_1	0.0110	0.5548	0.9319	0.0208	0.2358	0.9459	−0.0014	0.1215	0.9459	−0.0022	0.0600	0.9540
	β_2	0.0162	0.5112	0.9294	−0.0387	0.2387	0.9598	−0.0146	0.1188	0.9419	−0.0026	0.0601	0.9480
	a	0.0895	0.1032	0.8808	0.0221	0.0236	0.9047	0.0157	0.0106	0.9186	0.0031	0.0051	0.9402
	q	7.2823	278.8213	0.7405	1.8410	39.7247	0.8661	0.6285	5.8665	0.9174	0.2307	0.7891	0.9271
0.75	β_0	−0.0315	0.3760	0.8898	−0.0288	0.1656	0.9420	−0.0176	0.0747	0.9519	−0.0105	0.0355	0.9400
	β_1	0.0148	0.5461	0.9138	0.0315	0.2502	0.9438	0.0304	0.1122	0.9479	−0.0025	0.0512	0.9680
	β_2	−0.0009	0.5899	0.8976	0.0203	0.2742	0.9260	0.0021	0.1159	0.9578	0.0058	0.0543	0.9540
	a	0.0992	0.1544	0.8672	0.0220	0.0300	0.8718	0.0061	0.0126	0.9052	0.0111	0.0059	0.9348
	q	7.6120	296.0891	0.7005	2.4525	55.8594	0.8676	0.7319	8.5158	0.9213	0.1721	0.5829	0.9243
0.90	β_0	−0.0671	0.4158	0.9076	−0.0318	0.1970	0.9379	−0.0254	0.1029	0.9218	−0.0143	0.0494	0.9499
	β_1	0.0011	0.5499	0.8976	0.0124	0.2502	0.9500	−0.0047	0.1172	0.9500	−0.0006	0.0651	0.9400
	β_2	0.0309	0.5538	0.9217	0.0129	0.2524	0.9399	0.0240	0.1063	0.9720	0.0138	0.0627	0.9520
	a	0.0785	0.1009	0.8841	0.0198	0.0252	0.8879	0.0156	0.0110	0.9247	0.0086	0.0054	0.9367
	q	7.8836	334.5644	0.7218	3.2214	104.7364	0.8667	0.5753	4.9575	0.9278	0.1697	0.4890	0.9218

Table 1.7: Monte Carlo simulation results for the Dagum with $a = 2$ and $p = 4$.

γ		$n = 50$			$n = 100$			$n = 200$			$n = 400$		
		Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
Scenario 3: $(a, p) = (2, 4)$													
	β_0	0.0338	0.0586	0.9280	0.0093	0.0310	0.9280	−0.0033	0.0139	0.9420	0.0033	0.0067	0.9500
	β_1	0.0019	0.0911	0.9420	0.0040	0.0553	0.9198	0.0077	0.0240	0.9320	−0.0057	0.0117	0.9460
0.10	β_2	−0.0168	0.0926	0.9337	0.0018	0.0505	0.9238	0.0131	0.0225	0.9420	0.0037	0.0107	0.9600
	a	0.1593	0.2064	0.9374	0.0414	0.0782	0.8465	0.0247	0.0382	0.8351	0.0062	0.0195	0.8041
	q	14.2420	580.1449	0.4824	9.4429	331.8145	0.6199	5.2306	181.5796	0.6988	1.7060	36.9713	0.7672
	β_0	0.0152	0.0535	0.9419	0.0221	0.0274	0.9360	0.0013	0.0137	0.9399	0.0031	0.0071	0.9619
	β_1	0.0108	0.0967	0.9297	−0.0071	0.0510	0.9120	0.0072	0.0233	0.9459	−0.0004	0.0112	0.9600
0.25	β_2	−0.0169	0.0916	0.9357	−0.0060	0.0531	0.9238	−0.0022	0.0229	0.9400	−0.0032	0.0115	0.9560
	a	0.1354	0.2213	0.9299	0.0781	0.0915	0.8705	0.0323	0.0360	0.8403	0.0065	0.0201	0.7895
	q	14.1197	563.1669	0.5139	8.8008	342.9496	0.6250	4.1960	143.2611	0.7759	1.5172	27.7484	0.7964
	β_0	0.0101	0.0584	0.9220	−0.0050	0.0274	0.9539	−0.0009	0.0140	0.9420	0.0006	0.0059	0.9580
	β_1	−0.0055	0.0963	0.9140	0.0093	0.0510	0.9200	−0.0070	0.0241	0.9220	0.0091	0.0108	0.9420
0.5	β_2	−0.0133	0.1017	0.9200	0.0002	0.0468	0.9339	0.0059	0.0246	0.9400	−0.0058	0.0101	0.9639
	a	0.1515	0.2937	0.9041	0.0660	0.0768	0.8887	0.0383	0.0408	0.8584	0.0097	0.0180	0.8337
	q	14.9605	621.5463	0.5129	8.2909	306.5194	0.6462	4.4736	161.4415	0.7660	1.5362	29.5129	0.8207
	β_0	−0.0123	0.0695	0.9340	−0.0087	0.0308	0.9479	0.0043	0.0154	0.9300	−0.0083	0.0075	0.9560
	β_1	0.0005	0.1031	0.9259	0.0016	0.0482	0.9400	−0.0108	0.0243	0.9300	0.0086	0.0121	0.9400
0.75	β_2	−0.0141	0.1125	0.9058	−0.0141	0.0453	0.9459	−0.0019	0.0253	0.9218	0.0010	0.0109	0.9460
	a	0.1727	0.3346	0.9261	0.0575	0.0730	0.8879	0.0227	0.0410	0.8040	0.0062	0.0195	0.8172
	q	14.8322	607.1782	0.4954	9.4776	373.5821	0.6648	5.2483	174.3589	0.7050	1.8124	31.4739	0.7807
	β_0	−0.0380	0.0896	0.9034	−0.0094	0.0451	0.9177	−0.0040	0.0195	0.9517	0.0040	0.0104	0.9380
	β_1	0.0186	0.1161	0.9177	0.0083	0.0519	0.9277	0.0057	0.0201	0.9579	−0.0011	0.0119	0.9320
0.90	β_2	−0.0287	0.1070	0.9138	−0.0085	0.0540	0.9158	−0.0084	0.0237	0.9418	−0.0029	0.0122	0.9320
	a	0.1729	0.2768	0.9248	0.0537	0.1027	0.8491	0.0189	0.0318	0.8442	0.0012	0.0162	0.8462
	q	14.3650	630.0775	0.5225	9.8336	400.5561	0.5989	4.8901	179.3580	0.7324	1.9847	44.6904	0.8132

1.4.2 Empirical distribution of residuals

Here we show the performance of GCS and RQ residuals. We analyse the results with descriptive statistics (mean, median, standard deviation, coefficient of skewness and coefficient of kurtosis); see Tables 1.8 and 1.9 for the results of the Singh-Maddala and Dagum models, respectively.

The reference values of mean, median, Sd, skewness and kurtosis are 1, 0.69, 1, 2 and 6, respectively, for GCS residual, and 0, 0, 1, 0 and 0, respectively, for RQ residual. It is possible to verify that, as the sample size increases, the values tend to the expected results for each τ. Therefore, we can use the both residuals to verify the fit of the proposed models.

Table 1.8: Residual statistics for Singh-Maddala Monte Carlo simulations.

n		0.1		0.25		0.5		0.75		0.9	
		GCS	RQ	GCS	RQ	GCS	RQ	GCS	RQ	GCS	RQ
Scenario 1: $(a, q) = (0.5, 1)$											
$n = 50$	Mean	1.0000	0.0018	1.0000	0.0014	1.000	0.002	1.0001	0.0024	1.0000	0.0017
	Median	0.7013	0.0065	0.6979	0.0028	0.6917	0.0050	0.6989	0.0037	0.6993	0.0043
	Sd	0.9929	1.0081	0.9947	1.0079	0.9885	1.0076	0.9891	1.0065	0.9937	1.0079
	Skewness	1.6515	0.0224	1.6536	0.0097	1.5961	0.0211	1.6110	0.0105	1.6510	0.0163
	Kurtosis	3.1795	0.2304	3.1896	0.2503	2.7743	0.2814	2.9559	0.3034	3.1531	0.2384
$n = 100$	Mean	1.0000	0.0008	1.0000	0.0002	1.0000	0.0004	1.0000	0.0009	1.0000	0.0007
	Median	0.6974	0.0040	0.6922	−0.0025	0.7001	0.0072	0.6971	0.0034	0.6945	0.0003
	Sd	0.9969	1.0035	0.9996	1.0045	0.9995	1.0045	0.9971	1.0037	0.9980	1.0038
	Skewness	1.7996	−0.0008	1.8044	0.0005	1.8159	−0.0039	1.8020	−0.0052	1.8082	−0.0022
	Kurtosis	4.2238	−0.1470	4.1943	−0.1262	4.2526	−0.1108	4.2592	−0.1253	4.2570	−0.1302
$n = 200$	Mean	1.000	0.0006	1.000	0.0001	1.0000	0.0001	1.0000	0.0004	1.0000	0.0002
	Median	0.6967	0.0038	0.6940	0.0004	0.6932	0.0006	0.6977	0.0051	0.6937	0.0000
	Sd	0.9963	1.0016	0.9980	1.0023	1.0003	1.0024	0.9974	1.0020	1.0025	1.0027
	Skewness	1.8654	0.0034	1.8680	0.0022	1.9028	0.0023	1.8694	0.0036	1.9224	0.0021
	Kurtosis	4.7536	0.0847	4.7035	0.0775	5.0575	0.0469	4.7309	0.0734	5.2122	0.0403
$n = 400$	Mean	1.0000	0.0004	1.000	0.0001	1.0000	0.0001	1.0000	0.0002	1.0000	0.0003
	Median	0.6935	0.0001	0.6940	0.0007	0.6925	0.0011	0.6938	0.0005	0.6938	0.0005
	Sd	0.9979	1.0006	0.9994	1.0010	0.9998	1.0011	0.9990	1.0009	0.9980	1.0007
	Skewness	1.9308	0.0013	1.9405	0.0008	1.9485	0.0004	1.9357	0.0004	1.9255	0.0005
	Kurtosis	5.3601	0.0464	5.4317	0.0403	5.5376	0.0349	5.3534	0.0422	5.2617	0.0498
Scenario 2: $(a, q) = (2, 1)$											
$n = 50$	Mean	0.9999	0.0007	0.9999	0.0010	0.9999	0.0003	1.0000	0.0007	1.0000	0.0005
	Median	0.6947	−0.0011	0.6929	−0.0033	0.6953	−0.0007	0.6901	−0.0070	0.6950	−0.0007
	Sd	0.9973	1.0087	0.9935	1.0076	0.9999	1.0093	0.9981	1.0083	0.9993	1.0085
	Skewness	1.6709	−0.0017	1.6184	0.0036	1.7033	−0.0007	1.6740	0.0049	1.6712	0.0072
	Kurtosis	3.1842	−0.2574	2.8183	−0.3035	3.3739	−0.2379	3.1942	−0.2667	3.1597	−0.2583
$n = 100$	Mean	1.0000	0.0003	1.0000	0.0007	1.0000	0.0006	1.0000	0.0003	1.0000	0.0004
	Median	0.6923	−0.0024	0.6965	0.0028	0.6921	−0.0026	0.6934	−0.0011	0.6949	0.0009
	Sd	0.9992	1.0044	0.9963	1.0034	0.9976	1.0034	0.9986	1.0041	0.9976	1.0037
	Skewness	1.8166	0.0003	1.7803	0.0052	1.7992	0.0061	1.8079	0.0033	1.7921	0.0048
	Kurtosis	4.2162	−0.1299	3.9743	−0.1723	4.1587	−0.1612	4.1613	−0.1445	4.0383	−0.1594
$n = 200$	Mean	1.0000	0.0004	1.0000	0.0001	1.0000	0.0003	1.0000	0.0002	1.0000	0.0002
	Median	0.6951	0.0018	0.6897	−0.0050	0.6924	−0.0016	0.6917	−0.0025	0.6902	−0.0044
	Sd	0.9975	1.0017	0.9997	1.0022	0.9992	1.0018	0.9985	1.0021	0.9983	1.0019
	Skewness	1.8743	0.0018	1.9035	0.0004	1.9019	0.0027	1.8801	−0.0011	1.8757	0.0004
	Kurtosis	4.8043	−0.0908	5.0267	−0.0656	5.0307	−0.0741	4.8043	−0.0664	4.7595	−0.0776
$n = 400$	Mean	1.0000	0.0003	1.0000	0.0002	1.0000	0.0001	1.0000	0.0001	1.0000	−0.0001
	Median	0.6937	0.0004	0.6918	−0.0021	0.6940	0.0007	0.6920	−0.0018	0.6939	0.0007
	Sd	0.9975	1.0006	0.9983	1.0009	0.9998	1.0012	0.9992	1.0012	1.0006	1.0014
	Skewness	1.9111	0.0004	1.9224	−0.0012	1.9468	−0.0016	1.9365	−0.0028	1.9519	−0.0009
	Kurtosis	5.1005	−0.0554	5.2040	−0.0424	5.4577	−0.0253	5.3639	−0.0270	5.4123	−0.0211
Scenario 3: $(a, q) = (5, 1)$											
$n = 50$	Mean	1.0000	0.0008	0.9999	0.0009	1.0000	0.0009	0.9999	0.0009	0.9999	0.0007
	Median	0.6971	0.0017	0.6982	0.0033	0.6969	0.0015	0.6958	0.0005	0.6972	0.0020
	Sd	0.9957	1.0083	0.9948	1.0085	0.9953	1.0076	0.9940	1.0077	0.9965	1.0083
	Skewness	1.6576	0.0003	1.6465	−0.0059	1.6524	0.0068	1.6322	0.0046	1.6551	0.0028
	Kurtosis	3.0795	0.2799	3.0374	−0.2695	3.0727	−0.3022	2.898	−0.308	3.0403	−0.2767
$n = 100$	Mean	1.0000	0.0008	1.0000	0.0006	1.0000	0.0004	1.0000	0.0004	1.0000	0.0002
	Median	0.6935	−0.0010	0.6939	−0.0004	0.6967	0.0031	0.6929	−0.0017	0.6947	0.0005
	Sd	0.9956	1.0036	0.9970	1.0033	0.9988	1.0039	0.9993	1.0041	0.9995	1.0048
	Skewness	1.7737	−0.0034	1.7968	0.0071	1.8140	0.0038	1.8166	0.0019	1.8147	−0.0018
	Kurtosis	3.9289	−0.1425	4.1221	−0.1741	4.2349	−0.1394	4.2652	−0.1286	4.2104	−0.1198
$n = 200$	Mean	1.0000	0.0004	1.0000	−0.0001	1	0	1.0000	0.0002	1.0000	0.0001
	Median	0.6924	−0.0016	0.6947	0.0013	0.6922	−0.0019	0.6931	−0.0008	0.6934	−0.0004
	Sd	0.9977	1.0012	1.0018	1.0026	1.0007	1.0026	0.9998	1.0020	0.9998	1.0021
	Skewness	1.8822	0.0071	1.9305	0.0013	1.9171	−0.0010	1.9035	0.0021	1.8989	0.0028
	Kurtosis	4.8935	−0.1072	5.2166	−0.0483	5.1201	−0.0514	5.0378	−0.0679	4.9661	−0.0731
$n = 400$	Mean	1.0000	0.0003	1.0000	0.0003	1.0000	0.0002	1.0000	0.0002	1.0000	0.0002
	Median	0.6937	0.0003	0.6934	0.0000	0.6940	0.0007	0.6928	−0.0007	0.6927	−0.0008
	Sd	0.9980	1.0006	0.9981	1.0007	0.9989	1.0008	0.9988	1.0008	0.9992	1.0007
	Skewness	1.9260	0.0018	1.9295	0.0002	1.9321	0.0013	1.9375	0.0014	1.9408	0.0022
	Kurtosis	5.2908	−0.0573	5.3495	−0.0493	5.2939	−0.0451	5.3932	−0.0493	5.4139	−0.0436

Table 1.9: Residual statistics for Dagum Monte Carlo simulations.

n		0.1		0.25		0.5		0.75		0.9	
		GCS	RQ	GCS	RQ	GCS	RQ	GCS	RQ	GCS	RQ
Scenario 1: $(a, p) = (0.5, 1)$											
$n = 50$	Mean	0.9987	0.0006	0.9986	0.0006	1.0001	0.0021	0.9983	0.0008	1.0000	0.0038
	Median	0.6924	0.0033	0.6913	0.0054	0.6918	0.0032	0.6918	0.0030	0.6940	0.0009
	Sd	1.0049	1.0076	0.9995	1.0081	1.0097	1.0084	1.0062	1.0071	1.0074	1.0089
	Skewness	1.7158	0.0295	16.762	0.0093	1.7387	0.0272	1.7372	0.0273	1.7149	0.0156
	Kurtosis	3.411	0.260	3.1572	0.2527	3.5641	0.2138	3.5596	0.2255	3.3486	0.2127
$n = 100$	Mean	0.9989	-0.0008	0.9993	-0.0008	0.9991	-0.0006	0.9997	0.0000	0.9985	-0.0010
	Median	0.6926	-0.0021	0.6951	0.0011	0.6879	-0.0080	0.6903	-0.0050	0.6927	-0.0021
	Sd	0.9986	1.0038	1.0018	1.0042	0.9988	1.0040	0.9980	1.0046	0.9965	1.0034
	Skewness	1.8072	0.0050	1.8376	0.0106	1.8127	0.0030	1.7910	-0.0058	1.7973	0.0033
	Kurtosis	4.1823	-0.1393	4.4115	-0.1314	4.2703	-0.1345	4.0338	-0.1181	4.1396	-0.1475
$n = 200$	Mean	0.9998	0.0002	0.9996	0.0003	1.0001	0.0002	0.9997	0.0003	0.9990	0.0005
	Median	0.6922	0.0019	0.6926	0.0013	0.6943	0.0008	0.6926	0.0014	0.6926	0.0014
	Sd	1.0001	1.0022	0.9991	1.0020	0.9993	1.0025	0.9990	1.0021	0.9968	1.0015
	Skewness	1.9058	0.0019	1.8915	0.0018	1.8917	0.0049	1.8844	0.0032	1.8790	0.0007
	Kurtosis	5.0530	0.0642	4.9261	0.0748	4.9281	0.0548	4.8232	0.0829	4.8702	0.0782
$n = 400$	Mean	0.9999	0.0000	0.9996	0.0001	1.0000	0.0000	0.9997	0.0002	0.9994	0.0005
	Median	0.6936	0.0003	0.6928	0.0008	0.6940	0.0007	0.6927	0.0010	0.6940	0.0008
	Sd	0.9992	1.0011	0.9973	1.0008	0.9996	1.0012	0.9993	1.0009	0.9992	1.0006
	Skewness	1.9408	0.0022	1.9086	0.0021	1.9362	0.0008	1.9348	0.0015	1.9516	0.0036
	Kurtosis	5.4142	0.0301	5.0752	0.0487	5.3120	0.0325	5.3081	0.0412	5.5609	0.0417
Scenario 2: $(a, p) = (1, 2)$											
$n = 50$	Mean	1.0009	0.0001	0.9993	-0.0001	0.9998	-0.0001	0.9998	-0.0002	1.0005	0.0000
	Median	0.7000	0.0051	0.7021	0.0081	0.7022	0.0079	0.6927	-0.0040	0.6996	0.0048
	Sd	1.0006	1.0100	0.9894	1.0085	0.994	1.009	0.9948	1.0091	0.9992	1.0098
	Skewness	1.6550	0.0112	1.5900	-0.0062	1.6206	-0.0004	1.6094	0.0018	1.6446	0.0055
	Kurtosis	3.2276	-0.2941	2.8009	-0.3398	3.0132	-0.3114	2.9064	-0.3095	3.0804	-0.2858
$n = 100$	Mean	1.0001	0.0000	0.9994	0.0000	0.9992	-0.0003	0.9988	-0.0002	0.9994	-0.0001
	Median	0.6954	0.0014	0.6933	-0.0014	0.6972	0.0035	0.6959	0.0018	0.6965	0.0027
	Sd	1.0003	1.0048	0.9933	1.0041	0.9964	1.0041	0.9905	1.0036	0.9941	1.0042
	Skewness	1.8142	0.0012	1.7473	-0.0089	1.7875	-0.0051	1.7339	-0.0126	1.7610	-0.0084
	Kurtosis	4.2819	-0.1285	3.8424	-0.1614	4.1362	-0.1410	3.7739	-0.1724	3.9694	-0.1557
$n = 200$	Mean	0.9991	-0.0002	0.9993	-0.0001	1.0001	0.0000	0.9995	-0.0001	0.9986	-0.0003
	Median	0.6934	-0.0003	0.6980	0.0053	0.6939	0.0001	0.6978	0.0051	0.6946	0.0010
	Sd	0.9947	1.0016	0.9952	1.0018	1.0029	1.0026	0.9964	1.0019	0.9901	1.0010
	Skewness	1.8581	-0.0071	1.8568	-0.0071	1.9258	0.0024	1.8636	-0.0062	1.8250	-0.0138
	Kurtosis	4.7315	-0.0890	4.6862	-0.0845	5.2870	-0.0396	4.7281	-0.0736	4.5162	-0.1077
$n = 400$	Mean	0.9996	-0.0001	0.9997	-0.0001	0.9996	-0.0001	0.9990	-0.0003	0.9997	-0.0001
	Median	0.6956	0.0027	0.6930	-0.0005	0.6939	0.0005	0.6946	0.0014	0.6939	0.0006
	Sd	0.9975	1.0008	0.9977	1.0009	0.9976	1.0008	0.9949	1.0002	0.9991	1.0009
	Skewness	1.9307	-0.0035	1.9226	-0.0034	1.9270	-0.0025	1.9270	-0.0059	1.9514	-0.0014
	Kurtosis	5.3197	-0.0427	5.2584	-0.0421	5.2883	-0.0469	5.3821	-0.0599	5.5624	-0.0346
Scenario 3: $(a, p) = (2, 4)$											
$n = 50$	Mean	1.0059	0.0014	1.0061	0.0015	1.0047	0.0013	1.0049	0.0014	1.0039	0.0011
	Median	0.6955	-0.0004	0.6917	-0.0055	0.6879	-0.0100	0.6926	-0.0044	0.6905	-0.0069
	Sd	1.0139	1.0137	1.0144	1.0139	1.0074	1.0127	1.0113	1.0130	1.0047	1.0122
	Skewness	1.6462	0.0487	1.6223	0.0457	1.6121	0.0338	1.6511	0.0367	1.6107	0.0272
	Kurtosis	3.0694	-0.3480	2.9151	-0.3301	2.9161	-0.3471	3.1137	-0.3180	2.8872	-0.3450
$n = 100$	Mean	1.0008	0.0004	1.0023	0.0005	1.0012	0.0004	1.0008	0.0003	1.0012	0.0004
	Median	0.6946	0.0002	0.6931	-0.0018	0.6940	-0.0006	0.6962	0.0021	0.6936	-0.0011
	Sd	0.9955	1.0050	1.0068	1.0064	0.9999	1.0055	0.9974	1.0051	0.9995	1.0054
	Skewness	1.7423	0.0035	1.8133	0.0199	1.7731	0.0063	1.7640	0.0049	1.7721	0.0073
	Kurtosis	3.8270	-0.2055	4.2793	-0.1516	4.0014	-0.1671	3.9665	-0.1897	4.0262	-0.1760
$n = 200$	Mean	1.0003	0.0002	0.9999	0.0001	1	0	1.0000	0.0001	0.9996	-0.0001
	Median	0.6959	0.0025	0.6921	-0.0021	0.6965	0.0034	0.6934	-0.0006	0.6940	0.0003
	Sd	0.9990	1.0024	0.9975	1.0022	1.0008	1.0023	0.9976	1.0023	0.9963	1.0019
	Skewness	1.8890	0.0014	1.8751	-0.0017	1.9183	0.0022	1.8684	-0.0016	1.8748	-0.0037
	Kurtosis	5.0242	-0.0870	4.8970	-0.0859	5.2448	-0.0681	4.8039	-0.0875	4.8971	-0.0907
$n = 400$	Mean	0.9999	0.0000	0.9996	0.0000	0.9999	0.0000	0.9994	0.0000	0.9996	-0.0001
	Median	0.6942	0.0009	0.6937	0.0002	0.6926	-0.0011	0.6958	0.0028	0.6952	0.0021
	Sd	0.9982	1.0010	0.9972	1.0008	0.9989	1.0011	0.9950	1.0006	0.9976	1.0008
	Skewness	1.9236	-0.0024	1.9277	-0.0042	1.9335	-0.0017	1.9138	-0.0077	1.9327	-0.0029
	Kurtosis	5.2610	-0.0411	5.3413	-0.0462	5.3322	-0.0357	5.2572	-0.0551	5.4142	-0.0476

1.5 Application to income data

We use the 2016 Chilean household income data set, provided by the National Institute of Statistics in Chile[1] to illustrate the proposed parametric quantile regression models. This data set was also used by Sánchez et al. (2021b), who introduced the Birnbaum-Saunders quantile regression. While the Birnbaum-Saunders is not a distribution commonly used for income data, Singh-Maddala and Dagum are, so we assess if these models can produce better fits than the BS model.

The household income is the response variable (T), whereas the covariates are the total income due to salaries (X_1), the total income due to independent work (X_2) and the total income due to retirements (X_3). The original dataset contains 107 variables, including the aforementioned, but these were selected based on economic and statistical criteria in relation to the response variable and descriptive analysis conducted by Sánchez et al. (2021b). Moreover, all incomes are expressed in thousands of Chilean pesos[2].

We report in Table 1.10 descriptive statistics for the household income (T). Figure 1.3 shows the histogram along with usual and adjusted box plots (Rousseeuw et al., 2016). We observe that the household income data have a unimodal and right-skewed behavior, which i the precise needed scenario to uphold the usage of asymmetric distribution. Figure 1.4 shows scatterplots (with correlation) between the household income (T) and the covariates (X_1, X_2 and X_3). We observe that correlations are reasonable and significant, meanwhile the covariates have almost no linear correlation between each other.

Table 1.10: Descriptive statistics for the household income data (in thousands of Chilean pesos).

Mean	Median	Sd	Coef. Variation	Coef. Skewness	Coef. Kurtosis	minimum	maximum	n
698.80	938.10	837.52	0.89	2.45	11.03	70	5369.90	100

[1] Available at https://www.ine.cl/estadisticas/sociales/ingresos-y-gastos/encuesta-suplementaria-de-ingresos.

[2] See http://www.bancocentral.cl for their equivalence in American dollars.

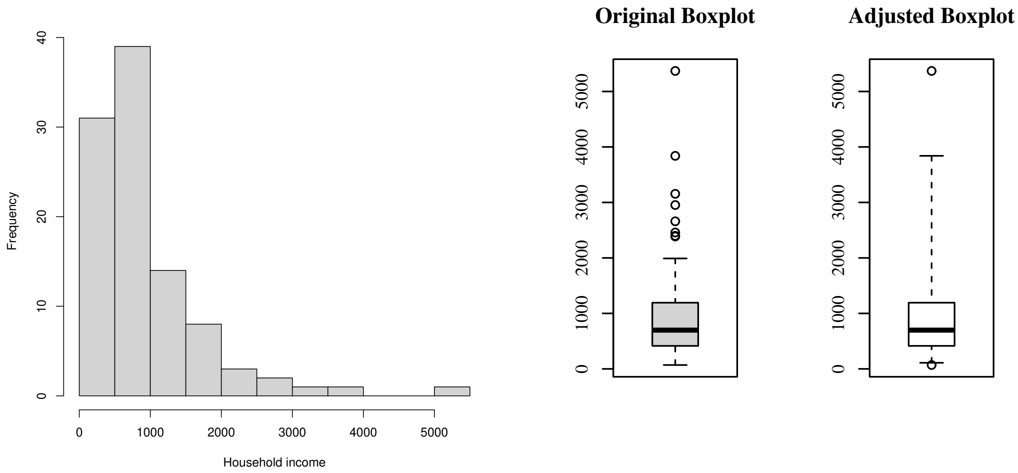


Figure 1.3: Histogram and boxplots for the household income data (in thousands of Chilean pesos).

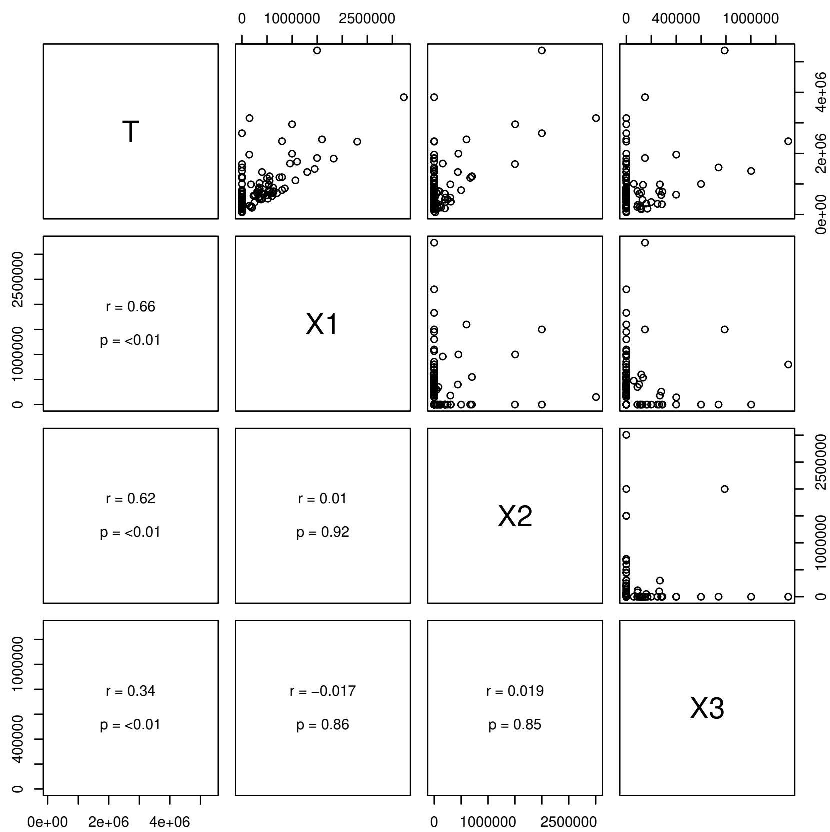


Figure 1.4: Scatterplots and correlations between variables T, X_1, X_2 and X_3.

We then analyze the household income data using the Singh-Maddala and Dagum quantile regression models, with regression structure expressed as

$$\gamma_i = \exp\left(\beta_0(\tau) + \beta_1(\tau)x_{1i} + \beta_2(\tau)x_{2i} + \beta_3(\tau)x_{3i}\right),$$

for $i, 1, \ldots, 100$. The proposed models are fitted using the function `IncomeReg.fit`, implemented in the `R` software by the authors. The codes are available upon request.

Table 1.11 presents the ML estimates, computed by the BFGS quasi-Newton method, standard errors (SEs) and Akaike (AIC) and Bayesian information (BIC) criteria values, for the Singh-Maddala and Dagum quantile regression models with $\tau = 0.50$. As mentioned earlier, the results of the Birnbaum-Saunders quantile regression are presented as well. The results of Table 1.11 reveal that the proposed Singh-Maddala and Dagum models provide better adjustments than the Birnbaum-Saunders model based on the values of log-likelihood, AIC and BIC. Particularly, the Singh-Maddala model has the lowest AIC and BIC values. The QQ plots with simulated envelope of the GCS and RQ residuals for the models considered in Table 1.11 confirm the results presented in Table 1.11; see Figure 1.5. Similar results are obtained when considering $\tau = \{0.10, \ldots, 0.90\}$.

1.6 Concluding remarks

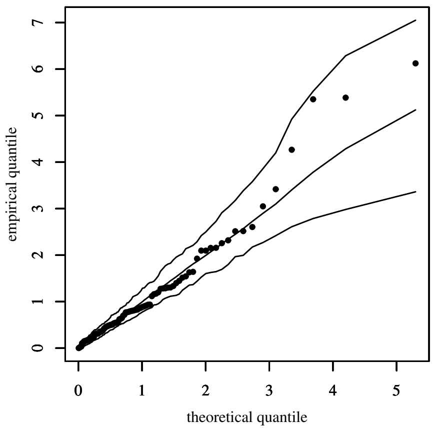
In this paper, we have proposed parametric quantile regression models based on the Singh-Maddala and Dagum regressions. The proposed models are based on reparametrizations of the original distributions, by including the distribution quantile as a parameter. The maximum likelihood method was used to estimate model parameters and Monte Carlo simulation studies were conducted in order to evaluate the performance of the estimators and the empirical distribution of the generalized Cox-Snell and random quantile residuals. Results showed that the estimates had good performance, and the residuals presented good agreement with their reference distributions. We applied the proposed models to a real data set, where we have modeled

Table 1.11: ML estimates (with standard errors in parentheses) and model selection measures for the income data.

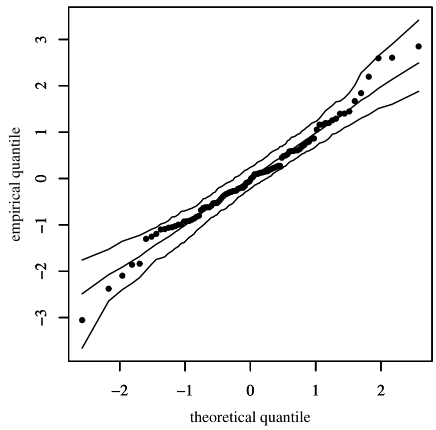
	Birnbaum-Saunders ($\tau = 0.50$)	Dagum ($\tau = 0.50$)	Singh-Maddala ($\tau = 0.50$)
β_0	198.0903*	150.8307*	137.8478*
	(22.3166)	(3.0771)	(3.2826)
β_1	1.0440*	1.1173*	1.1252*
	(0.0870)	(0.0636)	(0.0569)
β_2	1.1090*	1.2424*	1.2805*
	(0.1502)	(0.1172)	(0.1103)
β_3	1.0865*	1.1562*	1.1730*
	(0.1759)	(0.1395)	(0.1382)
α	0.3646*		
	(0.0087)		
a		4.3720	8.3380
		(0.5692)	(1.4720)
p or q		2.2100	0.4034
		(1.0290)	(0.1144)
Log-lik.	-692.8373	-686.9182	-685.2337
AIC	1395.675	1385.836	1382.467
BIC	1408.701	1386.740	1383.371

* significant at 5% level. ** significant at 10% level.

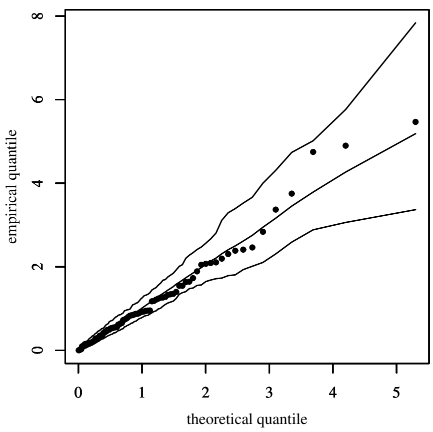
the household income as a function of the following covariates: total income due to salaries, total income due to independent work and total income due to retirements. The results were compared to those obtained by Sánchez et al. (2021b), who proposed the Birnbaum-Saunders quantile regression model. We showed that both Singh-Maddala and Dagum models have better fit to data than Birnbaum-Saunders model, with Singh-Maddala also showing a slight superior performance than Dagum. Therefore, results were favorable to the usage of Singh-Maddala and Dagum quantile regression models. As part of future research, it will be of interest to introduce zero-inflated quantile models. In addition, influence diagnostic tools can be investigated and also multivariate models can be studied. Work on these problems is currently in progress and we hope to report these findings in future papers.



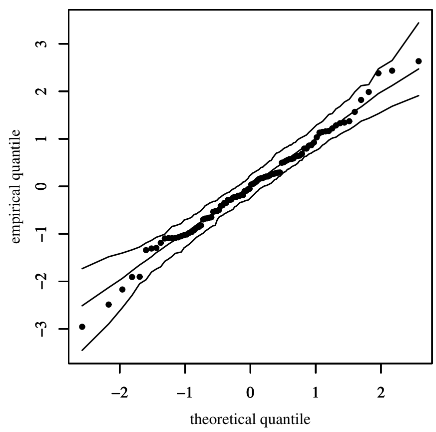
(a) Dagum (GCS)



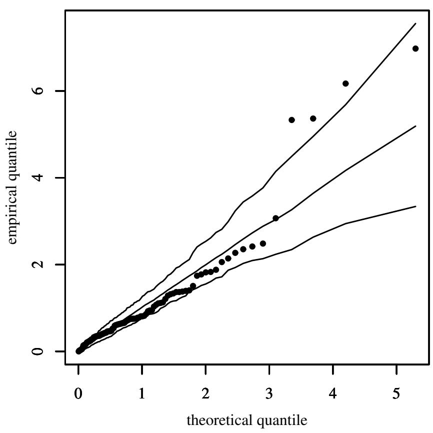
(b) Dagum (RQ)



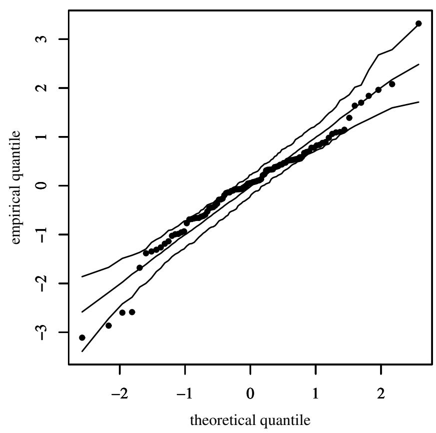
(c) Singh-Maddala (GCS)



(d) Singh-Maddala (RQ)



(e) Birnbaum-Saunders (GCS)



(f) Birnbaum-Saunders (RQ)

Figure 1.5: QQ plot and its envelope for the GSC and RQ residuals in the indicated models for the income data ($\tau = 0.50$).

Appendix A

Proofs

A.1 Proofs of Subsection 1.2.2

Proof of Lemma 1. The Gini coefficient is a measure of statistical dispersion and is defined as the half of the relative mean absolute difference:

$$G = \frac{1}{2\mu}\int_0^\infty \int_0^\infty f(x)f(y)|x-y|\,\mathrm{d}x\mathrm{d}y.$$

This measure is written as

$$G = \frac{1}{2\mu}\int_0^\infty \int_y^\infty f(x)f(y)(x-y)\,\mathrm{d}x\mathrm{d}y - \frac{1}{2\mu}\int_0^\infty \int_0^y f(x)f(y)(x-y)\,\mathrm{d}x\mathrm{d}y.$$

Equivalently,

$$\begin{aligned} G &= \frac{1}{2\mu}\int_0^\infty \big[\mathbb{E}(X\mathbb{1}_{\{X>y\}}) - y\mathbb{P}(X>y) - \mathbb{E}(X\mathbb{1}_{\{X\leqslant y\}}) + y\big] f(y)\,\mathrm{d}y \\ &= \frac{1}{\mu}\int_0^\infty \big[\mathbb{E}(X\mathbb{1}_{\{X>y\}}) - y\mathbb{P}(X>y)\big] f(y)\,\mathrm{d}y = \frac{1}{\mu}(I-J). \end{aligned}$$

The proof is complete. □

Proof of Theorem 1. First, we will find a closed expression for the integral I of Lemma 1. By using formula of truncated moments given in Property (SM4), we have

$$I = \int_0^\infty \mathbb{E}(X\mathbb{1}_{\{X>y\}}) f_{\mathrm{QSM}}(y; a, \gamma, q)\,\mathrm{d}y, \quad X \sim \mathrm{QSM}(a, \gamma, q),$$
$$= \int_0^\infty \frac{aq\gamma(\gamma/y)^{aq-1}}{(aq-1)c_q^q}\, {}_2F_1\left(1+q, q-\frac{1}{a}; q-\frac{1}{a}+1; -\frac{(\gamma/y)^a}{c_q}\right) \frac{aqc_q(y/\gamma)^{a-1}}{\gamma[1+c_q(y/\gamma)^a]^{1+q}}\,\mathrm{d}y, \quad aq > 1.$$

By taking the change of variables $z = c_q(y/\gamma)^a$ and $\mathrm{d}z = ac_q(y/\gamma)^{a-1}\mathrm{d}y/\gamma$, the above integral is written as

$$= \frac{aq^2\gamma}{(aq-1)c_q^{1/a}} \int_0^\infty \frac{1}{z^{q-1/a}}\, {}_2F_1\left(1+q, q-\frac{1}{a}; q-\frac{1}{a}+1; -\frac{1}{z}\right) \frac{1}{(1+z)^{1+q}}\,\mathrm{d}z.$$

The following integral representation is well-known: ${}_2F_1(\alpha, \beta; \gamma; z) = \mathrm{B}^{-1}(\beta, \gamma-\beta) \int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}\mathrm{d}t$, $\gamma > \beta > 0$; see Eq. (9.111) in Gradshteyn and Ryzhik (2015). Employing this integral representation, the above integral is

$$= \frac{aq^2\gamma}{(aq-1)c_q^{1/a}} \int_0^\infty z^{1+1/a} \left[\frac{1}{\mathrm{B}(q-\frac{1}{a}, 1)} \int_0^1 t^{-2-1/a} \frac{1}{[1+(1/t)z]^{1+q}}\,\mathrm{d}t\right] \frac{1}{(1+z)^{1+q}}\,\mathrm{d}z.$$

By Fubini's theorem (Rosenthal, 2006) we can change the order of integration of last iterated integral (which is valid, because the integrands are positive) so that the expression above is

$$= \frac{q^2\gamma}{(q-\frac{1}{a})\mathrm{B}(q-\frac{1}{a}, 1)c_q^{1/a}} \int_0^1 t^{-2-1/a} \left[\int_0^\infty \frac{z^{1+1/a}}{[1+(1/t)z]^{1+q}(1+z)^{1+q}}\,\mathrm{d}z\right] \mathrm{d}t.$$

By using the identity: $\int_0^\infty x^{\lambda-1}(1+ax^p)^{-\mu}(1+bx^p)^{-\nu}\mathrm{d}x = p^{-1}a^{-\lambda/p}\mathrm{B}(\lambda/p, \mu+\nu-\lambda/p)\, {}_2F_1(\nu, \lambda/p; \mu+\nu; 1-b/a)$, $p > 0$ and $0 < \lambda < 2(\mu+\nu)$; see Eq. (3.259.3) in Gradshteyn and Ryzhik (2015); the last integral is

$$= \frac{q^2\gamma\mathrm{B}(2+\frac{1}{a}, 2q-\frac{1}{a})}{(q-\frac{1}{a})\mathrm{B}(q-\frac{1}{a}, 1)c_q^{1/a}} \int_0^1 {}_2F_1\left(1+q, 2+\frac{1}{a}; 2+2q; 1-t\right) \mathrm{d}t, \quad 2a(1+2q) > 1.$$

But, $(q-1/a)\mathrm{B}(q-1/a,1)=1$. Finally, by employing the known identity of derivative of hypergeometric functions: $\mathrm{d}\{{}_2F_1(\alpha,\beta;\gamma;z)\}/\mathrm{d}z=(\alpha\beta/\gamma)\,{}_2F_1(\alpha+1,\beta+1;\gamma+1;z)$; see Eq. (9.103.1) in Gradshteyn and Ryzhik, 2015; we have

$$\begin{aligned} I &= \frac{2q^2\gamma\mathrm{B}(2+\frac{1}{a},2q-\frac{1}{a})}{(2+\frac{1}{a})c_q^{1/a}}\int_1^0\frac{\mathrm{d}}{\mathrm{dt}}\,{}_2F_1\left(q,1+\frac{1}{a};1+2q;1-t\right)\mathrm{d}t \\ &= \frac{2q^2\gamma(1+\frac{1}{a})\mathrm{B}(1+\frac{1}{a},2q-\frac{1}{a})}{(2+\frac{1}{a})(1+2q)c_q^{1/a}}\left[{}_2F_1\left(q,1+\frac{1}{a};1+2q;1\right)-1\right],\quad 2a(1+2q)>1, \end{aligned}$$

where in the second equality we have used the simple recurrence formula $\mathrm{B}(x+1,y)=x\mathrm{B}(x,y)/(x+y)$, the Fundamental Theorem of Calculus and the identity ${}_2F_1(\alpha,\beta;\gamma;0)=1$. But, by Gauss's summation theorem; see Eq. (9.122.1) of Gradshteyn and Ryzhik, 2015: ${}_2F_1(a,b;c;1)=\Gamma(c)\Gamma(c-a-b)/[\Gamma(c-a)\Gamma(c-b)]$, $c>a+b$, the integral I is

$$I=\frac{2q^2\gamma(1+\frac{1}{a})\mathrm{B}(1+\frac{1}{a},2q-\frac{1}{a})}{(2+\frac{1}{a})(1+2q)c_q^{1/a}}\left[\frac{\Gamma(1+2q)\Gamma(q-\frac{1}{a})}{\Gamma(1+q)\Gamma(2q-\frac{1}{a})}-1\right],\quad 2a(1+2q)>1, aq>1. \tag{A.1}$$

Second, the integral J in Lemma 1 is simply determined. Indeed, since $X\sim\mathrm{QSM}(a,\gamma,q)$, by using Equation (1.2) with $b=\gamma c_q^{-1/a}$, we get

$$\begin{aligned} J=\int_0^\infty y\mathbb{P}(X>y)f_{\mathrm{QSM}}(y;a,\gamma,q)\,\mathrm{d}y &= \int_0^\infty y\,[1+c_q(y/\gamma)^a]^{-q}\,\frac{aqc_q(y/\gamma)^{a-1}}{\gamma[1+c_q(y/\gamma)^a]^{1+q}}\,\mathrm{d}y \\ &= \gamma\int_0^\infty (y/\gamma)\,\frac{aqc_q(y/\gamma)^{a-1}}{\gamma[1+c_q(y/\gamma)^a]^{1+2q}}\,\mathrm{d}y \\ &= \frac{q\gamma}{c_q^{1/a}}\int_0^\infty\frac{z^{1/a}}{(1+z)^{1+2q}}\,\mathrm{d}z, \end{aligned}$$

where in the last equality we used the change of variables $z=c_q(y/\gamma)^a$ and $\mathrm{d}z=ac_q(y/\gamma)^{a-1}\mathrm{d}y/\gamma$.

By using Eq. (3.251.11) in Gradshteyn and Ryzhik (2015), the above integral is

$$= \frac{q\gamma}{c_q^{1/a}} \, \mathrm{B}\left(1+\frac{1}{a}, 2q-\frac{1}{a}\right), \quad 2aq > 1.$$

Therefore,

$$J = \frac{q\gamma}{c_q^{1/a}} \, \mathrm{B}\left(1+\frac{1}{a}, 2q-\frac{1}{a}\right), \quad 2aq > 1. \tag{A.2}$$

Finally, by combining (A.1) and (A.2) with Lemma 1, the proof of Formula (1.4) follows.

□

A.2 Proofs of Subsection 1.2.4

Proof of Property (QDA4). If $Y \sim \mathrm{QDA}(a, \gamma, p)$ then

$$\mathbb{E}(Y^r \mathbb{1}_{\{Y>x\}}) = \int_x^\infty y^r \, \frac{ap(y/\gamma)^{ap-1}}{\gamma e_p^p [1 + e_p^{-1}(y/\gamma)^a]^{1+p}} \, \mathrm{d}y.$$

Taking the change of variables $z = e_p^{-1}(y/\gamma)^a$ and $\mathrm{d}z = a e_p^{-1}(y/\gamma)^{a-1}\mathrm{d}y/\gamma$ we get

$$= p\gamma^r e_p^{ar+p-1} \int_{e_p^{-1}(x/\gamma)^a}^\infty \frac{z^{ar+p-1}}{(1+z)^{1+p}} \, \mathrm{d}z.$$

By using the identity: $\int_u^\infty x^{a-1}(1+bx)^{-\nu}\mathrm{d}x = u^{a-\nu} b^{-\nu} (\nu - a)^{-1} \, {}_2F_1(\nu, \nu - a; \nu - a + 1; -1/(bu))$, $\nu > a$; see Eq. (3.194.2) in Gradshteyn and Ryzhik (2015); the last integral is

$$= \frac{p\gamma^r (\gamma/x)^{a(1-ar)}}{(1-ar) e_p^{-p}} \, {}_2F_1\left(1+p, 1-ar; 2-ar; -\frac{(\gamma/x)^a}{e_p^{-1}}\right), \quad ar < 1.$$

Then the proof follows. □

Proof of Theorem 2. By Lemma 1 the Gini coefficient is given by $G = (I - J)/\mu$.

By using formula of truncated moments given in Property (QDA4),

$$
\begin{aligned}
I &= \int_0^\infty \mathbb{E}(X\mathbb{1}_{\{X>y\}}) f_{\mathrm{QDA}}(y;a,\gamma,p)\,\mathrm{d}y, \quad X\sim \mathrm{QDA}(a,\gamma,p),\\
&= p\int_0^\infty \frac{p\gamma(\gamma/y)^{a(1-a)}}{(1-a)e_p^{-p}}\, {}_2F_1\biggl(1+p,1-a;2-a;-\frac{(\gamma/y)^a}{e_p^{-1}}\biggr)\, \frac{ap(y/\gamma)^{ap-1}}{\gamma e_p^{p}[1+e_p^{-1}(y/\gamma)^a]^{1+p}}\,\mathrm{d}y\\
&= \frac{p\gamma e_p^{a(a-1)}}{(1-a)e_p^{-p}}\int_0^\infty \frac{1}{z^{1+a(1-a)-p}}\, {}_2F_1\biggl(1+p,1-a;2-a;-\frac{1}{z}\biggr)\, \frac{1}{(1+z)^{1+p}}\,\mathrm{d}z, \quad a<1,
\end{aligned}
$$

where in the last equality the change of variables $z=e_p^{-1}(y/\gamma)^a$ and $\mathrm{d}z=ae_p^{-1}(y/\gamma)^{a-1}\mathrm{d}y/\gamma$ was considered. Employing the integral representation: ${}_2F_1(\alpha,\beta;\gamma;z)=\mathrm{B}^{-1}(\beta,\gamma-\beta)\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}\mathrm{d}t$, $\gamma>\beta>0$; see Eq. (9.111) in Gradshteyn and Ryzhik (2015); the above integral is

$$
\begin{aligned}
&= \frac{p\gamma e_p^{a(a-1)}}{(1-a)e_p^{-p}}\int_0^\infty \frac{1}{z^{1+a(1-a)-p}}\biggl[\frac{1}{\mathrm{B}(1-a,1)}\int_0^1 t^{-a}(1-tz)^{-(1+p)}\,\mathrm{d}t\biggr]\frac{1}{(1+z)^{1+p}}\,\mathrm{d}z\\
&= \frac{p\gamma e_p^{a(a-1)}}{(1-a)\mathrm{B}(1-a,1)e_p^{-p}}\int_0^1 t^{-a}\biggl[\int_0^\infty \frac{z^{p-a(1-a)-1}}{(1-tz)^{1+p}(1+z)^{1+p}}\,\mathrm{d}z\biggr]\mathrm{d}t,
\end{aligned}
$$

where we used Fubini's theorem (Rosenthal, 2006) to change the order of integration. By using the identity: $\int_0^\infty x^{\lambda-1}(1+ax^s)^{-\mu}(1+bx^s)^{-\nu}\mathrm{d}x = s^{-1}a^{-\lambda/s}\mathrm{B}(\lambda/s,\mu+\nu-\lambda/s)\,{}_2F_1(\nu,\lambda/s;\mu+\nu;1-b/a)$, $s>0$ and $0<\lambda<2(\mu+\nu)$; see Eq. (3.259.3) in Gradshteyn and Ryzhik (2015); the integral above is written as

$$
\begin{aligned}
&= \frac{p\gamma\mathrm{B}(p-a(1-a),2+p+a(1-a))}{(1-a)\mathrm{B}(1-a,1)e_p^{a(1-a)-p}}\int_0^1 t^{-a}\,{}_2F_1\bigl(1+p,p-a(1-a);2(1+p);1+t\bigr)\,\mathrm{d}t\\
&= \frac{p\gamma\mathrm{B}(p-a(1-a),2+p+a(1-a))}{e_p^{a(1-a)-p}}\, I(a,p),
\end{aligned}
$$

because $(1-a)\mathrm{B}(1-a,1)=1$. Therefore,

$$I = \frac{p\gamma\mathrm{B}(p-a(1-a), 2+p+a(1-a))}{e_p^{a(1-a)-p}}\, I(a,p), \quad a<1. \tag{A.3}$$

On the other hand, the integral J in Lemma 1 is determinated of following form. By using (1.5) with $b=\gamma e_p^{1/a}$,

$$\begin{aligned}
J &= \int_0^\infty y\mathbb{P}(X>y) f_{\mathrm{QDA}}(y;a,\gamma,p)\,\mathrm{d}y, \quad X\sim \mathrm{QDA}(a,\gamma,p),\\
&= \mathbb{E}(Y) - \int_0^\infty y\big[1+\{e_p^{-1}(y/\gamma)^a\}^{-1}\big]^{-p}\,\frac{ap(y/\gamma)^{ap-1}}{\gamma e_p^p[1+e_p^{-1}(y/\gamma)^a]^{1+p}}\,\mathrm{d}y\\
&= \mathbb{E}(Y) - p\gamma e_p^{1/a}\int_0^\infty \frac{z^{2p+(1/a)-1}}{(1+z)^{1+2p}}\,\mathrm{d}z,
\end{aligned}$$

where we used the change of variables $z=e_p^{-1}(y/\gamma)^a$ and $\mathrm{d}z = ae_p^{-1}(y/\gamma)^{a-1}\mathrm{d}y/\gamma$. Eq. (3.251.11) in Gradshteyn and Ryzhik (2015) allows writing the last integral as

$$= \mathbb{E}(Y) - p\gamma e_p^{1/a}\mathrm{B}\left(2p+\frac{1}{a}, 1-\frac{1}{a}\right), \quad a<1.$$

Therefore, by Property (DA2),

$$J = \gamma e_p^{1/a}\mathrm{B}\left(a+\frac{1}{p}, 1-\frac{1}{p}\right) - p\gamma e_p^{1/a}\mathrm{B}\left(2p+\frac{1}{a}, 1-\frac{1}{a}\right), \quad a<1, p>1. \tag{A.4}$$

Hence, by combining (A.3) and (A.4) with Lemma 1, the proof of Formula (1.6) follows. □

References

Allison, Paul D (1978). “Measures of inequality”. *American sociological review*, pp. 865–880.

Berrebi, Z Moshe and Silber, Jacques (1985). “Income inequality indices and deprivation: A generalization”. *The Quarterly Journal of Economics* 100.3, pp. 807–810.

Chen, Chau-Nan, Tsaur, Tien-Wang, and Rhai, Tong-Shieng (1982). “The Gini coefficient and negative income”. *Oxford Economic Papers* 34.3, pp. 473–478.

Cramer, Jan Salomon (1971). *Empirical econometrics*. North-Holland Amsterdam.

Dagum, Camilo (1973). *Un modèle nonlinéaire de répartition fonctionnelle du revenu*. Faculté des sciences sociales. Dép. de science économique, Université d’Ottawa.

— (1975). “A model of income distribution and the conditions of existence of moments of finite order”. *Bulletin of the International Statistical Institute* 46.1, pp. 199–205.

— (2008). “Modeling income distributions and Lorenz curves”. In: ed. by Duangkamon Chotikapanich. Vol. 5. Springer Science & Business Media. Chap. A New Model of Personal Income Distribution: Specificationand Estimation.

Dey, Sanku, Al-Zahrani, Bander, and Basloom, Samerah (2017). “Dagum Distribution: Properties and Different Methods of Estimation”. *International Journal of Statistics and Probability* 6, pp. 74–92.

Fellman, Johan et al. (2012). “Estimation of Gini coefficients using Lorenz curves”. *Journal of Statistical and Econometric Methods* 1.2, pp. 31–38.

Galarza, Christian E, Zhang, Panpan, and Lachos, Víctor H (2020). "Logistic quantile regression for bounded outcomes using a family of heavy-tailed distributions". *Sankhya B*, pp. 1–25.

Gradshteyn, I.S. and Ryzhik, I.M. (2015). *Table of Integrals, Series and Products Eighth Edition*. Academic Press, San Diego: Elsevier.

Hajargasht, Gholamreza et al. (2012). "Inference for income distributions using grouped data". *Journal of Business & Economic Statistics* 30.4, pp. 563–575.

Kleiber, Christian (1996). "Dagum vs. Singh-Maddala income distributions". *Economics Letters* 53.3, pp. 265–268.

— (2008). "A guide to the Dagum distributions". In: *Modeling income distributions and Lorenz curves*. Springer, pp. 97–117.

Klugman, Stuart A., Panjer, Harry H., and Willmot, Gordon E (2019). *Loss models : from data to decisions*. Vol. Fifth edition. Wiley Series in Probability and Statistics: Wiley.

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

Krämer, Walter and Ziebach, Thorsten (2002). *The weak Pareto law and regular variation in the tails*. Tech. rep. Universität Dortmund.

Kumar, Devendra (2017a). "The Singh–Maddala distribution: properties and estimation". *International journal of system assurance engineering and management* 8.2, pp. 1297–1311.

— (2017b). "The Singh–Maddala distribution: properties and estimation". *International Journal of System Assurance Engineering and Management* 8, pp. 1297–1311.

Mittelhammer, Ron C, Judge, George G, and Miller, Douglas J (2000). *Econometric foundations pack with CD-ROM*. Cambridge University Press.

Pareto, Vilfredo (1897). *Cours d'économie politique*. Vol. 1. F. Pichou.

Pundir, Sudesh, Arora, Sangeeta, and Jain, Kanchan (2005). "Bonferroni curve and the related statistical inference". *Statistics & probability letters* 75.2, pp. 140–150.

Reed, William J (2003). "The Pareto law of incomes—an explanation and an extension". *Physica A: Statistical Mechanics and its Applications* 319, pp. 469–486.

Rosenthal, J.S. (2006). *A First Look at Rigorous Probability Theory*. World Scientific. ISBN: 9789812703705. URL: `https://books.google.com.br/books?id=PYNAgzDffDMC`.

Rousseeuw, P. et al. (2016). *Robustbase: basic robust statistics. R package version 0.92-6.* URL: `http://CRAN.Rproject.org/package=robustbase`.

Sánchez, Luis et al. (2021a). "A new quantile regression model and its diagnostic analytics for a Weibull distributed response with applications". *Journal of Computational and Applied Mathematics* 9, p. 2768.

Sánchez, Luis et al. (2021b). "Birnbaum-Saunders quantile regression and its diagnostics with application to economic data". *Applied Stochastic Models in Business and Industry* 37.1, pp. 53–73.

Saulo, H. et al. (2021). "Log-symmetric quantile regression models". *Statistica Neerlandica* 76, pp. 124–163.

Shirras, G Findlay (1935). "The Pareto Law and the distribution of income". *The Economic Journal* 45.180, pp. 663–681.

Singh, S.K. and Maddala, G. S. (1976). "A model of income distribution and the conditions of existence of moments of finite order". *Econometrica* 44, pp. 963–970.

Stich, Andreas (1996). "Inequality and negative income". *Journal of the Italian Statistical Society* 5.3, pp. 297–305.