

Canavire Bacarreza, Gustavo J.; Rios-Avila, Fernando; Sacco-Capurro, Flavia

Working Paper

Recovering Income Distribution in the Presence of Interval-Censored Data

IZA Discussion Papers, No. 15921

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Canavire Bacarreza, Gustavo J.; Rios-Avila, Fernando; Sacco-Capurro, Flavia (2023) : Recovering Income Distribution in the Presence of Interval-Censored Data, IZA Discussion Papers, No. 15921, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/272548>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

DISCUSSION PAPER SERIES

IZA DP No. 15921

**Recovering Income Distribution in the
Presence of Interval-Censored Data**

Gustavo Canavire-Bacarreza
Fernando Rios-Avila
Flavia Sacco-Capurro

FEBRUARY 2023

DISCUSSION PAPER SERIES

IZA DP No. 15921

Recovering Income Distribution in the Presence of Interval-Censored Data

Gustavo Canavire-Bacarreza

The World Bank and IZA

Fernando Rios-Avila

Levy Institute at Bard College

Flavia Sacco-Capurro

The World Bank

FEBRUARY 2023

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Recovering Income Distribution in the Presence of Interval-Censored Data*

This paper proposes a method to analyze interval-censored data, using multiple imputation based on a heteroskedastic interval regression approach. The proposed model aims to obtain a synthetic data set that can be used for standard analysis, including standard linear regression, quantile regression, or poverty and inequality estimation. The paper presents two applications to show the performance of the method. First, it runs a Monte Carlo simulation to show the method's performance under the assumption of multiplicative heteroskedasticity, with and without conditional normality. Second, it uses the proposed methodology to analyze labor income data in Grenada for 2013–20, where the salary data are interval-censored according to the salary intervals prespecified in the survey questionnaire. The results obtained are consistent across both exercises.

JEL Classification: C150, C340, J3

Keywords: interval-censored data, Monte Carlo simulation, heteroskedastic interval regression, wages

Corresponding author:

Gustavo Canavire-Bacarreza

The World Bank

1818 H Street

NW Washington

DC 20433

USA

E-mail: gcanavire@worldbank.org

* The findings, interpretations, and conclusions expressed in this paper do not necessarily reflect the views of the World Bank, the Executive Directors of the World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The authors would like to thank Ximena del Carpio, Leonardo Luchetti, Carlos Ospino, Daniel Mahler and the participants at the 5th IZA Labor Statistics Workshop: The Measurement of Incomes, Living Costs and Standards of Living and the 2022 Stata Conference for helpful comments and suggestions.

1. Introduction

Labor force surveys are a useful data source to understand employment dynamics in both developing and developed countries. These surveys provide vast information on the labor market status at higher frequency levels than living conditions surveys. In some cases, they are the only source of information to describe and examine the structure of the labor markets. In the Latin American and the Caribbean region, countries like Bolivia, Costa Rica, Ecuador, Jamaica, Mexico, Peru, and Uruguay collect their labor force surveys quarterly as opposed to a yearly basis, which is the case of most household and living standard surveys.

One of the key features of these labor surveys is that they provide information on the wages and salaries of workers. This allows to estimate job market trends and obtain inequality measures of labor income among workers. However, the full income distribution in many countries cannot be retrieved because labor income is reported in brackets. Because of this, the estimation of inequality or poverty measures, as well as regression type analysis, is difficult. This is the case of the labor force survey for all countries in the Organization of Eastern Caribbean States (OECS).

This is not unique to the Caribbean region. Countries like Colombia, Germany, Australia, New Zealand, Bosnia and Herzegovina, North Macedonia and Serbia, among others, have similar data collection protocols for their micro census (Walter and Weimer 2018). In the U.S, the current population survey (CPS) collects detailed family income only once a year, in the March supplement, but collects family income in brackets on a monthly basis.

One argument in favor of using interval-censored questions to collect information on income is the higher response rate compared to questions asking to report exact amounts (Wang et al., 2013). This happens because income information is considered “sensitive”, and people are reluctant to report actual earnings, and may choose not to respond to those questions at all (Moore et al., 2000; Hagenaars and Vos, 1988). Field tests conducted in the past have shown that asking follow-up income questions in a series of unfolding brackets achieves superior results in terms of response rates for income amounts, as was the case of the National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System Survey (BRFSS), both administered by the Center for Disease Control and Prevention of the United States (Angelov and Ekstrom 2018, Yan et al. 2018).

However, even though this form of data collection solves the problem of underreporting or misreporting, it raises a problem for recovering the full wage (income) distribution, which is key to understanding and analyzing inequality.

To better use the information from these types of surveys, we propose an imputation approach to simulate the distribution of the data that is only available in brackets. The method is an extension on the imputation approach described in Royston (2007), which considers heteroskedastic errors to model the conditional distribution of the censored data. The estimated conditional distribution is then used to impute the data using draws from the estimated conditional distribution. Once the imputed data is obtained, standard aggregation methods (Rubin, 1987) can be used to analyze the censored data as if it were fully observed. For example, it can be used to calculate poverty or inequality measures, as well as perform regression analysis. To demonstrate the flexibility of this approach, we use a Monte Carlo simulation to analyze the sensitivity of our method. As an empirical example, we use the approach to analyze wage inequality in Grenada utilizing the country's Labor Force Survey.

Other approaches exist in the literature, and have been used for analyzing this kind of data. To measure income inequality with right-censored (top-coded) data, Jenkins et al. (2011) propose multiple-imputation methods for estimation and inference where censored observations are imputed using draws from a flexible parametric model fitted to the censored distribution, such as Generalized Beta of the second kind (GB2), Singh-Maddala or Dagum distributions. Chen (2017) provides a generalized approach for the estimation of parametric income distributions using grouped data, showing its consistency through complementary simulation results. More recently, Walter and Weimer (2018) propose an iterative kernel density algorithm that generates pseudo samples from the interval-censored income variable to estimate poverty and inequality indicators. While the interval regression approach we propose fits with the models described in Chen (2017), Jenkins et al (2011), and Walter and Weimer (2018), these papers focus on recovering the unconditional distribution of income, without considering the relationship with explanatory variables.

Zhou et al. (2017) and Chih-Yuan et al. (2021) propose methodologies for the estimation of conditional quantile regressions using interval censored data, under different distributional

assumptions. While this approach can be used for analyzing interval censored data, it only focuses on estimating conditional quantile regressions, requiring specialized software that are not readily available. In contrast, the method we propose can be applied not only for the estimation of conditional quantile regressions, but also for the estimation of unconditional distribution statistics.

Other studies, like the one proposed by Han et al. (2020), construct new measures of the income distribution and estimate poverty in the U.S. using data from the monthly Current Population Survey (CPS). They address the problem of censored income data using draws from the empirical income distribution observed in the last March supplement. A similar method is proposed by Parolin and Wimer (2020), who produce monthly updates of the Supplemental Poverty Measure (SPM) rates with demographic data from the CPS and poverty data from the previous March supplement of the CPS. However, these studies seek to obtain income estimates using the uncensored distribution of previous years, which is not always available with other data sources, like the ones analyzed in this paper.

Buutner and Rassler (2008) propose a multiple imputation approach, similar to ours, to analyze wages from the German Institute of Employment Research (IAB) employment survey. While their method focuses on the analysis top coded data, we expand the approach to analyze data with a more generalized censoring structure.

The paper is organized as follows. Section 2 introduces the model and the econometric issues associated with the imputation method; Section 3 provides a Monte Carlo simulation exercise to analyze the performance of the methodology; Section 4 uses the methodology to analyze labor income distribution changes in Grenada using the 2013-2020 series of the Labor Force Survey. Section 5 concludes.

2. Methodology

To address the problem of interval-censored data, we propose a multiple imputation approach based on a heteroskedastic interval-regression model. An interval-regression model is a generalization of the Tobit model that allows using a mixture of censored and completely observed data, even if the censoring thresholds are unique to each individual. The goal of the model is to

find a set of parameters that maximizes the probability that, given a set of characteristics, the predicted latent earnings fall within the declared earning threshold. Imputations are obtained using random draws of the estimated conditional distributions.

2.1. Interval regression model

Assume that (log) earned income (y_i) has a data generating process such that:

$$y_i = \mu(x_i) + v_i\sigma(x_i) \quad (1)$$

Where v_i is a homoscedastic i.i.d. error, with mean 0 and standard deviation 1, that is independent of the characteristics x . $\mu(x_i)$ and $\sigma(x_i)$ are flexible functions of x_i . $\mu(x_i)$ represents the conditional mean of y_i , and $\sigma(x_i)$ is a strictly positive function that represents the conditional standard deviation of y_i . Following Machado and Santos-Silva (2019), the conditional mean $\mu(x_i)$ captures location shift effects of characteristics on the outcome, whereas $\sigma(x_i)$ capture the scale shifts, which relate to how much of the spread is explained by differences in characteristics. Following the standard setup of interval-regression models (Stewart, 1983), we impose the assumption that v_i follows a standard normal distribution, so that $y_i|x_i$ is also normally distributed with mean $\mu(x_i)$ and standard deviation $\sigma(x_i)$.¹

$$\text{if } v_i \sim N(0,1) \rightarrow y_i|x_i \sim N(\mu(x), \sigma(x)) \quad (2)$$

Under this assumption, equation 1 can be estimated via maximum likelihood by maximizing the following function:

$$L_i(\mu(x), \sigma(x)) = f_{y|x}(\mu(x), \sigma(x)) = \frac{1}{\sigma(x)} \phi\left(\frac{y_i - \mu(x)}{\sigma(x)}\right) \quad (3a)$$

¹ While this assumption is unnecessary for the estimation of standard linear regression models, imposing some distribution assumption on the errors is necessary when estimating models via maximum likelihood. Nevertheless, as described in MacDonald, Stoddard and Walton (2018), it is possible to relax this assumption using more flexible distributions.

$$\mu(x), \sigma(x) = \max \frac{1}{N} \sum \log(L_i) \quad (3b)$$

Under these conditions, and assuming a flexible enough model specification to capture the conditional mean and conditional variance, estimating equation (1) allows us to recover the whole distribution of the dependent variable y_i .

When y_i is fully observed, this variable can be directly used for estimating any measure of poverty or inequality, or to analyze the relationship between observed characteristics X and the outcome y , using standard statistical methods. Often, however, due to survey design, one may only have access to data reported in brackets. In other words, rather than observing y_i , one may only observe that reported income by individual i is within some lower (ll_i) and upper (uu_i) thresholds, which may be different for each individual. In this case, unless $ll_i = uu_i$, the likelihood function defined by Equations 3a and 3b is not defined.

An alternative for estimating a model with this type of data is the use of what is known as interval regression. Interval regression is a generalization of the censored regression estimators like the Tobit model (see Cameron and Trivedi (2010, ch 16) for a discussion of censored regressions), where data can be a mixture of left-censored, right-censored, interval-censored, or fully observed. For simplicity, we refer to the case with interval-censored data.

When the data is interval-censored, rather than modeling the outcome itself, the approach focuses on modeling the probability that an individual i reports income to be within the underlying income brackets:

$$P(ll_i \leq y_i < uu_i | x_i) \quad (4)$$

Using the data generating process (d.g.p.) defined by equation 1, and the normality assumption of the error v_i , equation (4) can be rewritten as:

$$P\left(\frac{ll_i - \mu(x_i)}{\sigma(x_i)} \leq v_i < \frac{uu_i - \mu(x_i)}{\sigma(x_i)} \mid x_i\right) = P\left(v_i < \frac{uu_i - \mu(x_i)}{\sigma(x_i)}\right) - P\left(v_i < \frac{ll_i - \mu(x_i)}{\sigma(x_i)}\right) \quad (5a)$$

$$= \Phi\left(\frac{uu_i - \mu(x_i)}{\sigma(x_i)}\right) - \Phi\left(\frac{ll_i - \mu(x_i)}{\sigma(x_i)}\right) \quad (5b)$$

Where $\Phi(\cdot)$ is the cumulative normal density function. Using equation (5b), the loglikelihood function that is maximized to identify the parameters $\mu(x_i)$ and $\sigma(x_i)$ is defined as:

$$L_i(\mu(x), \sigma(x)) = \Phi\left(\frac{uu_i - \mu(x_i)}{\sigma(x_i)}\right) - \Phi\left(\frac{ll_i - \mu(x_i)}{\sigma(x_i)}\right) \text{ if data is interval - censored} \quad (6a)$$

$$L_i(\mu(x), \sigma(x)) = \Phi\left(\frac{uu_i - \mu(x_i)}{\sigma(x_i)}\right) \text{ if data is left - censored} \quad (6b)$$

$$L_i(\mu(x), \sigma(x)) = 1 - \Phi\left(\frac{ll_i - \mu(x_i)}{\sigma(x_i)}\right) \text{ if data is right - censored} \quad (6c)$$

$$L_i(\mu(x), \sigma(x)) = \frac{1}{\sigma(x_i)} \Phi\left(\frac{ll_i - \mu(x_i)}{\sigma(x_i)}\right) \text{ if data is fully observed} \quad (6d)$$

Which can be used to obtain estimates for $\mu(x)$ and $\sigma(x)$ using maximum likelihood estimation.

2.2. Model imputation

As previously described, when dealing with interval-censored data, we have limited access to the observed distribution of the variable of interest. This is in contrast with standard multiple imputation analysis, where the variable of interest is fully unobserved. This distinction has implications on the imputation strategy because it determines the appropriate draw of the imputed error.

Consider the d.g.p stated in equation 1 and define y_i^* to be the true but unobserved variable of interest. By definition, if the data is interval-censored, the range of values that can be potentially used to impute y_i^* are bounded between the lower and upper threshold of a given interval. In addition, conditional on the observed characteristics x , and the parameters $\mu(x_i)$ and $\sigma(x_i)$, it implies that the unobserved error v_i^* is also bounded:

$$v_i^* \in \left[\frac{ll_i - \mu(x_i)}{\sigma(x_i)}, \frac{uu_i - \mu(x_i)}{\sigma(x_i)} \right] \quad (7)$$

Furthermore, under the assumption that v_i follows a standard normal distribution, we can impute values for y_i^* , by simply getting random draws for v_i^* from a truncated random normal distribution:

$$\tilde{v}_i = \Phi^{-1}(r_i), \text{ where } r_i \sim \text{Uniform} \left(\Phi \left(\frac{ll_i - \mu(x_i)}{\sigma(x_i)} \right), \Phi \left(\frac{uu_i - \mu(x_i)}{\sigma(x_i)} \right) \right) \quad (8)$$

Where $\Phi^{-1}(r_i)$ corresponds to the r_{th} quantile for the standard normal distribution. Finally, the imputed value for the outcome of interest y_i^* is given by:

$$\tilde{y}_i = \mu(x_i) + \tilde{v}_i \sigma(x_i) \quad (9)$$

Because the population parameters $\mu(x_i)$ and $\sigma(x_i)$ are unknown, we use the sample equivalents that are estimated using the interval regression estimator via maximum likelihood.² To account for the uncertainty of the regression estimation, we obtain random draws from the following joint normal distribution:

$$\begin{bmatrix} \tilde{\mu}(x) \\ \tilde{\sigma}(x) \end{bmatrix} \sim N \left(\begin{bmatrix} \hat{\mu}(x) \\ \hat{\sigma}(x) \end{bmatrix}, \tilde{\Omega} \right); \quad \tilde{\Omega} = \hat{\Omega} * \frac{n}{\tilde{n}}; \quad \tilde{n} \sim \chi_n^2 \quad (10)$$

Where $\hat{\Omega}$ is the ML variance-covariance matrix estimate, n is the number of observations in the sample, and \tilde{n} is a random draw from a chi squared distribution n degrees of freedom. Finally, the imputation for y_i^* will be given by:

$$\tilde{y}_i = \tilde{\mu}(x_i) + \tilde{v}_i \tilde{\sigma}(x_i) \quad (11a)$$

$$\tilde{v}_i = \Phi^{-1}(\tilde{r}_i), \text{ where } \tilde{r}_i \sim \text{Uniform} \left(\Phi \left(\frac{ll_i - \tilde{\mu}(x_i)}{\tilde{\sigma}(x_i)} \right), \Phi \left(\frac{uu_i - \tilde{\mu}(x_i)}{\tilde{\sigma}(x_i)} \right) \right) \quad (11b)$$

² For numerical purposes, it is also important to emphasize that $\sigma(x_i)$ is not estimated directly, but $\ln \sigma(x_i)$ is estimated instead.

Where \tilde{v}_i is used in (11a) instead of \tilde{v}_i , to account for the role of the estimated parameters on the error \tilde{v} .

In summary, the imputation algorithm is as follows:

1. Estimate the parameters associated with $\mu(x)$ and $\sigma(x)$ using a heteroskedastic interval regression approach via maximum likelihood, as well as the variance covariance matrix Ω .
2. Obtain \tilde{n} from a random draw from χ_n^2 , and estimate $\tilde{\Omega}$.
3. Obtain a random draw for $\tilde{\mu}(x)$ and $\tilde{\sigma}(x)$ from $N\left(\begin{matrix} \hat{\mu}(x) \\ \hat{\sigma}(x) \end{matrix}, \tilde{\Omega}\right)$.
4. Obtain random draws for \tilde{v}_i , conditional on $\tilde{\mu}(x)$ and $\tilde{\sigma}(x)$, for each observation i .
5. Get the full sample of imputed data \tilde{y}_i .
6. Repeat steps 2-4 M times and obtain M sets of imputed samples.

Steps 2-4 corresponds to simulating from the posterior distribution, similar to what is described in Gelman et al. (2014).

2.3. Model estimation and inference

Once the M imputed data sets have been obtained, statistical analysis can be done by independently implementing the desired model estimation across all M imputed samples. The aggregation and summary from the M estimated models could then be done applying the combination rules described in Rubin (1987).

Let β be the set of parameters of interest, and $\hat{\beta}_m$ and \hat{V}_m be the set of estimated coefficients and corresponding variance-covariance matrix obtained using simulated sample m . The multiple imputation estimates $\hat{\beta}_M$ for the parameter of interest is given by:

$$\hat{\beta}_M = \frac{1}{M} \sum_{m=1}^M \hat{\beta}_m \quad (13)$$

Whereas the variance-covariance estimate \hat{V}_M is given by:

$$\hat{V}_M = \frac{1}{M} \sum_{m=1}^M V_m + \left(\frac{M+1}{M} \right) \frac{(\hat{\beta}_m - \hat{\beta}_M)'(\hat{\beta}_m - \hat{\beta}_M)}{M-1} \quad (14)$$

3. Simulation studies

3.1. Setup

We examine the performance of our proposed estimator under several simulation scenarios, using data structures with explicit multiplicative heteroskedasticity, similar to the ones proposed in Machado and Santos-Silva (2019), and with a varying coefficient model structure, as in Hsu, Wen and Chen (2021). In both cases, the goal is to simulate data that would show heterogeneity when using conditional quantile regressions for the estimation. This structure is flexible enough to also allow the estimation of other distribution based regressions such as unconditional quantile regressions (Firpo, Fortin and Lemieux, 2009) and Recentered Influence function regressions in general (Rios-Avila, 2020).

The first set of simulations is designed to study the performance of the estimator under the assumption of multiplicative heteroskedasticity assuming the following functional form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v\sigma(x_1, x_2) \quad (15)$$

Where $x_1 \sim \text{bernulli}(0.5)$ and $x_2 \sim \chi_5^2/5$. Following Machado and Santos-Silva (2019), we use two different functional forms for $\sigma(x_1, x_2)$:

$$\sigma_1(x_1, x_2) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 \quad (16a)$$

$$\sigma_2(x_1, x_2) = e^{\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2} \quad (16b)$$

In both cases, we require that $\sigma(x_1, x_2)$ to be strictly positive. The first case, equation (16a), imposes the assumption of linear heteroskedasticity and provides a closed form solution for the corresponding quantile coefficients. The second option, equation (16b), guarantees standard deviation to be strictly positive, but does not have a closed form solution for the corresponding conditional quantile regression coefficients. As described in Machado and Santos-Silva (2019), this data generating process also guarantees that quantiles will not cross, and thus the

corresponding coefficients can be estimated directly using standard conditional quantile regression estimators.

Using this data structure, we consider four different distributions for the error v : Normal distribution, logistic distribution, chi square distribution with 5 degrees of freedom, and uniform distribution. All of them were adjusted to have a mean 0 and standard deviation 1. Whereas the first two distributions are meant to show how sensitive is the estimator to the normality assumption, the third and fourth aim to show how sensitive the results are to cases where the error has a skew distribution, or a distribution with limited range. With this considerations, the data generating process is defined as:

$$y = x_1 + x_2 + v * (1 - 0.5x_1 + 0.2x_2) \quad (17a)$$

$$y = x_1 + x_2 + v * e^{0.6-0.5+0.2x_2} \quad (17b)$$

The second set of simulations use a data generating process following a varying coefficient approach, based on the percentile τ an observation belongs to. In this setup, we assume that τ is defined by a random draw from a uniform distribution, and that y is given by:

$$y = \beta_0(\tau) + \beta_1(\tau)x_1 + \beta_2(\tau)x_2 \quad (18)$$

Following Hsu, et al (2021), the coefficients $\beta(\tau)$'s are defined as:

$$\beta_0(\tau) = 1 + 0.5\Phi^{-1}(\tau); \beta_1(\tau) = 0.4 + 1.2\Phi^{-1}(\tau); \beta_2(\tau) = 0.6 + 0.5\Phi^{-1}(\tau) \quad (19a)$$

$$\beta_0(\tau) = \beta_1(\tau) = \beta_2(\tau) = 0.5(1 + \Phi^{-1}(\tau) - \log(1 - \tau)) \quad (19b)$$

Equation (19a) imposes a structure that is similar to the multiplicative normality under linear heteroskedasticity (equation 17a), whereas the second equation imposes a skew conditional distribution of the outcome.

In all scenarios, we assume that data is subject to interval censoring, such that $ll_i = \lfloor y_i \rfloor$ & $uu_i = \lceil y_i \rceil$, where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ represent the nearest integers that is lower or higher than y_i respectively. In addition, we also assume if $y_i < -1$ or $y_i > 10$, the lower and upper thresholds, respectively, will be undefined.

For the implementation and analysis, we use 2500 replications, with a sample size of 1,000 observations. Replications using sample sizes of 500 and 2,000 are provided in the appendix, with results that are qualitatively similar. We focus on the comparison of conditional quantile regressions for the 10th, 50th and 90th quantiles, as well as for the 10th, 50th and 90th unconditional quantiles. Quantile regressions were estimated using the fast algorithm developed in Chernozhukov et al (2022) and implemented via the Stata command `-qrprocess-`, whereas the unconditional quantile regressions were estimated following Firpo, Fortin and Lemieux (2009) and implanted via the Stata command `-rifhdreg-` (Rios-Avila, 2020). Finally, the simulation was implemented using `-parallel-` (Vega Yon and Quistorff, 2019). Finally, our imputation method is implanted with a new user-written program `-intreg_mi-`, which is available upon request.

While population parameters for conditional quantile regressions for some of the data generating exists, there are no close form solutions for the population parameters corresponding to the RIF regressions. Because of this, we assume that average estimates using fully observed data represent the population parameters, for the calculation of the relevant statistics. Thus, by construction, the bias of the model estimations using fully observed data is zero.

3.2. Results

Tables 1 to 3 provide a summary of the results for the Monte Carlo simulations using the different data generating processes. In each table, we present the bias of our imputation procedure compared to the average parameters using fully observed data as if they were the asymptotic population parameters. For both conditional and unconditional quantile regressions, the bias observed using the multiple imputation data is small when the homoscedastic error is assumed to follow a symmetric bell curve distribution, regardless of the type of heteroskedasticity implied by the dgp.

When the errors follow a chi2 distribution or uniform distribution, we observe some bias, especially for the lower quantile coefficients. The bias, however, is considerably smaller if the data generating process assumes an functional form with exponential heteroskedasticity. Finally, the results using the varying coefficient structure reveal low bias in both cases. In all simulations, the bias magnitude did not depend on the sample size (see appendix).

In terms of the mean absolute error (MAE), we present the ratio between the MAE for the imputed data and the MAE for the fully observed data. Except for cases when the bias is large, the MAE for the imputed data is somewhat smaller than the one using fully observed data, by almost 10%. It is possible that this gain in the precision of the point estimate may be simulation specific, since we are indirectly using parametric structures for the estimation of the quantile regressions. In terms of standard errors ratio, which compares the average standard errors of the imputed data to fully observed data, we observe that the standard errors for imputed data are about 15% larger on average, than the standard errors based on fully observed data. This is expected given the information loss due to the nature of the interval censored data.

Table 1. Monte Carlo Simulation: N=1000, Linear Heteroskedasticity

$y = x\beta + u * \gamma x$		$u \sim \text{normal}$				$u \sim \text{logistic}$				$u \sim \text{Chi2}$				$u \sim \text{uniform}$			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	2.011	0.008	-0.014	0.235	1.955	-0.030	-0.094	0.156	1.848	0.321	3.886	0.915	2.094	0.258	2.073	0.690
	x2	0.798	0.002	-0.030	0.155	0.803	-0.006	-0.063	0.134	0.831	0.123	1.490	0.377	0.786	0.052	0.293	0.296
	cons	-2.381	-0.008	0.006	0.292	-2.247	0.022	-0.084	0.207	-1.989	-0.471	5.425	1.035	-2.572	-0.264	1.946	0.749
CQR-Q50	x1	1.000	0.001	-0.055	0.070	1.003	-0.001	-0.056	0.086	1.168	0.020	-0.029	0.094	0.996	0.004	-0.053	0.043
	x2	1.001	-0.002	-0.055	0.071	0.997	-0.001	-0.025	0.088	0.969	0.000	-0.048	0.084	0.999	0.001	-0.074	0.041
	cons	-0.001	0.002	-0.042	0.072	-0.002	0.002	-0.060	0.088	-0.385	0.023	-0.032	0.096	0.005	-0.005	-0.042	0.045
CQR-Q90	x1	-0.009	0.000	-0.049	0.103	0.041	0.005	-0.066	0.096	-0.051	-0.011	-0.051	0.048	-0.097	-0.008	-0.019	0.146
	x2	1.199	0.001	-0.041	0.110	1.191	0.000	-0.067	0.109	1.215	-0.004	-0.054	0.047	1.216	-0.001	0.024	0.145
	cons	2.383	-0.001	-0.059	0.099	2.252	0.007	-0.061	0.095	2.486	-0.003	-0.048	0.046	2.573	-0.042	0.057	0.141
UQR-Q10	x1	2.097	0.008	-0.005	0.159	1.915	-0.021	-0.065	0.114	1.840	0.188	0.903	0.287	2.539	0.236	0.760	0.459
	x2	0.611	0.001	-0.021	0.060	0.602	-0.008	-0.067	0.047	0.684	0.079	0.312	0.252	0.651	0.027	0.075	0.191
	cons	-2.537	-0.006	0.006	0.130	-2.342	0.016	-0.063	0.108	-2.370	-0.174	0.557	0.332	-3.046	-0.163	0.352	0.401
UQR-Q50	x1	1.006	0.000	-0.083	0.130	1.026	-0.007	-0.059	0.159	1.165	0.039	-0.011	0.161	0.945	0.012	-0.070	0.066
	x2	0.929	0.000	-0.058	0.110	0.919	-0.002	-0.061	0.128	0.945	0.000	-0.067	0.144	0.921	0.007	-0.096	0.062
	cons	0.131	0.001	-0.069	0.122	0.120	0.004	-0.072	0.140	-0.199	0.020	0.001	0.137	0.190	-0.004	-0.076	0.073
UQR-Q90	x1	0.052	-0.001	-0.076	0.103	0.106	0.004	-0.061	0.108	0.014	-0.004	-0.055	0.034	-0.003	-0.004	-0.045	0.166
	x2	1.466	0.001	-0.053	0.192	1.492	-0.004	-0.089	0.213	1.455	-0.004	-0.087	0.111	1.484	-0.015	0.052	0.231
	cons	2.263	-0.001	-0.075	0.129	2.134	0.010	-0.063	0.134	2.369	0.000	-0.060	0.068	2.314	0.010	0.003	0.176

Table 2 Monte Carlo Simulation: N=1000, exponential Heteroskedasticity

$y = x\beta + u * e^{\gamma x}$		$u \sim \text{normal}$				$u \sim \text{logistic}$				$u \sim \text{Chi2}$				$u \sim \text{uniform}$			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	1.639	-0.001	-0.032	0.168	1.603	-0.009	-0.066	0.143	1.536	-0.030	0.301	0.484	1.691	-0.012	0.228	0.320
	x2	0.743	0.004	-0.040	0.173	0.758	0.003	-0.078	0.162	0.788	0.018	0.258	0.465	0.726	0.046	0.501	0.268
	cons	-1.280	-0.003	-0.026	0.160	-1.209	-0.013	-0.081	0.129	-1.072	-0.050	0.434	0.456	-1.384	-0.008	0.352	0.344
CQR-Q50	x1	1.000	0.000	-0.076	0.115	0.999	0.000	-0.045	0.158	1.102	0.025	0.009	0.152	1.001	-0.001	-0.123	0.045
	x2	0.999	0.000	-0.060	0.114	0.998	-0.003	-0.027	0.158	0.959	-0.009	-0.012	0.154	1.002	0.007	-0.104	0.043
	cons	0.000	0.001	-0.078	0.119	0.004	0.003	-0.032	0.162	-0.204	0.053	0.195	0.154	0.000	-0.008	-0.105	0.045
CQR-Q90	x1	0.364	-0.002	-0.066	0.170	0.392	0.009	-0.062	0.154	0.331	-0.003	-0.075	0.053	0.306	0.014	0.148	0.270
	x2	1.255	-0.001	-0.065	0.162	1.239	-0.002	-0.040	0.157	1.269	-0.007	-0.078	0.035	1.274	-0.012	0.152	0.253
	cons	1.279	0.001	-0.066	0.164	1.216	0.012	-0.061	0.147	1.340	-0.010	-0.083	0.047	1.386	-0.040	0.268	0.266
UQR-Q10	x1	1.613	-0.002	-0.184	0.256	1.478	0.018	-0.097	0.258	1.273	0.071	0.059	0.265	1.734	-0.125	0.171	0.089
	x2	0.582	0.001	-0.097	0.107	0.565	0.006	-0.052	0.100	0.648	-0.003	-0.106	0.180	0.670	-0.084	0.219	0.042
	cons	-1.533	-0.001	-0.135	0.274	-1.390	-0.019	-0.043	0.256	-1.417	-0.042	-0.138	0.348	-1.843	0.200	0.496	0.167
UQR-Q50	x1	1.003	0.002	-0.072	0.246	1.025	-0.015	-0.053	0.273	1.099	0.037	0.045	0.284	0.922	0.072	0.152	0.181
	x2	0.850	0.001	-0.079	0.183	0.853	-0.009	-0.074	0.200	0.860	-0.031	0.037	0.219	0.839	0.042	0.077	0.139
	cons	0.170	0.000	-0.079	0.207	0.152	0.016	-0.047	0.221	0.023	0.059	0.191	0.233	0.245	-0.077	0.144	0.182
UQR-Q90	x1	0.430	-0.002	-0.047	0.077	0.446	-0.006	-0.068	0.050	0.383	-0.020	-0.075	0.000	0.429	0.009	-0.010	0.129
	x2	1.624	-0.001	-0.124	0.484	1.612	-0.024	-0.105	0.433	1.585	-0.064	-0.055	0.295	1.630	0.041	-0.083	0.558
	cons	1.212	0.002	-0.134	0.323	1.190	0.027	-0.127	0.290	1.305	0.066	-0.073	0.163	1.217	-0.043	-0.078	0.377

Table 3 Monte Carlo Simulation: N=1000, Varying coefficient structure

$y = x\beta(t)$		Type 1				Type 2			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	-1.140	0.000	-0.030	0.103	-0.092	0.011	0.033	0.154
	x2	-0.035	0.010	-0.038	0.120	-0.086	0.010	0.016	0.156
	cons	0.356	-0.010	-0.046	0.182	-0.086	-0.043	0.178	0.170
CQR-Q50	x1	0.404	-0.002	-0.052	0.067	0.845	-0.009	-0.021	0.056
	x2	0.601	-0.001	-0.050	0.065	0.841	-0.004	-0.016	0.045
	cons	0.998	0.003	-0.031	0.085	0.853	0.029	0.011	0.059
CQR-Q90	x1	1.938	0.001	-0.048	0.098	2.282	-0.001	-0.045	0.045
	x2	1.236	-0.005	-0.038	0.098	2.280	-0.010	-0.067	0.054
	cons	1.644	0.004	-0.046	0.139	2.309	-0.002	-0.066	0.073
UQR-Q10	x1	-1.211	0.002	-0.046	0.170	-0.097	0.018	-0.027	0.174
	x2	-0.044	0.002	-0.023	0.072	-0.078	0.012	-0.031	0.157
	cons	0.482	-0.004	-0.025	0.066	-0.074	-0.056	0.078	0.162
UQR-Q50	x1	0.418	-0.003	-0.059	0.119	0.900	-0.008	-0.023	0.031
	x2	0.535	-0.003	-0.071	0.112	0.737	-0.005	-0.015	0.026
	cons	0.899	0.007	-0.061	0.124	0.757	0.017	-0.031	0.028
UQR-Q90	x1	1.982	0.001	-0.079	0.150	2.214	-0.002	-0.039	0.051
	x2	1.296	0.002	-0.060	0.136	2.321	0.001	-0.022	0.077
	cons	1.778	-0.003	-0.093	0.141	2.499	-0.006	-0.031	0.057

4. Wage inequality in Grenada

This illustration focuses on an empirical application of our proposed method for the case of Grenada, focusing on the description of wage inequality trends in the country between 2013 and 2020 using the annual Labor Force Survey (LFS). This survey provides the only source of information that can be used to describe the status of the labor market and the distribution of labor income in the country.

One major limitation of this survey, however, is the collection of labor income data. Compared to standard household surveys or labor force surveys in most developed countries, labor income recorded in the LFS in Grenada is only available in brackets. Furthermore, there is a large proportion of the employed population who do not declare their labor income. Table 4 provides an overview of the labor income distribution across time.

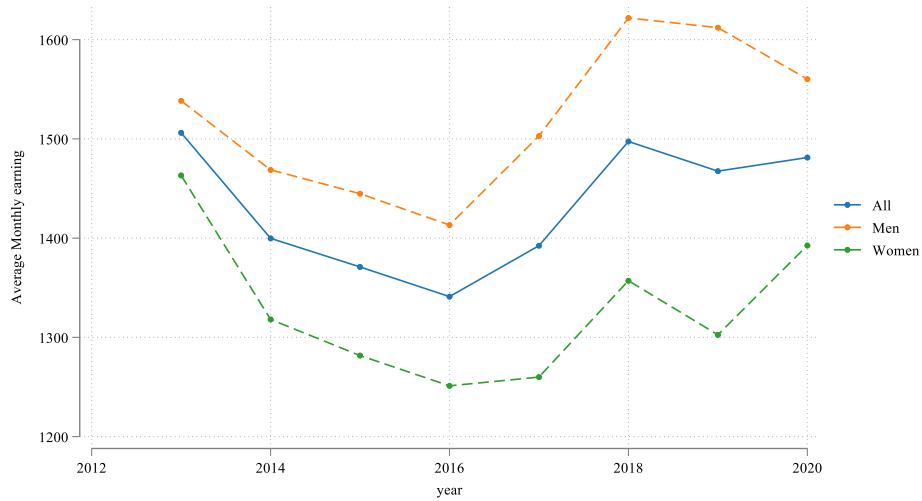
Table 4 Labor Income distribution by year

Year	2013	2014	2015	2016	2017	2018	2019	2020
>200	3.0	1.2	3.7	3.5	1.4	0.2	0.0	0.4
200-399	6.9	5.8	6.3	5.3	4.1	1.6	1.2	1.1
400-799	15.4	15.9	12.3	14.2	13.7	9.0	8.3	10.3
800-1199	19.1	20.0	18.3	18.7	21.1	20.4	23.8	24.6
1200-1999	17.7	17.4	13.9	13.1	18.4	14.7	14.9	15.9
2000-3999	15.6	11.3	11.2	11.5	10.5	9.7	12.8	11.8
4000-5999	2.6	2.4	2.4	2.2	2.2	1.6	1.2	2.1
6000+	2.0	1.2	0.6	0.6	0.7	1.0	1.0	0.5
Not stated	17.7	24.8	31.3	30.9	27.9	41.8	36.7	33.2

In this case, we face two types of problems. On the one hand, we only have access to interval-censored data, which is insufficient to analyze changes in the distribution of earnings in the country, and, on the other hand, we have an increasing proportion of individuals who do not declare income. We apply the imputation procedure previously described to address both problems, estimating the interval-censored regression for each year, with a set of household-level characteristics and job type characteristics. The sample of interest includes all adults who declared to be employed, even if they did not state their income.

We make the simplifying assumption that people who did not state income are randomly distributed conditional on observed characteristics. To account for the fact that characteristics may differ across those who did or did not state their incomes, an inverse probability weighting strategy is used to estimate the interval regression model. Finally, the imputation procedure is implemented as discussed in section 3 but assuming no lower and upper bounds for the imputed wages. Nevertheless, the maximum imputed wage for those who do not state their income is capped at the maximum predicted among those who declare their income. In all cases, imputed earnings are adjusted by inflation.

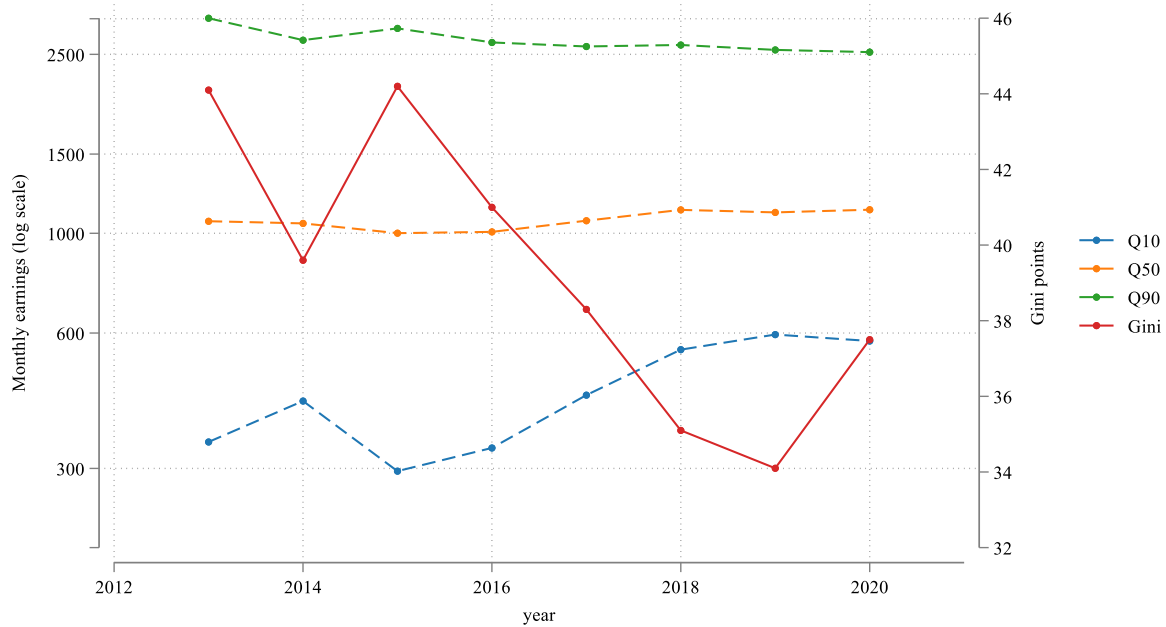
Figure 1 Average Monthly Earnings by Year and Gender



The results suggest that after a small decline in average real monthly earnings from 2013 to 2016, there was a slight improvement in the following two years, with a small decline in 2019, with average wages remaining at stable levels in 2020, despite the Covid-19 pandemic.⁷ The results also suggest that the gender earnings gap has shown a somewhat increasing trend between 2013 and 2019, although it is predicted to decline a little in 2020.

⁷ This estimate does not take into account the decline in labor force participation observed during the pandemic.

Figure 2 Selected Quantiles and Gini coefficient across Years



In terms of inequality, the estimates suggest that it has declined substantially across the years. The estimated Gini coefficient fell from 44.2 Gini points in 2015 to 34.1 in 2019, with a significant increase in 2020. This decline in inequality seems to have been driven by faster growth in the lower and middle sections of the wage distribution and a small decline in the upper section of the distribution.

5. Conclusion

In this paper, we present an imputation strategy that can be used to analyze interval-censored data. Our method proposes that a flexible enough interval regression model can be used to impute interval-censored data, which allows to recover the full distribution of data, and can be further analyzed using standard statistical methods.

The main limitation of our strategy is the assumption of conditional normality, which is required for the estimation of the interval regression model using standard software. The principles of the imputation approach, however, could be extended to allow for more flexible moment specifications, as well as error distributions.

Nevertheless, the Monte Carlo simulation suggests that as long as the latent error has a symmetric bell shaped distribution, regression analysis using the imputed data show small bias, with performance that is comparable to analyzing the uncensored data. Furthermore, when the heteroskedasticity structure is given by an exponential function, biases are small even when the latent error follows a skew or a limited distribution.

For the specific case of Grenada, we only had access to interval-censored data, which is insufficient to analyze changes in the distribution of earnings in the country, and, on the other hand, we have an increasing proportion of individuals who do not declare income. We apply the imputation procedure to address both problems, estimating the interval-censored regression for each year, with a set of household-level characteristics and job type characteristics. The results suggest that earned income inequality in this country has declined, which coincides with other economic performance indicators in the country.

References

- Angelov, A. G., & Ekström, M. (2018). Maximum likelihood estimation for survey data with informative interval censoring. *AStA Advances in Statistical Analysis*, 103(2), 217-236.
- Büttner, T., & Rässler, S. (2008). Multiple imputation of right-censored wages in the German IAB Employment Sample considering heteroscedasticity.
- Cameron, A. Colin and Trivedi, Pravin K. (2010). *Microeconometrics: Methods and Applications*. Cambridge University press.
- Chen, Y.-T. (2017). A unified approach to estimating and testing income distributions with grouped data. *Journal of Business & Economic Statistics*, pages 1–18.
- Chen Y, Zhao Y (2021) Efficient sparse estimation on interval-censored data with approximated L0 norm: Application to child mortality. *PLoS ONE* 16(4): e0249359. <https://doi.org/10.1371/journal.pone.0249359>
- Chernozhukov, V., Fernández-Val, I. & Melly, B. Fast algorithms for the quantile regression process. *Empir Econ* **62**, 7–33 (2022). <https://doi.org/10.1007/s00181-020-01898-0>
- Chih-Yuan, H., Chi-Chung, W. and Yi-Hau, C. (2021). Quantile function regression analysis for interval censored data, with application to salary survey data. *Japanese Journal of Statistics and Data Science* 72. DOI: 10.1007/s42081-021-00113-3
- Demirtas, H., S. A. Freels, and R. M. Yucel. 2008. "Plausibility of Multivariate Normality Assumption When Multiply Imputing Non-Gaussian Continuous Outcomes: A Simulation Assessment." *Journal of Statistical Computation and Simulation* 78 (1): 69–84.
- Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, 77(3), 953–973. <http://www.jstor.org/stable/40263848>
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.
- Hagenaars, A. and De Vos, K. (1988). The Definition and Measurement of Poverty. *Journal of Human Resources*, 23, 211-221. <http://dx.doi.org/10.2307/145776>
- Han, J., Meyer, B. D., & Sullivan, J. X. (2020). *Income and Poverty in the COVID-19 Pandemic* (No. w27729). National Bureau of Economic Research.
- Jann, B. (2003). The Swiss Labor Market Survey 1998 (SLMS 98). *Schmollers Jahrbuch : Zeitschrift für Wirtschaftsund Sozialwissenschaften*, 123(2), 329-335. <https://nbn-resolving.org/urn:nbn:de:0168-ss0ar-409467>
- Jenkins, S., Burkhauser, R., Feng, S., & Larrimore, J. (2011). Measuring inequality using censored data: A multiple-imputation approach to estimation and inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 174(1), 63-81.

- McDonald, J., Stoddard, O. & Walton, D. (2018) On using interval response data in experimental economics. *Journal of Behavioral and Experimental Economics* 72 (2018), 9–16.
- Machado, José A.F. & Santos Silva, J.M.C., 2019. "Quantiles via moments," *Journal of Econometrics*, Elsevier, vol. 213(1), pages 145-173.
- Moore, J. C., L. Stinson and E. Welniak. "Income Measurement Error in Surveys: A Review." *Journal of Official Statistics* 16 (2000): 331-362.
- Parolin, Z., & Wimer, C. (2020). Forecasting estimates of poverty during the COVID-19 crisis. *Poverty and Social Policy Brief*, 4(8).
- Rios-Avila, F. (2020). Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition. *The Stata Journal*, 20(1), 51–94. <https://doi.org/10.1177/1536867X20909690>
- Royston, P. (2007). Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. *Stata Journal*, 7: 445–464.
- Rubin, D. B. 1987. *Multiple Imputation for Non-response in Surveys*. New York: Wiley.
- Ting Yan, Liangqiang Qu, Zhaohai Li, Ao Yuan. "Conditional kernel density estimation for some incomplete data models." *Electron. J. Statist.* 12 (1) 1299 - 1329, 2018. <https://doi.org/10.1214/18-EJS1423>
- Vega Yon, G. G., & Quistorff, B. (2019). parallel: A command for parallel computing. *The Stata Journal*, 19(3), 667–684. <https://doi.org/10.1177/1536867X19874242>
- Walter, P., & Weimer, K. (2018). *Estimating poverty and inequality indicators using interval censored income data from the german microcensus* (No. 2018/10). *Diskussionsbeiträge*.
- Wang, X., Chen, MH. & Yan, J. Bayesian dynamic regression models for interval censored survival data with application to children dental health. *Lifetime Data Anal* 19, 297–316 (2013). <https://doi.org/10.1007/s10985-013-9246->
- Xiuqing Zhou, Yanqin Feng & Xiuli Du (2017) Quantile regression for interval censored data, *Communications in Statistics - Theory and Methods*, 46:8, 3848-3863, DOI: [10.1080/03610926.2015.1073317](https://doi.org/10.1080/03610926.2015.1073317)
- Yi-Ting Chen (2018) A Unified Approach to Estimating and Testing Income Distributions With Grouped Data, *Journal of Business & Economic Statistics*, 36:3, 438-455

Appendix

Table A1. Monte Carlo Simulation: N=500, Linear Heteroskedasticity

$y = x\beta + u * \gamma x$		$u \sim \text{normal}$				$u \sim \text{logistic}$				$u \sim \text{Chi2}$				$u \sim \text{uniform}$			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	2.005	0.023	-0.013	0.249	1.962	-0.028	-0.070	0.177	1.843	0.330	2.664	0.807	2.093	0.265	1.397	0.587
	x2	0.808	0.006	-0.036	0.162	0.802	-0.004	-0.041	0.146	0.838	0.123	0.816	0.335	0.783	0.052	0.249	0.249
	cons	-2.385	-0.025	-0.021	0.303	-2.251	0.020	-0.068	0.227	-1.988	-0.481	3.696	0.918	-2.563	-0.274	1.395	0.638
CQR-Q50	x1	1.003	0.001	-0.063	0.062	0.998	-0.001	-0.031	0.080	1.160	0.021	-0.042	0.095	1.009	0.000	-0.059	0.054
	x2	1.004	0.000	-0.057	0.063	0.999	0.000	-0.039	0.080	0.971	0.000	-0.044	0.082	0.996	0.000	-0.069	0.055
	cons	-0.004	0.000	-0.062	0.064	0.005	0.001	-0.040	0.081	-0.379	0.023	-0.044	0.104	-0.002	-0.002	-0.064	0.056
CQR-Q90	x1	-0.011	0.000	-0.050	0.105	0.040	0.005	-0.070	0.106	-0.047	-0.012	-0.050	0.054	-0.090	-0.010	-0.021	0.120
	x2	1.205	-0.001	-0.064	0.112	1.190	-0.004	-0.069	0.119	1.212	-0.002	-0.069	0.056	1.209	0.000	0.040	0.118
	cons	2.376	0.000	-0.053	0.103	2.255	0.011	-0.063	0.104	2.488	-0.006	-0.054	0.052	2.569	-0.039	0.038	0.115
UQR-Q10	x1	2.071	0.022	0.017	0.165	1.907	-0.018	-0.024	0.114	1.829	0.201	0.715	0.295	2.461	0.264	0.724	0.437
	x2	0.613	0.005	-0.029	0.074	0.594	-0.006	-0.063	0.049	0.687	0.087	0.262	0.258	0.631	0.038	0.096	0.220
	cons	-2.528	-0.016	0.021	0.149	-2.327	0.012	-0.051	0.114	-2.366	-0.190	0.436	0.348	-2.983	-0.196	0.333	0.413
UQR-Q50	x1	1.023	0.002	-0.058	0.129	1.035	-0.009	-0.065	0.158	1.169	0.039	-0.025	0.151	0.956	0.011	-0.081	0.064
	x2	0.941	0.003	-0.074	0.106	0.934	-0.001	-0.053	0.127	0.965	-0.004	-0.056	0.137	0.921	0.011	-0.093	0.058
	cons	0.111	-0.003	-0.077	0.116	0.107	0.004	-0.051	0.136	-0.218	0.023	-0.004	0.131	0.183	-0.006	-0.080	0.068
UQR-Q90	x1	0.042	0.000	-0.064	0.098	0.100	0.006	-0.077	0.104	0.018	-0.006	-0.055	0.039	0.000	-0.004	-0.049	0.164
	x2	1.470	-0.005	-0.057	0.180	1.469	-0.004	-0.074	0.207	1.430	-0.006	-0.074	0.109	1.481	-0.012	0.042	0.220
	cons	2.267	0.001	-0.046	0.118	2.158	0.010	-0.062	0.125	2.389	0.002	-0.057	0.064	2.313	0.009	-0.006	0.167

Table A2 Monte Carlo Simulation: N=500, exponential Heteroskedasticity

$y = x\beta + u * e^{\gamma x}$		$u \sim \text{normal}$				$u \sim \text{logistic}$				$u \sim \text{Chi2}$				$u \sim \text{uniform}$			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	1.643	0.000	-0.053	0.172	1.608	-0.008	-0.091	0.159	1.535	-0.029	0.217	0.422	1.692	-0.011	0.244	0.258
	x2	0.749	0.004	-0.076	0.174	0.762	0.003	-0.076	0.175	0.788	0.018	0.196	0.404	0.728	0.046	0.372	0.211
	cons	-1.287	-0.004	-0.057	0.169	-1.219	-0.013	-0.077	0.147	-1.071	-0.049	0.253	0.400	-1.383	-0.011	0.344	0.275
CQR-Q50	x1	1.001	0.001	-0.068	0.105	0.999	0.000	-0.039	0.146	1.101	0.022	-0.035	0.139	0.997	0.000	-0.124	0.052
	x2	1.002	-0.002	-0.066	0.103	1.002	-0.005	-0.042	0.143	0.958	-0.010	-0.003	0.139	0.998	0.006	-0.117	0.050
	cons	-0.003	0.001	-0.085	0.106	-0.002	0.006	-0.054	0.150	-0.202	0.055	0.090	0.140	0.003	-0.008	-0.121	0.052
CQR-Q90	x1	0.363	0.000	-0.066	0.173	0.392	0.004	-0.073	0.166	0.329	-0.003	-0.059	0.062	0.307	0.011	0.124	0.218
	x2	1.254	-0.004	-0.054	0.165	1.238	-0.004	-0.068	0.168	1.261	-0.006	-0.082	0.058	1.271	-0.014	0.144	0.201
	cons	1.278	0.003	-0.045	0.165	1.217	0.015	-0.069	0.160	1.346	-0.015	-0.067	0.061	1.385	-0.037	0.161	0.213
UQR-Q10	x1	1.596	-0.001	-0.143	0.204	1.466	0.022	-0.082	0.203	1.288	0.054	-0.057	0.209	1.724	-0.103	-0.007	0.081
	x2	0.581	0.001	-0.080	0.099	0.564	0.006	-0.052	0.094	0.654	-0.011	-0.073	0.149	0.671	-0.073	0.047	0.049
	cons	-1.527	-0.002	-0.090	0.235	-1.388	-0.022	-0.052	0.217	-1.427	-0.027	-0.135	0.288	-1.836	0.175	0.146	0.158
UQR-Q50	x1	1.020	0.004	-0.061	0.235	1.041	-0.016	-0.053	0.266	1.118	0.038	0.015	0.276	0.928	0.072	0.029	0.175
	x2	0.870	-0.002	-0.075	0.172	0.870	-0.010	-0.059	0.196	0.870	-0.027	-0.008	0.212	0.846	0.043	0.018	0.133
	cons	0.143	0.002	-0.057	0.188	0.125	0.020	-0.038	0.206	0.003	0.055	0.095	0.217	0.234	-0.076	0.043	0.166
UQR-Q90	x1	0.424	0.002	-0.042	0.087	0.436	-0.006	-0.066	0.062	0.373	-0.016	-0.074	0.011	0.421	0.005	-0.027	0.132
	x2	1.598	0.005	-0.083	0.400	1.604	-0.021	-0.108	0.367	1.572	-0.053	-0.089	0.253	1.606	0.031	-0.086	0.471
	cons	1.239	-0.006	-0.085	0.262	1.206	0.025	-0.120	0.238	1.325	0.049	-0.102	0.135	1.241	-0.032	-0.099	0.308

Table A3 Monte Carlo Simulation: N=500, Varying coefficient structure

$y = x\beta(t)$		Type 1				Type 2			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	-1.133	0.002	-0.030	0.110	-0.078	0.010	0.027	0.139
	x2	-0.036	0.008	-0.045	0.119	-0.078	0.009	0.036	0.146
	cons	0.359	-0.009	-0.067	0.169	-0.094	-0.041	0.134	0.152
CQR-Q50	x1	0.401	-0.001	-0.044	0.064	0.852	-0.009	-0.044	0.052
	x2	0.598	0.000	-0.052	0.059	0.848	-0.004	-0.002	0.042
	cons	1.002	0.001	-0.051	0.070	0.847	0.028	0.005	0.051
CQR-Q90	x1	1.933	0.001	-0.050	0.100	2.296	0.001	-0.040	0.047
	x2	1.234	-0.005	-0.039	0.099	2.276	-0.010	-0.046	0.067
	cons	1.649	0.004	-0.045	0.136	2.305	-0.001	-0.055	0.086
UQR-Q10	x1	-1.183	0.003	-0.063	0.152	-0.085	0.013	-0.015	0.166
	x2	-0.045	0.003	-0.046	0.075	-0.077	0.014	-0.018	0.149
	cons	0.471	-0.005	-0.029	0.066	-0.077	-0.053	0.036	0.152
UQR-Q50	x1	0.419	0.000	-0.061	0.122	0.914	-0.005	-0.026	0.037
	x2	0.542	-0.003	-0.053	0.109	0.746	-0.003	-0.025	0.030
	cons	0.892	0.004	-0.067	0.118	0.742	0.015	-0.030	0.032
UQR-Q90	x1	1.953	0.003	-0.055	0.123	2.215	0.000	-0.041	0.050
	x2	1.291	0.002	-0.044	0.130	2.305	-0.003	-0.033	0.076
	cons	1.800	-0.004	-0.061	0.112	2.517	-0.005	-0.037	0.049

Table A4. Monte Carlo Simulation: N=2000, Linear Heteroskedasticity

$y = x\beta + u * \gamma x$		$u \sim \text{normal}$				$u \sim \text{logistic}$				$u \sim \text{Chi2}$				$u \sim \text{uniform}$			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	2.012	0.001	-0.011	0.228	1.956	-0.032	-0.058	0.140	1.846	0.315	5.759	1.028	2.097	0.254	3.135	0.790
	x2	0.797	0.000	-0.040	0.150	0.813	-0.007	-0.060	0.121	0.831	0.120	2.261	0.414	0.785	0.053	0.433	0.343
	cons	-2.379	-0.002	-0.028	0.283	-2.253	0.025	-0.059	0.188	-1.990	-0.462	7.931	1.152	-2.576	-0.261	2.864	0.859
CQR-Q50	x1	1.000	0.001	-0.056	0.076	1.000	-0.001	-0.042	0.095	1.162	0.020	-0.016	0.090	0.997	0.001	-0.077	0.038
	x2	1.000	0.000	-0.057	0.076	1.001	-0.001	-0.044	0.099	0.967	0.000	-0.066	0.086	1.001	0.001	-0.077	0.035
	cons	-0.001	0.000	-0.059	0.076	-0.001	0.002	-0.047	0.096	-0.379	0.024	-0.025	0.087	0.002	-0.002	-0.078	0.040
CQR-Q90	x1	-0.013	0.001	-0.045	0.099	0.040	0.008	-0.073	0.092	-0.059	-0.011	-0.055	0.045	-0.099	-0.010	-0.007	0.180
	x2	1.204	-0.001	-0.047	0.107	1.193	-0.002	-0.055	0.099	1.207	-0.003	-0.059	0.041	1.217	0.000	0.036	0.178
	cons	2.380	-0.002	-0.057	0.096	2.251	0.008	-0.058	0.088	2.497	-0.003	-0.048	0.042	2.577	-0.042	0.108	0.177
UQR-Q10	x1	2.102	0.004	-0.011	0.174	1.928	-0.018	-0.091	0.123	1.849	0.168	1.117	0.282	2.605	0.208	0.772	0.504
	x2	0.611	0.001	-0.032	0.059	0.615	-0.010	-0.061	0.050	0.683	0.073	0.482	0.241	0.672	0.018	0.062	0.176
	cons	-2.537	-0.004	0.006	0.128	-2.357	0.014	-0.080	0.109	-2.374	-0.157	0.725	0.318	-3.104	-0.135	0.370	0.415
UQR-Q50	x1	0.997	0.002	-0.071	0.136	1.014	-0.010	-0.070	0.162	1.141	0.046	0.068	0.171	0.941	0.013	-0.069	0.065
	x2	0.918	0.000	-0.069	0.112	0.914	-0.003	-0.066	0.131	0.932	0.000	-0.080	0.151	0.924	0.007	-0.093	0.063
	cons	0.144	-0.001	-0.064	0.129	0.134	0.006	-0.055	0.147	-0.174	0.017	0.001	0.143	0.190	-0.003	-0.066	0.077
UQR-Q90	x1	0.051	0.001	-0.047	0.104	0.104	0.006	-0.073	0.110	0.010	-0.005	-0.051	0.035	-0.006	-0.003	-0.043	0.169
	x2	1.478	-0.001	-0.062	0.203	1.502	-0.002	-0.079	0.220	1.463	0.001	-0.083	0.127	1.474	-0.011	0.060	0.251
	cons	2.249	0.000	-0.065	0.142	2.126	0.008	-0.082	0.144	2.362	-0.003	-0.043	0.083	2.326	0.008	-0.010	0.190

Table A5 Monte Carlo Simulation: N=2000, exponential Heteroskedasticity

$y = x\beta + u * \gamma x$		$u \sim \text{normal}$				$u \sim \text{logistic}$				$u \sim \text{Chi2}$				$u \sim \text{uniform}$			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	1.639	0.000	-0.053	0.162	1.603	-0.009	-0.076	0.130	1.536	-0.030	0.442	0.541	1.693	-0.014	0.254	0.390
	x2	0.744	0.004	-0.051	0.169	0.758	0.002	-0.086	0.150	0.786	0.017	0.308	0.512	0.726	0.044	0.721	0.333
	cons	-1.280	-0.004	-0.043	0.153	-1.210	-0.013	-0.099	0.114	-1.072	-0.048	0.749	0.503	-1.386	-0.005	0.312	0.420
CQR-Q50	x1	0.999	0.001	-0.072	0.127	1.001	0.000	-0.047	0.170	1.103	0.024	0.071	0.157	1.001	-0.001	-0.118	0.038
	x2	1.000	0.000	-0.060	0.122	0.999	-0.003	-0.012	0.170	0.957	-0.008	-0.011	0.165	1.001	0.007	-0.129	0.033
	cons	0.000	-0.001	-0.044	0.129	0.000	0.004	-0.042	0.169	-0.203	0.053	0.375	0.161	-0.002	-0.008	-0.118	0.038
CQR-Q90	x1	0.363	-0.002	-0.041	0.165	0.397	0.006	-0.060	0.142	0.332	-0.003	-0.066	0.047	0.310	0.014	0.137	0.338
	x2	1.256	-0.002	-0.062	0.161	1.241	-0.002	-0.088	0.142	1.265	-0.003	-0.076	0.028	1.276	-0.012	0.161	0.317
	cons	1.279	0.002	-0.053	0.160	1.211	0.016	-0.075	0.140	1.341	-0.013	-0.063	0.045	1.383	-0.040	0.365	0.333
UQR-Q10	x1	1.621	-0.001	-0.219	0.316	1.485	0.022	-0.097	0.314	1.266	0.081	0.396	0.326	1.735	-0.141	0.500	0.106
	x2	0.587	0.000	-0.086	0.123	0.570	0.006	-0.064	0.115	0.643	0.002	-0.112	0.211	0.675	-0.091	0.542	0.037
	cons	-1.542	-0.001	-0.131	0.326	-1.401	-0.020	-0.049	0.301	-1.406	-0.052	-0.049	0.412	-1.850	0.216	1.038	0.181
UQR-Q50	x1	0.991	0.002	-0.082	0.258	1.014	-0.016	-0.038	0.289	1.082	0.036	0.143	0.298	0.916	0.070	0.329	0.196
	x2	0.843	-0.001	-0.086	0.194	0.842	-0.010	-0.046	0.219	0.844	-0.030	0.114	0.241	0.831	0.041	0.157	0.156
	cons	0.184	0.002	-0.083	0.228	0.167	0.019	-0.046	0.243	0.046	0.059	0.339	0.264	0.256	-0.074	0.330	0.200
UQR-Q90	x1	0.433	0.000	-0.028	0.072	0.455	-0.007	-0.051	0.046	0.388	-0.023	-0.071	-0.009	0.437	0.012	0.004	0.135
	x2	1.630	-0.001	-0.148	0.553	1.632	-0.027	-0.114	0.514	1.606	-0.081	-0.010	0.348	1.647	0.047	-0.100	0.682
	cons	1.201	0.001	-0.158	0.379	1.167	0.030	-0.143	0.358	1.282	0.083	-0.019	0.199	1.194	-0.049	-0.093	0.472

Table A6 Monte Carlo Simulation: N=2000, Varying coefficient structure

$y = x\beta(t)$		Type 1				Type 2			
		TRUE	Bias	MAE Ratio	StErr Ratio	TRUE	Bias	MAE Ratio	StErr Ratio
CQR-Q10	x1	-1.131	0.003	-0.074	0.101	-0.085	0.010	0.024	0.167
	x2	-0.041	0.010	-0.030	0.116	-0.090	0.011	0.021	0.158
	cons	0.358	-0.012	-0.060	0.187	-0.087	-0.043	0.208	0.184
CQR-Q50	x1	0.400	0.000	-0.041	0.072	0.849	-0.009	-0.023	0.059
	x2	0.599	-0.002	-0.057	0.072	0.845	-0.006	-0.027	0.047
	cons	0.999	0.004	-0.058	0.098	0.850	0.030	0.040	0.065
CQR-Q90	x1	1.937	0.000	-0.045	0.098	2.294	-0.007	-0.045	0.041
	x2	1.240	-0.006	-0.055	0.094	2.293	-0.015	-0.069	0.047
	cons	1.640	0.005	-0.055	0.134	2.291	0.006	-0.075	0.063
UQR-Q10	x1	-1.219	0.000	-0.087	0.194	-0.089	0.017	-0.030	0.179
	x2	-0.049	0.002	-0.037	0.069	-0.079	0.014	-0.009	0.159
	cons	0.492	-0.003	-0.039	0.067	-0.076	-0.057	0.165	0.166
UQR-Q50	x1	0.410	-0.003	-0.061	0.124	0.897	-0.007	-0.016	0.027
	x2	0.528	-0.006	-0.060	0.116	0.733	-0.006	-0.028	0.022
	cons	0.908	0.010	-0.066	0.129	0.765	0.016	-0.037	0.027
UQR-Q90	x1	1.999	0.002	-0.097	0.190	2.249	-0.002	-0.045	0.052
	x2	1.314	0.002	-0.061	0.159	2.348	-0.003	-0.031	0.081
	cons	1.747	-0.003	-0.109	0.191	2.458	-0.002	-0.044	0.072