
Finite sample valid probabilistic inference on quantile regression

ISIPTA 2023

Leonardo Cella: www.leocella.com



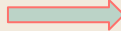
Introduction

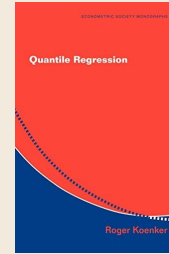
Setup: Data $Z^n = (Z_1, \dots, Z_n)$, $Z_i = (X_i, Y_i)$, are iid with distribution P .

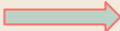
Quantile Regression:

- Let $Q_x(\tau)$ denote the τ -th quantile of Y given $X=x$
- $Q_x(\tau) = x^\top \theta$
- **Goal:** Make inferences on θ
 - distribution free
 - valid

Traditional methods

- Inferences through confidence regions 



- Notion of validity is familiar  coverage guarantees:

$$\sup_P P^n \{C_\alpha(Z^n) \not\subseteq \theta(P)\} \leq \alpha, \quad \alpha \in [0, 1].$$

- Traditional methods usually achieve this asymptotically

Can we do more?

Probabilistic inference

- Assign degrees of belief to general assertions about θ
- **Validity:** Control the assignment of **high** degrees of belief to **false** assertions

$$\sup_{P: \theta(P) \notin A} P^n \{ \Pi_{Z^n}(A) > 1 - \alpha \} \leq \alpha, \quad \alpha \in [0, 1].$$

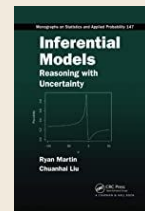
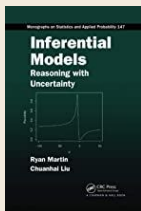
- Bayes comes to mind, but **False confidence theorem** says we need imprecision!

Satellite conjunction analysis and the false confidence theorem

Michael Scott Balch , Ryan Martin and Scott Ferson

Published: 17 July 2019 | <https://doi.org/10.1098/rspa.2018.0565>

“Degrees of belief that are
(additive) probabilities are at risk
of being invalid”



Inferential Models

Imprecise probabilistic approach to statistical inference that has validity as its main pillar

Overview

Setup: Data $Z^n = (Z_1, \dots, Z_n)$ is iid with distribution P_ω

Two-step IM construction

1. Choose an appropriate $h : (\mathbb{Z}^n \times \Omega) \rightarrow \mathbb{R}$ that determines a partial ordering of candidate values for ω given z^n , e.g., likelihood ratio:

$$h(z^n, \omega) = L_{z^n}(\omega) / L_{z^n}(\hat{\omega}_{z^n})$$

2. Compute the possibility contour

$$\pi_{z^n}(\omega) = P_\omega^n \{h(Z^n, \omega) \leq h(z^n, \omega)\}$$

Degrees of
belief

Confidence
Regions

Back to quantile regression...

- Data $Z^n = (Z_1, \dots, Z_n)$ are iid with distribution P
- $Z_i = (X_i, Y_i)$
- $Q_x(\tau) = x^T \theta \rightarrow \tau$ -th quantile of Y given $X=x$
- $\theta = \theta(P)$: **Inferential target**

IM construction

1- Choose an appropriate h that orders candidate values for θ given data

2 - Compute the contour $\pi_{z^n}(\theta) = P^n\{h(Z^n, \theta) \leq h(z^n, \theta)\}$

Theorem:

This construction yields valid degrees of belief for assertions about θ , as well as valid confidence regions for θ .

Problems....

1- Choose h



No model



No likelihood ratio

2 - Compute the contour

$$\pi_{Z^n}(\theta) = P^n\{h(Z^n, \theta) \leq h(z^n, \theta)\}$$



P is not known

Possible solution



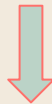
h in terms of relative risk

π approximated by the bootstrap

Due to the bootstrap, validity is just achieved asymptotically!

Basic idea

Choose an h whose distribution is known and independent of unknown quantities



$\pi_{z^n}(\theta) = P^n\{h(Z^n, \theta) \leq h(z^n, \theta)\}$ can be computed

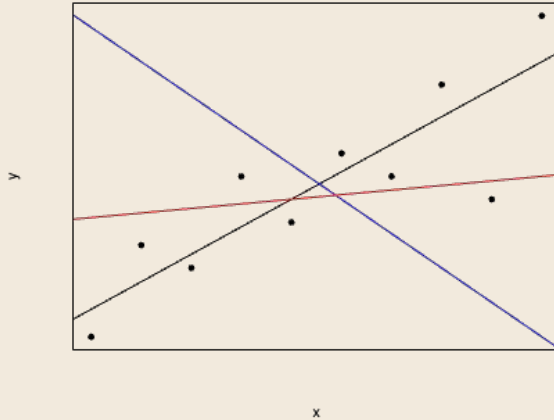
But how to obtain such h ?

- 1) Find a function γ of data and θ that is a pivot
- 2) Set h as the γ 's probability mass

Intuitively...

$$\gamma = \sum_{i=1}^n I_{(0,\infty)}(Y_i - x_i^\top \theta) \implies \gamma \sim \text{Bin}(n, 1 - \tau) \implies h = \binom{n}{\gamma} (1 - \tau)^\gamma \tau^{n-\gamma}$$

But not efficient...



New consideration: Discrete case first



Consider the independent binomials at each level of X separately!

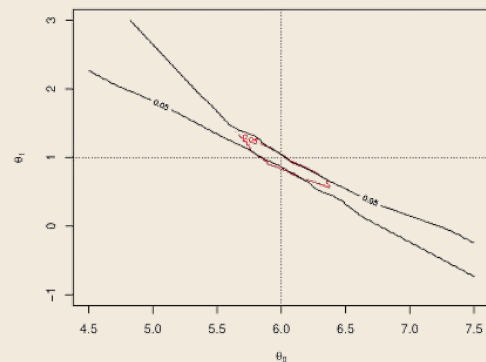
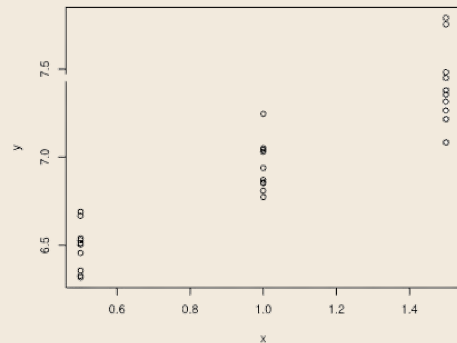


$$h = \prod_{i=1}^k \binom{n_i}{\gamma_i} (1-\tau)^{\gamma_i} \tau^{n_i-\gamma_i},$$

where

$$\gamma_i = \sum_{j=1}^{n_i} I_{(0,\infty)}(Y_j - x_i\theta), \quad i = 1, \dots, k.$$

Example: $n=30, \tau=0.5$



Continuous X

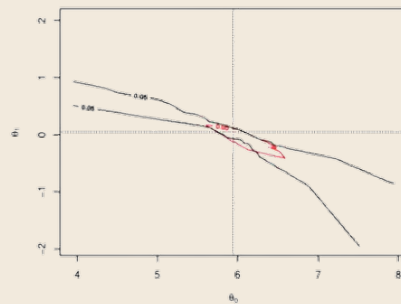
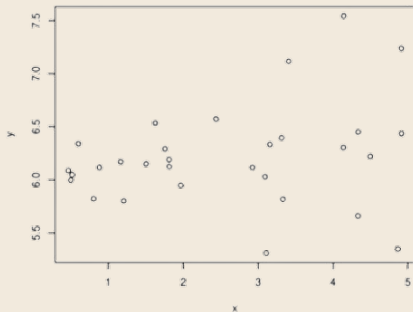
If X is continuous, **there is no replication** of Y for any given $X = x$



But we do have replications of Y in neighborhoods of X !

- Form k neighborhoods of X
- Consider each one of the k independent binomials separately
- h : product of their probability masses

Example: $n=30$, $\tau=0.3$, $k=2$



This was fast, but....

- More details
- More examples
- Comparisons with other methods
- Open questions

Finite sample valid probabilistic inference on quantile regression

Leonardo Cella

Department of Statistical Sciences, Wake Forest University, USA

CELLA@WFU.EDU



Wake Forest University
Department of Statistical Sciences



Finite sample valid probabilistic inference on quantile regression

Leonardo Cella | locella@wfu.edu

1 - Introduction

Data $\mathcal{D}^n = \{(x_i, Y_i) : i = 1, \dots, n\}$ of n covariate-response pairs are fit with distribution F . Seeking a τ -quantile about F .

- $Q_\tau(Y) = \tau^*$ is the τ -th quantile of Y given $X = x$.
- $\hat{Q}_\tau(Y)$ makes inference on τ that can accommodate pre and valid.
- Common solution: Inference through confidence regions.
- Notion of validity is flexible \rightarrow coverage guarantee:

$$\sup_{\theta \in \Theta} P^*(C_{\tau, \hat{Q}_\tau}(\theta) \ni \theta | F) \leq \alpha, \quad \alpha \in [0, 1]. \quad (1)$$

2 - Probabilistic Inference

Elemental confidence region, where we can assign degree of belief to false assertions.

- Validity: control the maximum of high degree of belief to false assertions:

$$\sup_{\theta \in \Theta} P^*(\Pi_{\tau, \hat{Q}_\tau}(\theta) \ni \theta | F) \leq \alpha, \quad \alpha \in [0, 1]. \quad (2)$$
- Bayesian approach? False Confidence Theorem says we need impractical.

3 - Inferential Models

Consider the parameter space where $\mathcal{Z}^n = \{Z_1, \dots, Z_n\}$ are iid with distribution $P_{\theta, \tau}$. IM approach [1] often valid probabilistic inference for ω .

Two-step IM construction:

1. Choose an appropriate $\theta : \mathcal{Z}^n \rightarrow \Theta$ that determines a partial ordering of candidate values for ω given Z^n , e.g. likelihood ratio:

$$h_{\theta}(Z^n, \omega) = \tau_{\theta}(Z^n, \omega) / \tau_{\theta}(Z^n, \omega_0).$$

2. Compute the possibility contour:

$$\tau_{\theta}(Z^n) = P^*(h_{\theta}(Z^n, \omega) \leq M_{\theta}^*(\omega) | Z^n)$$

$\theta \rightarrow$ valid probabilistic inference and confidence region for ω .

4 - Nonparametric IM for quantile regression

The idea here is to relax the construction above:

1. Choose ω that orders candidate values for θ given Z^n .

2. Compute the contour:

$$\tau_{\theta}(Z^n) = P^*(h_{\theta}(Z^n, \omega) \leq M_{\theta}^*(\omega) | Z^n), \quad \theta \in \Theta. \quad (3)$$

Theorem:

- The degree of belief obtained from (3) are valid in the sense of (2).
- $\{P^*(\theta \in \Theta | \tau_{\theta}(Z^n) > \alpha)\}$ is a valid confidence region in the sense of (1).

Challenges:

1. What if θ is not \rightarrow to likelihood ratio.
2. How to compute (3)? Found that P^* is unknown.

A possible solution:

- In [3], a bootstrap-based IM construction was proposed
- Validity is just achieved asymptotically.

But we want more \rightarrow IM that achieves validity for any sample size.

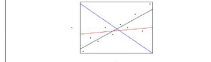
Strategy: Choose ω whose distribution is known and independent of unknown quantities, so (3) can be computed! More specifically:

- Find (θ, ω) that is a pivot
- $\omega \sim \tau$ probability mass

4.1 - An intuitive (but bad) solution

Data $\mathcal{D}^n = \{(x_i, Y_i) : i = 1, \dots, n\}$ of n covariate-response pairs are fit with distribution F . Seeking a τ -quantile about F .

- $Q_\tau(Y) = \tau^*$ is the τ -th quantile of Y given $X = x$.
- Very naïve! For example, for $\tau = 0.5$, one line that splits the data in half, e.g., the black red and blue lines, is equally maximally plausible.



4.2 - A better solution

Let X be discrete with k levels. The idea is to consider the likelihood for each level of X separately, and take θ as the product of their probability masses:

$$\theta = \prod_{i=1}^k \binom{n_i}{n_i}^{-1} \tau_i^{n_i} \omega_i^{-n_i}, \quad \text{where } \tau_i = \sum_{j=1}^k h_{\theta}(x_j, \omega) \omega_j. \quad (4)$$

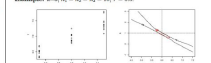


Figure 2. There are $\binom{n_i}{n_i}$ ways to split the data in half.

If X is continuous, there is no replication of Y for any given $X = x$. But we do have replications of Y in neighborhoods of X , so (4) can still be used!

Form θ neighborhoods of X .

• Consider each one of the k independent intervals separately.

• Use the product of their probability masses as the quantity of order θ .

Example: Simulations study with $n = 30$, $\tau = 0.5$ and $k = 2$ to compare the coverage probabilities and mean lengths of 95% nested confidence for the quantile regression confidence based on the IM and two other methods:

τ	IM	IM	IM
0.5	0.95	0.95	0.95
0.25	0.95	0.95	0.95
0.75	0.95	0.95	0.95

5 - Open questions

Other pivot options? How to best select the neighborhoods of a continuous X ? Specifically, does the number of neighborhoods and/or the manner of replication per neighborhood impact the efficiency of the IM?

References

- [1]. Cella and B. Martin. Exact and approximately valid probabilistic inference on a class of quantified functions. *International Journal of Approximate Reasoning*, 151:265–278, 2022.
- [2]. B. Martin and C. Liu. *Approximate Models: Reasoning with Uncertainty*. Monographs in Statistics and Applied Probability Series. Chapman & Hall/CRC Press, 2015.