Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operations Research

Comparing Finite Sample Confidence Intervals in Quantile Regression Models

Matthias Hofstede (612810)

Supervisor:	EAI. Herstad
Second assessor:	K. Gruber
Date final version:	29th June 2024

Abstract

This paper investigates the performance of different confidence intervals that remain valid in finite samples in quantile regression models. These intervals are compared with exogenous and endogenous data, with a focus on how the strength of the instrument affects performance. We make use of a Markov Chain Monte Carlo (MCMC) algorithm to construct confidence intervals that remain valid in finite samples. These intervals are compared with credible intervals of Bayesian methods. This comparison is based on coverage probabilities and average interval widths. We find that the Bayesian intervals are more informative when the data is exogenous. With strong instruments, both methods are equally informative. However, the Bayesian method obtains coverage probabilities closer to the nominal probability level. With weak instruments, only the Bayesian method results in reliable and informative intervals. The MCMC algorithm leads to confidence intervals that are at least 10576% wider than those of the Bayesian method. Also, its coverage probability is a least 30.2 percentage points below the nominal probability level of 95% when the instruments are weak.

1 Introduction

We can use confidence intervals to quantify the risk around an estimated coefficient in a quantile regression (QR) model. If these confidence intervals are based on asymptotic theory, they could be invalid and thus uninformative. Using them can then lead to unreliable inference. Thus, it is important to have confidence intervals in quantile regression models that remain valid in finite samples. These intervals are also called finite sample confidence intervals.

In this paper, we compare different methods that construct finite sample confidence intervals in quantile regression models. We analyse the performance of these different methods with both exogenous and endogenous data. We also investigate the effect of the strength of the instruments on the performance of the different methods.

The first method that we consider is introduced by Chernozhukov et al. (2009). They derive a Markov Chain Monte Carlo (MCMC) algorithm, a marginal grid search and a regular grid search to form finite sample confidence intervals in quantile regression models. They show that these algorithms can be used in the presence of endogeneity. Furthermore, they argue that these methods are robust against weak instruments.

A second way to obtain finite sample inference in quantile regression models is by using Bayesian quantile regression. In Bayesian quantile regression, we assume a likelihood function for the data and a prior density function for the parameters in the model. Then, we use Bayes' theorem to obtain an exact, finite sample posterior probability density function for the parameters in your model. Based on this posterior density, it is possible to construct finite sample credible intervals for the parameters in a quantile regression model. A credible interval is defined as an interval within which an unobserved parameter value falls with a particular probability. This is different from the definition of a confidence interval. However, credible intervals and confidence intervals have the same purpose and can still be compared.

Yu & Moyeed (2001) are the first to introduce a Bayesian quantile regression model. They show that using a likelihood function based on the asymmetric Laplace distribution is an effective choice for these models. Ogasawara & Kobayashi (2015) introduce an endogenous Bayesian quantile regression model based on the asymmetric Laplace distribution and the control function of Lee (2007). Their approach can be used to construct credible intervals for the parameters in an endogenous quantile regression model.

This paper is related to the literature on the comparison of Bayesian and frequentist statistics. An example of this is the paper by Gray et al. (2015). They compare the performance of Bayesian credible intervals and frequentist confidence intervals with different data-generating processes and sample sizes. They find that the Bayesian credible intervals often have a greater precision than the frequentist confidence intervals.

This paper is also related to the literature on finite sample inference in quantile regression models. Several articles investigate the finite sample performance of asymptotic confidence intervals in quantile regression models (Kocherginsky et al., 2005; Tarr, 2012). However, an extensive comparison of finite sample confidence intervals in quantile regression models does not exist in the current literature. Additionally, the effect of endogenous data with weak instruments on the performance of these confidence intervals has not been examined. Weak instruments can significantly impact the reliability of the inference in models with endogeneity (Andrews et al., 2019). This makes it crucial to understand their influence on the performance of these different methods.

So, we aim to address this gap in the literature by investigating the performance of different finite sample confidence intervals in quantile regression models. More concretely, we investigate the following research questions:

Which finite sample confidence interval results in the most reliable and informative inference with exogenous and endogenous data in a quantile regression model?

What is the influence of the strength of the instruments on the performance of these confidence intervals?

To answer these two questions, we construct credible intervals using the Bayesian quantile models of Yu & Moyeed (2001) and Ogasawara & Kobayashi (2015), in the presence of exogenous and endogenous data respectively. These credible intervals are then compared against the MCMC algorithm and the grid searches of Chernozhukov et al. (2009).

We compare the different methods based on their coverage probabilities and average interval widths using a simulation study. We find that Bayesian methods outperform the MCMC algorithm in almost all cases.

If the data is exogenous, both methods obtain conservative intervals. However, the Bayesian method has tighter average intervals and is thus more informative. When the data is endogenous, the performance is dependent on the strength of the instruments. If the instruments are strong, both methods obtain reliable and informative results. However, the Bayesian credible intervals yield a coverage probability closest to the nominal probability level. The Bayesian method is thus preferred if there are strong instruments. If the instruments are weak, the performance of the Bayesian method does not change. This method still obtains reliable and informative intervals. The performance of the MCMC algorithm does change. This method now obtains coverage probabilities at least 30.2 percentage points below the nominal probability level of 95%. Also, its average interval widths are at least 10576% wider than the average interval widths of the Bayesian method.

Finally, we look at two empirical studies. The results of the empirical studies support the claims of the simulation study. Furthermore, the empirical studies show that in a large sample, all methods converge to the asymptotic confidence intervals. This indicates that the performance of the different methods becomes identical in large samples. We also find that the algorithms of Chernozhukov et al. (2009) have the advantage that they are computationally more efficient than the Bayesian methods.

The remainder of this paper is structured as follows. Section 2 introduces the data used for the empirical studies in this paper. Section 3 introduces the quantile regression models, as well as the different algorithms to construct confidence intervals. Section 4 presents and discusses the research results. Section 5 concludes the paper.

2 Data

In this paper, we consider two empirical studies to answer our research questions. Similar to Chernozhukov et al. (2009), we consider the estimation of demand elasticity in a small sample and the impact of years of schooling on wages in a large sample.

2.1 Demand for Fish

The first data set we consider is from the Fulton fish market in New York. This dataset is first used by Graddy (1995) to test for imperfect competition in the market. This dataset ranges from December 2, 1991, to May 8, 1992, and contains 111 observations. This dataset contains the logarithm of the total quantity of fresh whiting sold on a day (quantity). This can be seen as the demand for the product. It also contains the logarithm of the average daily price of fresh whiting sold (price). Furthermore, it contains the dummy variables stormy, mixed and days. Stormy is a dummy variable that indicates the event of wave height greater than 4.5 feet and wind speed greater than 18 knots. Mixed is a dummy variable that indicates the event of wave height greater than 3.8 feet and wind speed greater than 13 knots. Days contains a dummy variable for each day of the week. We use this dataset to estimate demand elasticities.

2.2 Returns to Schooling

The second empirical study we consider is the effect of education on income. To investigate this, we use data from the 1980 US Census. This data is previously used by Angrist & Krueger (1991). This dataset contains information on 329509 men born between 1930 and 1939. From this dataset, we use the logarithm of weekly wages (WKLYWGE) and the years of completed schooling (EDUC). We also use the discrete variables SOB, YOB and QOB. SOB denotes the state of birth of an individual, YOB denotes the year of birth and QOB denotes the quarter of birth. Similar to Chernozhukov et al. (2009), we use SOB to create 51 state-of-birth dummies, YOB to create nine year-of-birth dummies and QOB to construct three quarter-of-birth dummies. We exclude the tenth YOB, fourth QOB and 52th SOB dummy to prevent multicollinearity.

3 Methodology

Suppose we have a dataset containing a dependent variable, an endogenous variable, instruments and covariates. Let y_i denote the dependent variable, d_i the endogenous variable, z_i the instruments and x_i the covariates, including a constant, for $i = 1, \ldots, N$. Let $\tau \in (0, 1)$ denote the quantile of interest. For notational convenience, let z_i^* denote a column vector of all instruments and all covariates, i.e., all regressors in a first-stage regression. Similarly, let d_i^* denote a column vector of the endogenous variable and all covariates, i.e., all regressors in a second-stage regression. We define all variables without subscripts as their matrix counterparts. When $d = z$, we have exogenous data and when $d \neq z$, there is endogeneity present. If $d = z$, a QR model for the τ-th quantile can be expressed as

$$Q_\tau(y_i|d_i^*) = q(d_i^*, \theta_\tau), \tag{1}$$

where $q(d_i^*, \theta_\tau)$ denotes a linear or non-linear function of the independent variables and coefficients. Note that $q(d_i^*, \theta_\tau) = d_i^{*'}\theta_\tau$ in the linear case. The coefficients θ_τ are estimated by minimizing

$$Q(\theta_\tau) = \sum_{i:y_i \geq q(d_i^*,\theta_\tau)}^{N} \tau|y_i - q(d_i^*, \theta_\tau)| + \sum_{i:y_i < q(d_i^*,\theta_\tau)}^{N} (1-\tau)|y_i - q(d_i^*, \theta_\tau)|, \tag{2}$$

as introduced in Koenker & Bassett Jr (1978).

If the data is endogenous, we rely on the inverse quantile regression (IQR) of Chernozhukov & Hansen (2005). The general idea behind this model is that we regress $y_i - d_i\alpha$ on both z_i and x_i using the QR minimisation defined in Equation 2. Next, we find the value of α that returns zero as the coefficient for each instrumental variable in this equation. More formally, for each α, we minimise with respect to β_τ and γ_τ

$$\begin{aligned} Q(\beta_\tau, \gamma_\tau) = & \sum_{i:y_i - d_i\alpha \geq q(z_i,x_i,\beta_\tau,\gamma_\tau)}^{N} \tau|y_i - d_i\alpha - q(z_i, x_i, \beta_\tau, \gamma_\tau)| \\ & + \sum_{i:y_i - d_i\alpha < q(z_i,x_i,\beta_\tau,\gamma_\tau)}^{N} (1-\tau)|y_i - d_i\alpha - q(z_i, x_i, \beta_\tau, \gamma_\tau)| \end{aligned} \tag{3}$$

where $q(z_i, x_i, \beta_\tau, \gamma_\tau) = z_i'\beta_\tau + x_i'\gamma_\tau$ in the linear case. So, we want to find α such that $\beta_\tau = \mathbf{0}$. This is done by defining a grid of values for α and minimising $Q(\beta_\tau, \gamma_\tau)$ for each α. Then, the value of α that minimises the Wald test statistic of excluding z, is the IQR estimate of α.

For the IQR model, we base the asymptotic confidence intervals on the asymptotic theory of the IQR estimator, as derived in Chernozhukov & Hansen (2006). For the QR model, we base the asymptotic confidence intervals on the asymptotic theory of the regular quantile regression estimator, as derived in Koenker & Bassett Jr (1978). For both of these theories, we make use of the Hall-Sheather bandwidth, as suggested in Koenker (2005).

3.1 Finite Sample Confidence Intervals Based on GMM

Chernozhukov et al. (2009) show that the generalized method-of-moments (GMM) formulation of a quantile regression is given

$$L_n(\theta) = \frac{1}{2}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_i(\theta_\tau)\right]' \left(\frac{1}{\tau(1-\tau)}\left[\frac{1}{N}\sum_{i=1}^{N} z_i^* z_i^{*\prime}\right]^{-1}\right)\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_i(\theta_\tau)\right], \qquad (4)$$

where $m_i(\theta_\tau) = [\tau - I(y_i \leq q(d_i^*, \theta_\tau)]\, z_i^*$. This formulation holds for both endogenous and exogenous quantile regression models.

The main result from Chernozhukov et al. (2009), is that $L_n(\theta)$ is in distribution equal to

$$l_n = \frac{1}{2}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\tau - B_i) z_i^{*\prime}\right] \left(\frac{1}{\tau(1-\tau)}\left[\frac{1}{N}\sum_{i=1}^{N} z_i^* z_i^{*\prime}\right]^{-1}\right)\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\tau - B_i) z_i^*\right], \qquad (5)$$

conditional on $(z_1^*, ..., z_N^*)$. Here, $(B_1, ..., B_N)$ are i.i.d. Bernoulli random variables with expectation equal to τ, independent of $(z_1^*, ..., z_N^*)$.

Using the distribution l_n, we can construct a finite sample critical value $c_n(\alpha)$. This procedure is defined as algorithm 1 in Chernozhukov et al. (2009).

First, we draw $U_{i,j}$ as i.i.d. Uniform variables for $i = 1, ..., N$. Next, we define $B_{i,j} = I(U_{i,j} \leq \tau)$, where $I(.)$ denotes the indicator function. As a third step, we compute l_n as in Equation 5. This procedure is repeated J times. Finally, we obtain $c_n(\alpha)$ as the α-quantile of the sample $(l_{n,j}, j = 1, ..., J)$.

Using a Markov Chain Monte Carlo (MCMC) algorithm, we can construct finite sample confidence intervals. This algorithm is defined as algorithm 2 in Chernozhukov et al. (2009). Suppose we have a symmetric proposal density $h(.)$ and a current value of the coefficients θ^t. In this paper, similar to Chernozhukov et al. (2009), we define $h(.)$ as a multivariate location scale t-distribution. This distribution has a mean equal to the previous estimate θ^t and a variance equal to $\frac{dof}{dof-2}V$, where V is a matrix specified by the user. The degree of freedom of this t-distribution is set to dof, which is defined as the sample size N minus the number of variables in d_i^*. In this paper, we call this algorithm the MCMC algorithm.

In the first step of the MCMC algorithm, we generate $\theta^t_{prop} \sim h(\theta^t)$. Second, we accept θ^t_{prop} as the new estimate θ^{t+1} with a probability equal to $\min(1, f(\theta^t_{prop})/f(\theta^t))$. Else we set $\theta^{t+1} = \theta^t$. In step three, we store both θ^t and $L_n(\theta^t)$. These steps should be repeated for $t = 1, .., T$. From this algorithm, we obtain a set of grid points, θ^t, and a set of values for the objective function, $L_n(\theta^t)$, for $t = 1, ..., T$. The resulting finite sample $100(1-\alpha)\%$ confidence interval then consists of all θ^t for which $L_n(\theta^t)$ is below the critical value $c_n(\alpha)$.

A second way to construct finite sample confidence intervals is by using the marginal approach. This is defined as algorithm 3 in Chernozhukov et al. (2009). This approach is suitable if only the confidence bounds for a single element of θ are of interest. For this method, we construct a grid search region for a single parameter of θ. Let us denote this element by θ_1. The remaining elements of θ are denoted by θ_{-1}. The reasoning behind the marginal approach is that θ_1 is inside the confidence interval as long as there exists at least one value of θ_{-1}, such that $L_n((\theta_1, \theta'_{-1})') \leq c_n(\alpha)$. So, for each element in the grid region of θ_1, the marginal approach finds $\theta^*_{-1} = \arg\inf_{\theta_{-1}} L_n((\theta_1, \theta'_{-1})')$. If $L_n((\theta_1, \theta^{*'}_{-1})') \leq c_n(\alpha)$, we include θ_1 in the confidence region and else we exclude it.

A third way to construct finite sample confidence intervals is by using a grid search. Going over different values for each element in θ again leads to a set of grid points, θ^t, and a set of values for the objective function, $L_n(\theta^t)$. From here, we can construct confidence intervals in the same way as for the MCMC algorithm.

3.2 Bayesian Statistics

The main idea behind Bayesian statistics is that we treat the parameters of a model as random variables. Rather than estimating fixed values for a parameter, we estimate the entire posterior distribution of the parameters. This is done based on an assumed prior distribution and a likelihood function. By Bayes' theorem, we know that

$$P(\theta \mid d^*, y) = \frac{P(d^*, y \mid \theta) \cdot P(\theta)}{P(d^*, y)} \propto P(d^*, y \mid \theta) \cdot P(\theta), \tag{6}$$

where $P(\theta \mid d^*, y)$ is the posterior distribution of θ, i.e., the distribution of interest. Furthermore, $P(\theta)$ is the prior distribution of θ and $P(d^*, y \mid \theta)$ is the joint likelihood function. It is often impossible to analytically derive the posterior distribution $P(\theta \mid d^*, y)$. However, it is possible to numerically estimate this density by making use of the result in Equation 6. The numerator can be calculated, using the assumed prior distribution and likelihood function. However, we cannot calculate the denominator. To sample from the posterior distribution, we only need to make use of the ratio of posterior densities. This means that the denominator cancels out. Thus, Bayes' theorem offers a way to sample from the posterior distribution, without the need to analytically derive it. This posterior density is not dependent on asymptotic theory and is thus valid in finite samples. So, we can use Bayesian statistics to construct finite sample credible intervals. In this paper, we use the No-U-Turn Sampler (Hoffman & Gelman, 2014) to sample from the posterior density. More details on this can be found in Appendix B.

3.2.1 Bayesian Quantile Regression

Yu & Moyeed (2001) show that the framework of Bayesian statistics can be used for quantile regression under the assumption of exogeneity. They introduce the Bayesian quantile regression model (BQR). One of the key assumptions of Bayesian statistics is a likelihood that correctly describes the data. Yu & Moyeed (2001) show that in a quantile regression setting, we can use the asymmetric Laplace distribution as the likelihood. They argue that this likelihood is a valid choice, even if the data does not follow this distribution. Thus, they suggest using the joint likelihood

$$P(d^*, y \mid \theta) = \tau^n (1 - \tau)^n \exp\left(-Q(\theta_\tau)\right), \tag{7}$$

where $Q(\theta_\tau)$ is defined as in Equation 2. Note that this slightly deviates from the regular asymmetric Laplace likelihood in the sense that the scale is fixed to one. Yu & Moyeed (2001) also show that it is possible to use an improper, flat prior distribution for θ. They prove that even in this scenario the BQR model obtains a joint posterior distribution proportional to the likelihood surface. In this paper, we specify flat priors for θ as

$$\theta \sim N(\mathbf{0}, k \cdot I), \tag{8}$$

where I is the identity matrix and k is a scalar that is set reasonably large to obtain our desired flat and uninformative priors.

3.2.2 Bayesian Instrumental Variable Quantile Regression

In this section, we introduce the Bayesian instrumental variable quantile regression model (BIVQR). For this, we rely on the Bayesian quantile regression model with endogenous variables introduced in Ogasawara & Kobayashi (2015). Consider the following instrumental variable model

$$y_i = d_i \beta_{d,q} + x_i' \beta_{x,q} + \epsilon_i, \tag{9}$$

$$d_i = x_i' \gamma_{1,\pi} + z_i' \gamma_{2,\pi} + v_i, \tag{10}$$

where q denotes the quantile of interest and π is an arbitrary quantile level, not necessarily equal to q. Ogasawara & Kobayashi (2015) show that the model introduced in Equations 9 and 10 can be written as

$$y_i = d_i \beta_{d,q} + x_i' \beta_{x,q} + \lambda_q(v_i) + \epsilon_i^*, \tag{11}$$

where $\epsilon_i^* = \epsilon_i - \lambda_q(v_i)$ and $\lambda_q(v_i)$ is the control function introduced in Lee (2007). In this paper, we assume a linear control function, i.e.,

$$\lambda_q(v_i) = (d_i - x_i' \gamma_{1,\pi} - z_i' \gamma_{2,\pi}) \eta_q. \tag{12}$$

Similar to the BQR model, we use an asymmetric Laplace distribution as the likelihood. So, we assume that ϵ_i^* follows an asymmetric Laplace distribution with shape parameter q and scale parameter one. Similarly, we assume that v_i follows an asymmetric Laplace distribution with shape parameter π and scale parameter one. We again define flat priors by using the normal distribution. Let $\beta_q = (\beta_{d,q}, \beta_{x,q}', \eta_q)'$ and $\gamma_\pi = (\gamma_{1,\pi}', \gamma_{2,\pi}')'$. Then, we assume the prior distributions

$$\gamma_\pi \sim N(\mathbf{0}, k \cdot I), \tag{13}$$

$$\beta_q \sim N(\mathbf{0}, k \cdot I), \tag{14}$$

$$\pi \sim U(0, 1), \tag{15}$$

where k is the same as in Equation 8.

3.2.3 Convergence Diagnostics

One critical aspect of Bayesian statistics is the convergence of the algorithm. In Bayesian statistics, MCMC methods are used to sample from a posterior distribution. However, if the algorithm did not converge, it implies that the samples might not be representative of the true posterior distribution. This could lead to incorrect inference. To check convergence, we use the Gelman-Rubin statistic (Gelman & Rubin, 1992). This statistic compares the variance between the different chains to the variance within the chains. If the chains converged, the between-chain variance should be close to zero. The Gelman-Rubin statistic is defined as

$$R = \frac{\hat{V}}{\hat{W}} = \frac{\frac{n-1}{n}\hat{W} + \frac{1}{n}\hat{B}}{\hat{W}},$$

where $\hat{B}$ is the estimated between chain variance and $\hat{W}$ is the estimated within chain variance averaged across all chains. The variable n denotes the number of draws. So, if $\hat{B}$ goes to 0 and $n \to \infty$, then R becomes one. Thus, we claim that a Bayesian model is converged if R is close to one. More information on the computation and definition of $\hat{W}$ and $\hat{B}$ can be found in Gelman & Rubin (1992).

4 Results

In this section, we present the results obtained from both empirical studies and the simulation study. In each subsection, we first explain the model and the variables that we use. Then we explain the initialisation and convergence of the algorithms. Finally, we introduce and explain the results. For all results, we consider three quantiles, namely $0.25, 0.5$ and 0.75, and we only look at 95% confidence intervals. So, alpha is set to 0.05 in algorithm 1, to obtain a finite sample critical value. For the MCMC algorithm, we take as a starting value for θ the parameters obtained for QR or IQR in the case of exogenous or endogenous data respectively. The covariance matrix of the quasi-posterior density $h(.)$ in Section 3.1 is set to the asymptotic covariance matrix, as introduced at the start of Section 3. For both BQR and BIVQR, we set the standard deviation of the prior distribution equal to a 100. So, we set $k = 100^2$ in Equations 8, 13 and 14. This approach leads to the desired flat priors.

4.1 Simulation Study

First, we present a simulation study to compare the different confidence intervals. For this simulation study, we only consider the MCMC algorithm and the Bayesian methods. The marginal approach is a special case of the MCMC algorithm and has similar performance. That is why we do not consider the marginal approach in this simulation study. The similarity of the two methods will be further discussed in Section 4.2, based on Figure 2. The MCMC algorithm and the Bayesian methods are compared based on average interval widths and coverage probabilities.

Ideally, we want a coverage probability of 95%, as we construct 95% confidence and credible intervals. This means that the method that achieves a coverage probability closest to this value is preferred. However, we do not want a coverage probability below the nominal probability level. Therefore, it is better to have a conservative interval than an undercovered interval.

In this paper, we investigate confidence intervals that remain valid in finite samples. So, we set the number of observations to a 100, to investigate the behaviour of the different methods in a small sample. For the simulation study, we use a 100 replications for the exogenous data and for the data with strong instruments. We increase the number of replications to 250 for the data with weak instruments to obtain more reliable results. To obtain the finite sample critical value, we make use of algorithm 1 and set the number of iterations to 2000. For the MCMC algorithm, we set the number of iterations to 10000. For the BIVQR and BQR method, we sample from 2 independent chains. Each chain uses 500 iterations to tune the parameters of the NUTS algorithm. These iterations are then discarded and can thus be considered as a burn-in period. After this, we draw 1000 samples from each chain.

In this simulation study, we consider three different data-generating processes. Namely, we consider an exogenous data set and we consider an endogenous dataset, with both weak and strong instruments. For the endogenous data set, the data-generating process is given by

$$y_i = -1 + d_i + \epsilon_i, \tag{16}$$

$$d_i = 2 + \pi z_{1,i} + \pi z_{2,i} + v_i, \tag{17}$$

where $z_{1,i} \sim N(0,1)$, $z_{2,i} \sim N(0,1)$, $\epsilon_i \sim N(0,1)$, $v_i \sim N(0,1)$ and the correlation between ϵ_i and v_i is 0.8. This data-generating process is based on the working paper of Chernozhukov et al. (2006). We consider this process for $\pi = 0.05$ and $\pi = 1$, i.e., in the case of weak and strong instruments respectively. To show the strength of the correlation between the instruments and the variable d, we perform an OLS regression in Equation 17. The average F-statistic of this regression is 1.39 and 105.16 for π equal to 0.05 and 1 respectively. This shows the presence of weak and strong correlations at the mean.

For the exogenous data set, we use the data-generating process

$$y_i = -1 + d_i + \epsilon_i,$$

where $d_i \sim N(0,1)$ and $\epsilon_i \sim N(0,1)$.

4.1.1 Weak Instruments

We first present the results for the endogenous data when $\pi = 0.05$. The coverage probabilities and average interval widths of the different methods are presented in Table 1.

Table 1: Simulation Study Results with Weak Instruments

	Coverage Probability			Average Width		
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
MCMC	0.6480	0.5400	0.5480	1461.4880	651.9350	1270.7900
BIVQR	0.9600	0.9680	0.9720	6.2300	6.1640	6.5390
Asymptotic	0.7600	0.7560	0.7920	55.9210	25.1010	44.3480

If the strength of the instruments is weak, the asymptotic confidence intervals do not obtain a correct coverage probability. This is expected, as in a sample size of a 100 observations, asymptotic theory is invalid. We also find that the intervals of the MCMC algorithm obtain a coverage probability below the desired 95% for all quantiles. The highest coverage probability that it obtains is 64.8%, which is 30.2 percentage points below the nominal probability level. These coverage probabilities are even lower than the coverage probabilities of the asymptotic confidence intervals. The BIVQR method obtains a coverage probability slightly above 95%. This indicates that the BIVQR method is the only method that results in reliable intervals if the instruments are weak.

The Bayesian method also results in the most informative intervals. It obtains an average interval width that is roughly a factor of four smaller than the asymptotic intervals. The MCMC algorithm obtains average intervals that are wide and thus uninformative. For example, when $\tau = 0.5$, the average interval width of the MCMC algorithm is 10576% larger than that of the BIVQR method. In Appendix C, we show that the results of the MCMC algorithm in Table 1 are not driven by the low number of iterations of the algorithm.

The low coverage probabilities and wide average intervals of the MCMC algorithm imply that this method can not handle endogenous data with weak instruments. At first sight, it looks like this contradicts Chernozhukov et al. (2009), as they prove that their algorithms are robust against weak instruments. However, the performance of the MCMC algorithm can be explained by its initialisation. The MCMC algorithm depends on two user-specified parameters. Namely the starting values and the covariance matrix of the proposal density. For the general MCMC algorithm, we use the point estimates of IQR as starting values and base the covariance matrix on asymptotic theory. However, these point estimates are not accurate for this data-generating process. For example, IQR returns $[32.42, -40.77]$ as the coefficients for d and the constant respectively for one of the iterations of the simulation study when $\tau = 0.75$. However, the true parameters are $[1, -1]$. For this particular instance, the MCMC algorithm returns the interval $[32.43, 14793.43]$. This interval has a width of 14761, which is 11.6 larger than the average interval width of the MCMC algorithm. This shows that the initialisation can have a large impact on the performance of the MCMC algorithm if the data contains weak instruments.

This impact can also be seen from the trace plot of the MCMC chain in this example. The trace plot is presented in Figure 3 in Appendix D. In this figure, we also present the trace if we increase the number of iterations to 300000. The trace plots show that the MCMC algorithm does not converge if it is initialised with inaccurate starting values. However, these results do not imply that the MCMC algorithm can not obtain reliable and informative inference if the instruments are weak.

To show this, we obtain intervals from the MCMC algorithm with different initialisations, using the same data as in Table 1. We consider four different initialisations. First, we initialise the algorithm with a starting value of zero for both parameters and we set the identity matrix as the covariance matrix of the proposal density. This is the naive estimator. Second, we consider the accurate estimator. This estimator takes the starting value of 0.975 for d and -1.025 for the constant. The covariance matrix is again the identity matrix. This is done to research what happens if the MCMC algorithm is initialised with accurate starting values. Third, we consider the covariance estimator. This estimator takes the same starting values as the accurate estimator. However, we set the covariance matrix to the asymptotic covariance matrix, to investigate the effect of the covariance matrix on the performance of the algorithm. Finally, we also use the MCMC2 algorithm (Chernozhukov et al., 2009). This is an algorithm that runs five independent chains of the MCMC algorithm, to better explore the parameter space. To create five different starting points, we make use of the asymptotic standard errors. By adding and/or subtracting the standard errors from the IQR estimates, we obtain five different coordinates. We then use these as the starting parameters of the MCMC2 algorithm. The reasoning for this approach is that the MCMC algorithm can get stuck at a local mode of the objective function. If this happens, the MCMC algorithm can not fully explore the parameter space. Running five independent chains allows for a better exploration of the parameter space and could potentially increase the coverage probability. Also, running five independent chains reduces the risk that the results are driven by the initialisation of the algorithm. The results of these different strategies are presented in Table 2. The rows in the table are in the same order as in which the strategies are explained. The final row of this table contains the regular MCMC algorithm. As usual, we initialise this algorithm with the IQR estimates as starting values and set the covariance matrix to the asymptotic covariance matrix of the IQR estimates. Due to computation constraints, we only consider a 100 replications.

Table 2: The MCMC Algorithm with Different Initialisations with Weak Instruments

	Coverage Probability			Average Width		
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Naive Estimator	0.6800	0.5900	0.3300	79.1700	61.9800	82.0000
Accurate Estimator	0.9600	0.9400	0.8600	39.4500	213.5200	62.6900
Covariance Estimator	0.7700	0.8300	0.2000	77.3500	62.7400	84.9300
MCMC2 Algorithm	0.8500	0.8900	0.7100	750.1200	1002.4900	867.8500
MCMC Algorithm	0.6600	0.5400	0.5300	636.1500	697.6000	770.1700

This table shows the effect of choosing different starting values and covariance matrices on the performance of the MCMC algorithm. The naive, accurate and covariance estimator obtain tighter intervals compared to the regular MCMC algorithm, for all quantiles. On top of that, all strategies outperform based on the coverage probability for τ equal to 0.25 and 0.5. The covariance estimator obtains the highest coverage probabilities. The covariance estimator even obtains correct coverage for the lowest quantile. This estimator also significantly outperforms the accurate estimator. The only difference between the two initialisations is the covariance matrix. This shows the importance of choosing a correct covariance matrix for the MCMC algorithm.

The MCMC2 algorithm significantly outperforms the MCMC algorithm, as it obtains higher coverage probabilities. This shows that the MCMC2 algorithm can somewhat elevate the problems of the MCMC algorithm if the instruments are weak. Table 2 also shows that the MCMC2 algorithm obtains wider average intervals than the MCMC algorithm. This indicates that the MCMC2 algorithm explores a larger part of the parameter space, compared to the MCMC algorithm. We will also see this in Section 4.3. Overall, Table 2 shows that the MCMC algorithm can obtain reliable confidence intervals if the instruments are weak. However, this performance is dependent on the initialisation of the algorithm. This makes the usage of the MCMC algorithm difficult when the instruments are weak. The MCMC2 algorithm can be used as an alternative to the MCMC algorithm if we have weak instruments. This algorithm leads to a better exploration of the parameter space and obtains a higher coverage probability compared to the MCMC algorithm. However, the coverage probability of the MCMC2 algorithm is still below 95%. So, based on Table 1 and 2, we can conclude that the BIVQR method is preferred, if there are weak instruments. This is the only method that obtains a correct coverage and informative intervals.

4.1.2 Strong Instruments

We now present the results for the endogenous data when $\pi = 1$ in Equation 17. Table 3 shows the coverage probabilities and the average interval widths of the different methods when the data contains strong instruments.

Table 3: Simulation Study Results with Strong Instruments

	Coverage Probability			Average Width		
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
MCMC	0.9900	0.9700	0.9900	0.7100	0.5900	0.7100
BIVQR	0.9700	0.9500	0.9600	0.7300	0.8400	0.5100
Asymptotic	0.8800	0.9300	0.9100	0.3900	0.3800	0.4000

The performance of the BIVQR method is similar to that in Table 1. The credible intervals of the Bayesian model again obtain a coverage probability slightly above 95%. This indicates that the performance of the Bayesian method is not dependent on the strength of the instruments. The results of the MCMC algorithm change substantially compared to Table 1. The MCMC algorithm now obtains coverage probabilities above the 95% level. The MCMC algorithm is even more conservative the the BIVQR method if the instruments are strong. This shows that the performance of the MCMC algorithm depends on the strength of the instruments.

Compared to Table 1, all methods obtain tighter average interval widths. This is expected, as uncertainty and variability in the estimates rise in the case of weak instruments. Strong instruments thus lead to smaller confidence intervals. The interval widths of the MCMC algorithm and the Bayesian model are very similar in Table 3. There is not one method that obtains consistently tighter intervals. This implies that both methods obtain equally informative intervals if the data contains strong instruments. This again shows that the performance of the MCMC algorithm changes substantially compared to the scenario of weak instruments.

The performance of the MCMC algorithm is independent of the initialisation if there are strong instruments. In Table 8 in Appendix E, we present the coverage probabilities and average interval widths for the naive, accurate and covariance estimator and for the MCMC2 and MCMC algorithm when data is generated with strong instruments. We find that all strategies obtain almost identical results. This indicates that the influence of the starting parameters on the performance of the MCMC algorithm also changes when the instruments are strong.

Based on Table 3, we can conclude that the BIVQR method is also preferred if there is a strong correlation between the instruments and the endogenous variable. The coverage probability of the BIVQR method is closest to the nominal level of 95%, without getting below it. This method is thus preferred to obtain finite sample intervals if there are strong instruments. However, the MCMC algorithm can also be used. This algorithm obtains reliable inference. Also, both methods are equally informative, as the average interval width of the two methods is comparable.

4.1.3 Exogenous Data

In this section, we introduce the results of the exogenous data-generating process. The coverage probabilities and average interval widths for the different methods are reported in Table 4.

Table 4: Simulation Study Results with Exogenous Data

	Coverage Probability			Average Width		
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
MCMC	1.0000	1.0000	1.0000	2.7400	2.5700	2.5400
BQR	1.0000	1.0000	1.0000	1.5200	1.7700	1.4900
Asymptotic	1.0000	1.0000	1.0000	1.6800	2.0800	2.0900

Both the confidence intervals of the MCMC algorithm and the credible intervals of the Bayesian model obtain a coverage probability of a 100%. This indicates that both methods are conservative with exogenous data. More surprisingly is that the asymptotic confidence intervals also obtain a coverage probability of a 100%, despite the small sample size. This can be explained by the normality of the data-generating process. The asymptotic confidence intervals are based on the assumption of normality. This assumption is valid if the data is normally distributed. The asymptotic confidence intervals thus perform well with this data-generating process.

The MCMC algorithm returns on average the widest confidence intervals for all quantiles. The BQR method returns the smallest average intervals when the data is exogenous. The average interval widths of the BQR method are even smaller than the average widths of the asymptotic confidence intervals. This finding will be further discussed in Section 4.2.

So, if the data is exogenous, both methods can be used to obtain reliable confidence intervals. However, the credible intervals of the BQR method are tighter and thus more informative.

4.2 Demand for Fish

In the subsection, we study the demand for fish on the Fulton fish market, using the data introduced in Section 2.1. In this empirical study, we consider a simple Cobb-Douglas demand model, similar to Chernozhukov et al. (2009). We regress the logarithm of demand on the logarithm of price. So, we are estimating the price elasticity for different quantiles, i.e., for different levels of demand. A potential problem in this model is simultaneity, as demand can influence the price. This leads to endogeneity and to solve this we make use of two instruments, namely stormy and mixed. These variables can be used as instruments since weather conditions at sea should influence the quantity of fish supplied at the market and thus the price. However, weather conditions should not influence the demand. A simple ordinary least squares regression of the logarithm of price on these two instruments and a constant results in a F-statistic of 15.83. This indicates that the instruments are correlated with the price, as noted by Chernozhukov et al. (2006).

We consider two scenarios. In the first scenario, we treat price as exogenous. In the second scenario, we treat price as endogenous, using both stormy and mixed as instruments. For both scenarios, we introduce the results with and without additional covariates, where the covariates are dummy variables for each day of the week (days). All models include a constant. To obtain a finite sample critical value, we make use of algorithm 1, where we set the number of iterations to 10000. For the MCMC algorithm, we use 50000 iterations. The grid search regions for both the marginal approach and the grid search are in Appendix F.

For both BQR and BIVQR, we sample four independent chains, where each chain uses 1000 iterations to tune the parameters of the NUTS algorithms. After this, we draw 2000 samples from each chain. All Gelman-Rubin statistics equal one, up to two decimal places. This indicates that all chains converged. The values of the statistic can be found in Table 9 in Appendix G.

In Table 5, we present the 95% level confidence intervals, obtained using the different methods, for $\tau = 0.25, 0.5, 0.75$. In Panel A, we introduce the results when the variable price is treated as exogenous and in Panel B when it is treated as endogenous. In Panels C and D, we do the same, but we add the covariates mentioned before. Similar to Chernozhukov et al. (2009), the results of the MCMC algorithm and the grid search are not reported for Panels C and D.

First, let us consider Panels A and C. Table 5 shows that the asymptotic intervals are always smaller than the intervals of the MCMC algorithm, the marginal approach and the grid search. However, these differences are not substantial, as noted by Chernozhukov et al. (2009). For example, at the median in Panel A, the difference between the width of the asymptotic interval and the width of the largest finite sample interval (Marginal) is only 0.322. These differences become larger at the 25th and 75th quantiles. When we include additional covariates in Panel C, the asymptotic confidence intervals become smaller, whereas the marginal confidence intervals become wider.

The Bayesian credible intervals are always smaller than the confidence intervals of the MCMC algorithm, the marginal approach and the grid search. In Panel A, these differences are small. However, in Panel C these differences grow, as the Bayesian credible intervals become smaller, whereas the marginal intervals become wider. The Bayesian intervals are often close to the asymptotic intervals. For example, in Panel A, when $\tau = 0.25$, the asymptotic interval has a width of 0.948. The width of the interval of the BQR method is 1.022, which is only 0.074 wider.

We also find that the width of the interval of the Bayesian method is sometimes smaller than the width of the asymptotic interval. This is similar to the result in Table 4. For example, in Panel A, when $\tau = 0.75$, the asymptotic interval has a width of 0.932 and the Bayesian interval has a width of 0.910. This can be explained by the fact that the asymptotic interval is forced to be symmetric. The Bayesian interval, however, is more flexible and can more accurately describe

Table 5: 95% Level Confidence Interval Estimates for the Demand for Fish.

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Panel A. Quantile regression			
Quantile regression (Asymptotic)	(-0.875, 0.073)	(-0.785, -0.037)	(-1.174, -0.242)
Finite sample (MCMC)	(-1.356, 0.33)	(-1.034, 0.02)	(-1.197, 0.073)
Finite sample (Grid)	(-1.375, 0.32)	(-1.015, 0.02)	(-1.195, 0.065)
Finite sample (Marginal)	(-1.37, 0.35)	(-1.03, 0.04)	(-1.21, 0.09)
Finite sample (Bayesian)	(-1.022, -0.0)	(-0.905, -0.018)	(-1.116, -0.206)
Panel B. IV Quantile Regression			
Quantile regression (Asymptotic)	(-2.483, -0.252)	(-1.766, 0.044)	(-2.033, -0.504)
Finite sample (MCMC)	(-4.186, 1.234)	(-3.569, 0.188)	(-5.263, 25.02)
Finite sample (Grid)	(-4.25, 40.0]	(-3.6, 0.2)	(-5.15, 24.85)
Finite sample (Marginal)	(-4.43, 1.0]	(-3.62, 0.2)	(-5.0, 1.0]
Finite sample (Bayesian)	(-2.978, -0.333)	(-1.896, 0.193)	(-1.962, -0.132)
Panel C. Quantile regression (covariates)			
Quantile regression (Asymptotic)	(-0.695, -0.016)	(-0.717, -0.059)	(-1.265, -0.329)
Finite sample (Marginal)	(-1.8, 0.63)	(-1.38, 0.36)	(-1.28, 0.43)
Finite sample (Bayesian)	(-0.853, -0.002)	(-0.819, 0.001)	(-1.126, -0.185)
Panel D. IV Quantile regression (covariates)			
Quantile regression (Asymptotic)	(-2.403, 0.325)	(-1.457, 0.267)	(-1.893, -0.465)
Finite sample (Marginal)	[-5.0, 1.0]	[-5.0, 1.0]	[-5.0, 1.0]
Finite sample (Bayesian)	(-2.588, 0.203)	(-2.58, 0.344)	(-2.175, 0.217)

the data if it is asymmetric. This can lead to more certainty for the Bayesian method and thus smaller intervals.

The close relation between the asymptotic confidence intervals and the Bayesian credible intervals can be attributed to the Bernstein-von Mises theorem (Doob, 1949). This theorem states that the posterior distribution of a Bayesian model converges to a multivariate normal distribution centred at the maximum likelihood estimator. This implies that the Bayesian credible intervals should converge to the frequentist asymptotic confidence intervals. We illustrate that this happens in, for example, Panel A, when $\tau = 0.75$. We do this by plotting the Bayesian posterior distribution together with the asymptotic distribution, for the variable price. The asymptotic distribution is a normal distribution, centred at the QR estimate. Its standard deviation equals the asymptotic standard error of the variable price. These two distributions, together with their 95% confidence intervals, are plotted in Figure 1.

Figure 1 shows that the two distributions obtain almost identical medians. However, the asymptotic distribution is symmetric, whereas the Bayesian posterior distribution is more flexible. This leads to a smaller interval for the Bayesian method, despite the almost identical medians. From this figure, we can also see that the two distributions are still different. This indicates that the Bayesian posterior distribution has not yet converged to the asymptotic distribution. This is not surprising as the sample size is only a 100 observations.

The results change in Panels B and D, where we treat price as endogenous. In this scenario, the differences between the intervals based on the methods of Chernozhukov et al. (2009) and the asymptotic intervals grow. These differences are again largest at the 25th and 75th quantile.

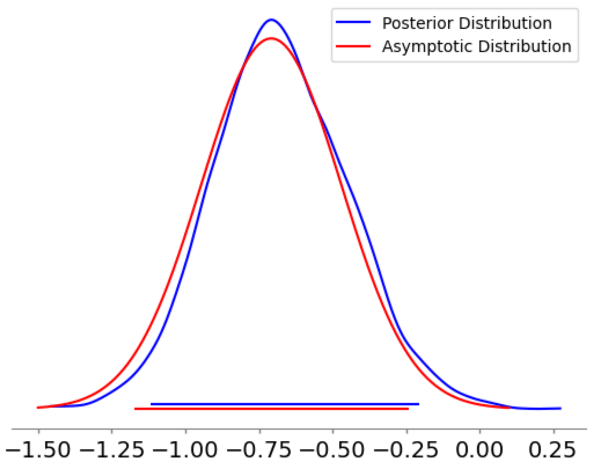

Figure 1: The Bayesian Posterior Distribution and Asymptotic Distribution for the Variable Price, when $\tau = 0.75$, in Panel A

When we add additional covariates in Panel D, the confidence intervals of the marginal approach become so wide that they include the entire grid region. The results of the Bayesian intervals stay similar to that in Panels A and C. The Bayesian intervals are again always smaller than the intervals of the MCMC algorithm, the grid search and the marginal approach. In Panel B, the Bayesian credible intervals remain close to asymptotic intervals. Even if we add additional covariates, the intervals of the BIVQR method stay relatively tight and close to the asymptotic intervals.

The findings in Table 5 support the claims of the simulation study in Section 4.1. Based on the simulation study, we expect the BIVQR method and the MCMC algorithm to perform similar if the instruments are strong. If the instruments are weak, we expect the BIVQR method to always result in the tightest intervals, as in Table 1. In this empirical study, the correlation between the instruments and the endogenous variable is stronger than in Section 4.1.1 but weaker than in Section 4.1.2. So, we expect to find that the Bayesian credible intervals are tighter than the intervals of the MCMC algorithm. However, we do not expect the differences to be as substantial as in Table 1. In Panel B of Table 5, we find that the Bayesian credible intervals are indeed always tighter than the intervals of the MCMC algorithm. However, these differences are not as large as in Table 1.

Finally, we illustrate the working of the MCMC algorithm and the grid search. Figure 2 presents the draws of the MCMC algorithm and the grid search at the median in Panel B. The left panel shows the draws of the MCMC algorithm, where the red dots are the accepted points, and the black dots fall outside the confidence interval. The right panel again shows the red dots as the accepted points of the MCMC algorithm. The solid black line is the confidence region obtained by the grid search.

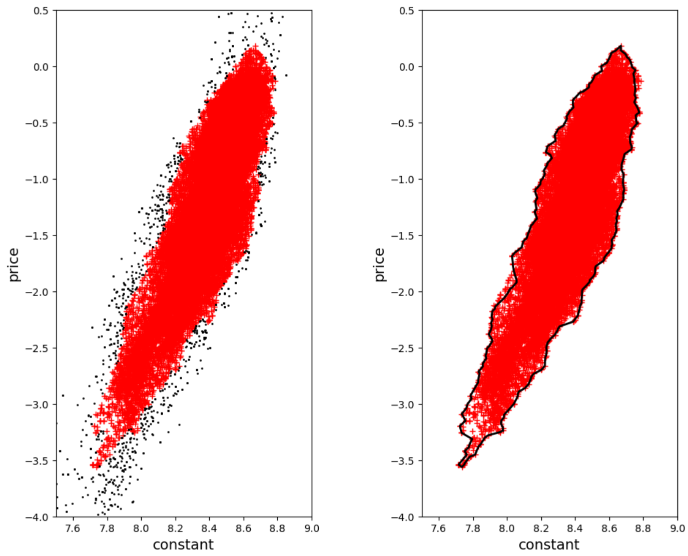

Figure 2: The Confidence Regions of the MCMC Algorithm and the Grid Search,
for Panel B, $\tau = 0.5$

The right panel shows that the confidence region of the grid search is almost identical to the border of accepted points of the MCMC algorithm. This indicates that the MCMC algorithm and the grid search obtain similar confidence regions, which shows the validity of the MCMC algorithm. However, in Table 5, we see that there are large differences between the results of the MCMC algorithm and the grid search in Panel B. These large differences may call into question the validity of the MCMC algorithm. However, as noted by Chernozhukov et al. (2009), the differences between the MCMC algorithm and the grid search can be resolved by using a different search algorithm. The differences between the marginal approach and the grid search in Panel B can be resolved by increasing the marginal grid region, as this was chosen to be $[-5, 1]$. The similarity of the two methods supports the choice of only looking at the MCMC algorithm in Section 4.1.

4.3 Returns to Schooling

In this subsection, we present the results from the second empirical study. The data of this study is introduced in Section 2.2. Here, we estimate the returns to schooling, using a simple linear quantile model, as defined in Section 3. As a dependent variable, we use the logarithm of weekly wages (WKLYWGE). The variable of interest is EDUC, i.e., the years of completed schooling. This regression suffers from endogeneity, as there is a latent variable that measures the ability of a person. This ability can potentially increase the years of schooling and the wage. Similar to Angrist & Krueger (1991) and Chernozhukov et al. (2009), we consider the quarter of birth dummies (QOB) as instruments. A simple least squares regression of EDUC on QOB and a constant results in an F-statistic of 34.01. This indicates that the correlation between the instruments and the endogenous variable is strong.

We again consider two scenarios. In the first scenario, we assume that EDUC is exogenous and in the second we assume it is endogenous. Each scenario is considered both with and without additional covariates. The additional covariates are YOB and SOB as defined in 2.2. To handle the high dimension of this dataset, we reduce the number of iterations in the different algorithms. First, to obtain a critical value, we set $J = 2000$ in algorithm 1. For the MCMC algorithm, we decrease the number of iterations to 10000. The grid search regions for both the marginal approach and the grid search can be found in Appendix F.

For the BQR model, we again sample four independent chains, where each chain uses a 1000 iterations to tune the parameters. After this, we draw 2000 samples from each chain. All Gelman-Rubin statistics equal one up to two decimal places for the BQR model. This indicates that all chains converged. The values of the statistic can be found in Table 10 in Appendix G. The BIVQR results are not reported for this empirical study. The BIVQR model has a likelihood structure that is more complex than the BQR model. This, together with a high sample size, leads to high run times and convergence issues.

In Table 6, we present the 95% level confidence intervals, obtained using the different methods, for $\tau = 0.25, 0.5, 0.75$. In Panel A, we introduce the results when the variable EDUC is treated as exogenous and in Panel B when it is treated as endogenous. In Panels C and D, we do the same but we add the covariates mentioned before. Due to computation times, the MCMC algorithm and the grid search are not reported for Panels C and D, similar to Chernozhukov et al. (2009). For the same reason, we also do not report the intervals of the Bayesian methods for these two panels. For Panel B, we also report the results of the MCMC2 algorithm, as introduced in Section 4.1.

Table 6: 95% Level Confidence Interval Estimates for the Returns to Schooling.

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Panel A. Quantile regression			
Quantile regression (Asymptotic)	(0.0715, 0.0731)	(0.0642, 0.0652)	(0.0637, 0.065)
Finite sample (MCMC)	(0.071, 0.074)	(0.064, 0.066)	(0.0637, 0.0656)
Finite sample (Grid)	(0.072, 0.074)	(0.0641, 0.065)	(0.0642, 0.0654)
Finite sample (Marginal)	(0.0706, 0.0742)	(0.0638, 0.0662)	(0.0634, 0.0658)
Finite sample (Bayesian)	(0.071, 0.074)	(0.064, 0.066)	(0.064, 0.065)
Panel B. IV Quantile Regression			
Quantile regression (Asymptotic)	(0.0767, 0.2082)	(0.0591, 0.168)	(0.0411, 0.1093)
Finite sample (MCMC)	(0.0799, 0.2094)	(0.0386, 0.1199)	(0.0120, 0.0726)
Finite sample (MCMC2)	(0.0613, 0.3796)	(0.0386, 0.1199)	(0.0115, 0.0726)
Finite sample (Grid)	(0.059, 0.197)	(0.078, 0.119)	(0.059, 0.073)
Finite sample (Marginal)	(0.05, 0.39)	(0.03, 0.13)	(0.0, 0.08)
Panel C. Quantile regression (covariates)			
Quantile regression (Asymptotic)	(0.0664, 0.0721)	(0.0645, 0.066)	(0.0628, 0.0663)
Finite sample (Marginal)	(0.0654, 0.0722)	(0.0618, 0.0678)	(0.0614, 0.0678)
Panel D. IV Quantile regression (covariates)			
Quantile regression (Asymptotic)	(0.0333, 0.1781)	(0.0479, 0.1159)	(0.0556, 0.122)
Finite sample (Marginal)	(-0.12, 1.0]	[-1, 1]	[-1.0, 0.36)

In Panels A and C, when EDUC is treated as exogenous, all methods result in similar confidence intervals. The intervals of the MCMC algorithm, the grid search and the marginal approach are almost identical to the asymptotic confidence intervals. This is expected, as asymptotic theory should perform well in such a large sample, as noted by Chernozhukov et al. (2009). Also, the Bayesian credible intervals are roughly identical to the asymptotic confidence intervals. Similar to in Section 4.2, this shows that the Bayesian credible intervals converge to the asymptotic confidence intervals.

The differences between the intervals become larger when EDUC is treated as endogenous. At the median, the intervals are still very similar. However, at the other quantiles, we see some larger differences. For example, in Panel B, the marginal approach results in an interval width of 0.34 at the lower quantile. The asymptotic interval has a width of only 0.1315. These differences become even larger when we include additional covariates. Panel D of Table 6 shows that the marginal confidence intervals include at least one bound of the grid search region for all quantiles. This implies that these bounds are uninformative.

These large differences can be explained by the shape of the objective function, as noted by Chernozhukov et al. (2009). The width of the interval is so small that it causes problems for the grid search. It also makes the transition of the MCMC algorithm difficult. The MCMC2 algorithm solves the problem of the MCMC algorithm. In Panel B, at the 25th quantile, we see that the MCMC2 algorithm better explores the parameter space and results in a wider confidence interval. This performance is similar to what we find in Table 2. There we also see that the MCMC2 algorithm can better explore the parameter space and results in wider confidence intervals. When τ equals 0.5 or 0.75, the MCMC2 and MCMC algorithm perform almost identical. This indicates that the MCMC algorithm does not get stuck at a local mode for these two quantiles.

The small width of the interval also causes problems for the BIVQR model. It causes high autocorrelation for the posterior distributions, slow sampling and convergence issues. It is known that MCMC samplers converge to a stationary distribution if the number of samples goes to infinity (Roy, 2020). This implies that the issues of the BIVQR model can be resolved by increasing the number of samples. However, this is not feasible due to computation constraints. This shows that Bayesian models have the disadvantage that they are computationally intensive, which could make them infeasible in high dimensional samples. So, in high dimensional data sets, the algorithms of Chernozhukov et al. (2009) are preferred to obtain finite sample intervals in endogenous quantile regression models.

5 Conclusion

This paper investigates the performance of different finite sample confidence intervals in quantile regression models. We consider multiple methods to generate finite sample confidence intervals. First, we consider the MCMC algorithm and the marginal approach of Chernozhukov et al. (2009). Second, we consider Bayesian techniques to obtain finite sample credible intervals.

We use a simulation study and two empirical studies to compare the different methods with exogenous and endogenous data. Of special interest is the effect of the strength of the instruments on the performance of the methods. For the simulation study, we compare the MCMC algorithm and the Bayesian methods based on coverage probabilities and average interval widths. We find that Bayesian methods outperform the MCMC algorithm in almost all cases.

If the data is exogenous, both the MCMC algorithm and the Bayesian method obtain conservative intervals. The Bayesian method results in the tightest confidence intervals on average and is thus more informative with exogenous data.

If the data is endogenous, the performance depends on the strength of the instruments. With strong instruments, both methods obtain similar interval widths. So, both methods are equally informative. Also, both methods obtain coverage probabilities above the nominal probability level. However, the coverage probabilities of the Bayesian method are closest to the target level. Thus, the Bayesian method results in more exact inference, if the instruments are strong.

If the instruments are weak, the intervals of the Bayesian method remain reliable and informative. This method still results in coverage probabilities slightly above the nominal level of 95%. So, the strength of the instruments does not influence the performance of the Bayesian credible intervals.

The MCMC algorithm does not handle weak instruments well. The highest coverage probability that it obtains is still 30.2 percentage points below the nominal probability level. The MCMC algorithm also obtains uninformative confidence intervals. The average interval width of the MCMC algorithm is at least 10576% wider than that of the Bayesian method. So, the strength of the instruments has a lot of influence on the MCMC algorithm.

This relatively bad performance of the MCMC algorithm is due to its starting parameters. These parameters have a high influence on the performance if the instruments are weak. This makes the MCMC algorithm less useful in practice. Running multiple independent chains of the MCMC algorithm improves its performance. This approach leads to a better exploration of the parameter space and a higher coverage probability. However, even this strategy does not result in a correct coverage probability. So, the Bayesian method is preferred to construct finite sample confidence intervals in quantile regression models if there are weak instruments.

When we assess the different methods on two empirical studies, we find results that align with the claims of the simulation study. We further find that the different methods perform almost identical in a large sample setting. This is because the intervals of all methods converge to the asymptotic intervals. Finally, we find that the algorithms of Chernozhukov et al. (2009) are preferred in high dimensional data sets, as they are less computationally intense.

To conclude, this research shows that Bayesian techniques are preferred over the MCMC algorithm to obtain finite sample inference in quantile regression models. However, this conclusion is based on a simulation study with only one data-generating process and one sample size. To show that this claim holds across different situations, one could extend this simulation study by considering smaller and larger sample sizes, and by considering different data-generating processes. Furthermore, this paper only consider Bayesian techniques and the algorithms of Chernozhukov et al. (2009). However, there are more techniques to obtain finite sample inference in quantile regression models. Cella (2024) introduces a distribution-free inferential model that allows for valid inference in quantile regression models in finite samples. It would be of interest for further research to see how the confidence intervals of this model compare with the intervals of the methods investigated in this paper.

References

Andrews, I., Stock, J. H. & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*, 727–753.

Angrist, J. D. & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.

Cella, L. (2024). Distribution-free inferential models: achieving finite-sample valid probabilistic inference, with emphasis on quantile regression. *International Journal of Approximate Reasoning*.

Chernozhukov, V. & Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, *73*(1), 245–261.

Chernozhukov, V. & Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, *132*(2), 491–525.

Chernozhukov, V., Hansen, C. & Jansson, M. (2006). Finite sample inference for quantile regression models. *MIT Department of Economics Working Paper*.

Chernozhukov, V., Hansen, C. & Jansson, M. (2009). Finite sample inference for quantile regression models. *Journal of Econometrics*, *152*(2), 93–103.

Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, *49*(4), 327–335.

Doob, J. L. (1949). Application of the theory of martingales. *Le calcul des probabilites et ses applications*, 23–27.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–472.

Graddy, K. (1995). Testing for imperfect competition at the fulton fish market. *The RAND Journal of Economics*, 75–92.

Gray, K., Hampton, B., Silveti-Falls, T., McConnell, A. & Bausell, C. (2015). Comparison of bayesian credible intervals to frequentist confidence intervals. *Journal of Modern Applied Statistical Methods*, *14*, 43–52.

Hoffman, M. D. & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Kocherginsky, M., He, X. & Mu, Y. (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, *14*(1), 41–55.

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

Koenker, R. & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.

Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics*, *141*(2), 1131–1158.

Ogasawara, K. & Kobayashi, G. (2015). The impact of social workers on infant mortality in inter-war tokyo: Bayesian dynamic panel quantile regression with endogenous variables. *Cliometrica*, *9*, 97–130.

Roy, V. (2020). Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, *7*, 387–412.

Tarr, G. (2012). Small sample performance of quantile regression confidence intervals. *Journal of Statistical Computation and Simulation*, *82*(1), 81–94.

Yu, K. & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*(4), 437–447.

A Programming code

All code used in this paper is attached in a zip file. In this section, we explain which files in this zip file create the tables and figures in this paper. To create these tables and figures, we only need to use the **notebooks** folder.

The direction **notebooks\Results Fish** contains notebooks to create the results of Section 4.2. It contains one notebook for each panel in Table 5. It also contains the notebooks **BIVQR-fish** and **BQR-fish** to create the Bayesian results for this table. The notebook **plot** contains the code to create Figure 2.

The direction **notebooks\Results School** contains notebooks to create the results of Section 4.3. This direction contains one notebook for each panel in Table 6. It also contains the notebook **BQR-school** to create the Bayesian credible intervals for this table.

The direction **notebooks\Simulation** contains notebooks to create the results of Section 4.1. The notebook **MCMC-convergence** contains the code to create the MCMC trace plots in Appendix D. The notebook **MCMC-initialisation** contains the code for Tables 2 and 8. The notebook **robustness-weak** contains the code for the robustness check in Table 7. The notebooks **Sim-weak**, **Sim-strong** and **Sim-exog** contain the results of the Tables 1, 3 and 4 respectively.

More details on the content of the different notebooks can be found in the **README** file provided in the zip file. This file also contains information on the location of the code for the different algorithms and models.

B Sampler for Bayesian Method

This section motivates the choice of the No-U-TURN sampler (NUTS) for the Bayesian methods introduced in Section 3.2. Metropolis-Hastings (MH) (Chib & Greenberg, 1995) is also a suitable method to sample from the posterior distribution. However, a potential problem with using MH is that it can exhibit random walk behaviour. This can make MH inefficient. To overcome this, we make use of the NUTS, introduced by Hoffman & Gelman (2014). The NUTS is an extension of Hamiltonian Monte Carlo (HMC). This is a type of MCMC that takes steps based on first-order gradient information. By doing this, it avoids the random walk behaviour of MH. The problem with HMC is that it is highly dependent on two user-specified parameters, namely the step size and the number of steps. The NUTS eliminates the need to manually set these, by automatically tuning the parameters while sampling. Similar to MH, the NUTS only uses the ratio of posterior distributions.

C Simulation Study Robustness Check

In this section, we show that the results in Table 1 are not driven by the number of iterations of the MCMC algorithm. To show this, we increase the number of iterations of the MCMC algorithm from 10000 to 50000. To obtain a finite sample critical value, we increase the number of iterations of algorithm 1 from 2000 to 10000. Table 7 presents the results of the MCMC algorithm using the same data-generating process as in Table 1.

Table 7: 95% Level Confidence Interval Estimates for the Robustness Check

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Coverage Probability(MCMC)	0.72	0.56	0.53
Average Width (MCMC)	1953.06	1758.62	2114.00

This table shows that the coverage probabilities of the MCMC algorithm stay insufficiently low if we increase the number of iterations. The coverage probabilities increase slightly compared to Table 1 but stay below the desired 95%. The average interval widths become larger than before. The MCMC algorithm thus stays uninformative.

D Trace Plot of MCMC

In this section, we introduce the trace plot of the example in Section 4.1.1. The left panel of Figure 3 shows the trace of the MCMC algorithm with the regular 10000 iterations. The right panel of Figure 3 shows the same trace but now with 300000 iterations.

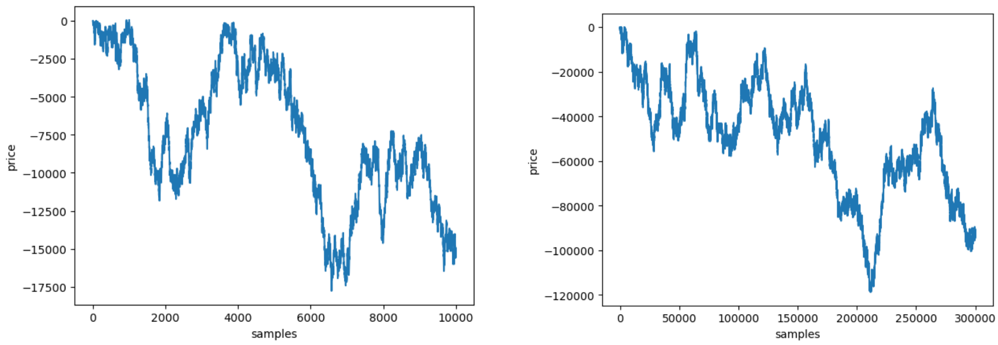

Figure 3: Trace Plot of the MCMC Algorithm with 10000 and 300000 Iterations

The left panel of Figure 3 shows that the MCMC algorithm fails to converge to the true parameter of one if we use 10000 samples. This result stays the same if we increase the number of samples to 300000. This indicates that the MCMC algorithm has convergence issues if it is initialised with inaccurate starting values.

E MCMC Algorithm Initialisation with Strong Instruments

In this section, we investigate the performance of the MCMC algorithm with different initialisations. Similar to in Table 2, we present the coverage probabilities and average interval widths for the naive, accurate and covariance estimator and for the MCMC2 and MCMC algorithm when data is generated with strong instruments. The results are in Table 8.

Table 8: The MCMC Algorithm with Different Initialisations with Strong Instruments

	Coverage Probability			Average Width		
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Naive Estimator	0.9900	0.9700	0.9700	0.6700	0.5500	0.6500
Accurate Estimator	0.9900	0.9700	0.9900	0.7200	0.5900	0.7100
Covariance Estimator	0.9900	0.9700	0.9800	0.6700	0.5400	0.6700
MCMC2 algorithm	0.9900	0.9700	0.9900	0.7400	0.6100	0.7400
MCMC algorithm	0.9900	0.9700	0.9900	0.7100	0.5900	0.7100

From here we see that the different initialisations obtain very similar coverage probabilities and average interval widths when we generate data with strong instruments. This indicates that the MCMC algorithm is only slightly influenced by its initialisations when the instruments are strong. We also find that the MCMC2 and MCMC algorithm obtain the exact same coverage probabilities. This indicates that the MCMC algorithm did not get stuck at a local mode and sufficiently explored the parameter space.

F Grid Search Regions

In this section, we present the grid search regions for the marginal approach and for the grid search. These regions are based on Chernozhukov et al. (2006). First, we present the regions of the demand for fish example, as in Section 4.2.

For the marginal approach, we need to specify a grid search region for the parameter of interest. In all scenarios, we consider an equally spaced grid of $[-5, 1]$ with a spacing of 0.01.

The regions for the grid search differ per scenario. When the variable price is treated as exogenous, we consider a grid of $[5, 10]$, with a spacing of 0.02 for the constant and $[-4, 2]$, with a spacing of 0.015 for the price. When we consider price to be endogenous, we use different regions for different quantiles. For $\tau = 0.25$, we use a grid of $[0, 10]$, with a spacing of .025, for the constant and an equally spaced grid of $[-40, 40]$ with a spacing of .25 for price. For $\tau = 0.5$, we use a grid region of $[6, 12]$ with a spacing of .0125 for the constant and a grid of $[-5, 5]$ with spacing .025 for the price. For $\tau = 0.75$, we use a grid of $[0, 30]$ with spacing .05 for the constant and a grid over $[-10, 30]$ with spacing .05 for the price.

Second, we present the regions for the returns to schooling example, as in Section 4.3. For the marginal procedure, we consider the grid $[0.055, 0.077]$ with intervals of 0.0004 for all quantiles in the exogenous scenario and the grid $[-1, 1]$, with a spacing of 0.01, for all quantiles in the endogenous scenario.

Finally, we present the search regions for the grid search approach. When EDUC is treated as exogenous, we use a grid of $[4.4, 5.6]$ for the constant, for all quantiles. The spacing is $0.01, 0.0001, 0.01$ for τ equal to $0.25, 0.5, 0.75$ respectively. For the coefficient of the variable EDUC, we use a grid of $[0.062, 0.077]$, where the spacing is $0.0001, 0.0005, 0.0001$ for τ equal to $0.25, 0.5, 0.75$ respectively. When EDUC is treated as endogenous, we consider the same grid for all quantiles. For the constant, we use a grid of $[3, 6]$, with a spacing of 0.01. For the coefficient of EDUC, we use a grid of $[0, 0.25]$, with a spacing of 0.001.

G Convergence Diagnostics

Table 9 displays the Gelman-Rubin statistic for all Bayesian models used to obtain the results in Table 5. All values are close to one, strongly indicating convergence and thus reliable results.

Table 9: Gelman-Rubin Statistic for the Bayesian Models in the Demand for Fish Study

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
BQR	1	1.0003	1.0006
BIVQR	1.0023	1.0016	1.0008
BQR (covariates)	1.0001	1.0003	1.0002
BIVQR (covariates)	1.0057	1.0095	1.0047

Table 10 shows the Gelman-Rubin statistic for all Bayesian models used to obtain the results in Table 6. All values are close to one, strongly indicating convergence and thus reliable results.

Table 10: Gelman-Rubin Statistic for the Bayesian Models in the Returns to Schooling Study

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
BQR	1.0008	1.0006	1.0008