An MCMC approach to classical estimation — V. Chernozhukov and H. Hong

Julien Pascal

Sciences Po Reading Group

07/10/2016

Introduction

V. Chernozhukov and H. Hong (2003) *An MCMC Approach to Classical Estimation* [2]

In this paper:

- Definition of a new type of estimators: Laplace type estimators (LTEs)
- Theoretically and computationally attractive
- Using MCMC methods to calculate LTEs

Outline

1. Estimation problem
2. LTEs as a solution
3. MCMC: Calculating LTEs
4. Theoretical properties of LTEs

Worked example: censored median regression, J. L. Powell (1984) [7]

Censored Median Regression

Setting: data in generated from:

$$t = 1, \dots T, \tilde{y}_t = x_t'\beta_0 + u_t$$

But the researcher only observes y_t and x_t:

$$t = 1, \dots T, y_t = max(0, x_t'\beta_0 + u_t)$$

With:

- the dependent variable y_t and the regression vector x_t are observed
- the parameter vector β_0 and the error term u_t are unobserved

Censored Median Regression

Powell (1984): censored LAD estimator

$$\hat{\beta} = argminS(\beta) = (1/T)\sum_{n=1}^{T} |y_t - max(0, x_t'\beta)|$$

- Under some conditions (see [7]), the censored LAD estimator is consistent and asymptotically normal
- Issue: difficult to find the minimum (not a $C1$ function)

LTEs as a solution

Heuristic: Use Bayesian methods to non-Bayesian problems.

Bayesian approach:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

With:

$$p(\theta|y) = \textit{posterior}$$

$$p(y|\theta) = \textit{likelihood}$$

$$p(\theta) = \textit{prior}$$

LTEs as a solution

"Quasi-Bayesian" approach:

$$p_n(\theta) \propto exp(Ln(\theta))\pi(\theta)$$

With:

$$p_n(\theta) = quasi - posterior$$

$$Ln(\theta) = objective$$

$$\pi(\theta) = prior$$

$Ln(\theta)$ objective function to be maximised.
With the censored median regression:

$$Ln(\theta) = -(1/T)\sum_{n=1}^{T} |y_t - max(0, x_t'\theta)|$$

LTEs as a solution

Transformation of $L_n(.)$:

$$p_n(\theta) = \frac{exp(Ln(\theta))\pi(\theta)}{\int_\Theta exp(Ln(\theta))\pi(\theta)d\theta}$$

$p_n(\theta)$ is a proper density over θ.
Can use Bayesian estimation tools (e.g. quasi-posterior mean).

Claim

Under mild existence and regularity conditions, the quasi-posterior can be used for inference. (see later).

LTE for the Censored Median Regression

We are after $p_n(\theta)$ to calculate:

- the quasi-posterior mean:

$$\hat{\theta} = \int_{\Theta} \theta p_n(\theta) d\theta$$

- the quasi-posterior median, etc.

Calculation of $p_n(\theta)$

Because of the denominator, $p_n(\theta)$ has a complicated form $\rightarrow$ use MCMC

Monte Carlo Method

We can approximate

$$\int_{\Theta} \theta p_n(\theta) d\theta$$

by drawing an i.i.d. sample from $p_n(\theta)$:

$$\theta^{(1)}, ..., \theta^{(n)}$$

and calculating:

$$(1/n) \sum_{i=1}^{n} \theta^{(i)}$$

$$(1/n) \sum_{i=1}^{n} \theta^{(i)} \xrightarrow[\infty]{n} \int_{\Theta} \theta p_n(\theta) d\theta$$

MCMC — Theory

Setting:

$$p_n(\theta) = \tilde{p}_n(\theta)/Z$$

- i.i.d. sampling from $p_n(\theta)$ is difficult because calculating Z is complicated
- $\tilde{p}_n(\theta)$ is easy to evaluate

Idea: Use $\tilde{p}_n(.)$ to generate a (correlated) series of draws $\theta^{(1)}, ..., \theta^{(n)}$ with stationary distribution $p_n(.)$

MCMC — Theory

Metropolis algorithm [6]: Assume $\theta \in \chi$ with χ a discrete space.

1. $S \leftarrow$ symmetric matrix
2. $\theta^{(1)} \leftarrow$ element of χ
3. for $i = 2, ..., n$
   - draw $\theta^{(*)} = S(\theta^{(i-1)}, .)$
   - draw u from $Uniform(0, 1)$
   - if $u < \tilde{p}_n(\theta^{(*)})/\tilde{p}_n(\theta^{(i-1)})$
     $\theta^{(i)} = \theta^{(*)}$
   - else
     $\theta^{(i)} = \theta^{(i-1)}$
4. Output $= \theta^{(1)}, \theta^{(2)}, ..., \theta^{(n)}$

Remark: In the paper, the authors use the Metropolis-Hastings algorithm [3]

MCMC — Theory

Claim: Let T be the transition matrix for the sequence of $\theta^{(1)}, ..., \theta^{(n)}$ defined above:

$$\forall (x, y) \in \chi^2, T(x, y) = P(X_{i+1} = x | X_i = y)$$

- the sequence $\theta^{(1)}, ..., \theta^{(n)}$ has stationary distribution $p_n(.)$
- if $\forall x \in \chi$, $p_n(x) > 0$: T is irreducible and aperiodic

From ergodic theory:

$$(1/n) \sum_{i=1}^{n} f(\theta^{(i)}) \xrightarrow{a.s} \int_{\Theta} f(\theta) p_n(\theta) d\theta$$

For $f : \Theta \rightarrow R$ a C^1 function.

MCMC — Example

Calculate $E(X)$ with $X \sim Beta(\alpha, \beta)$

$$Beta(\theta|\alpha,\beta) = (1/B(\alpha,\beta))\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Here:

$$\tilde{p_n}(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$Z = B(\alpha,\beta)$$

Simulation in Julia with $\alpha = 1/2$ and $\beta = 1/2$: Link

$iterations = 1.10^7$

MCMC — Example

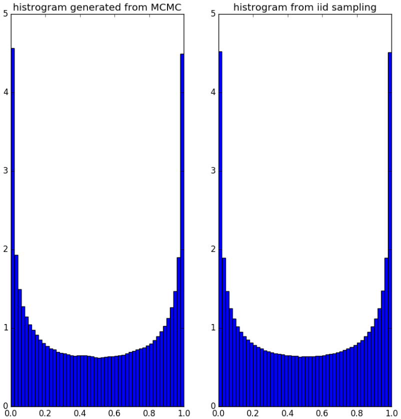

Figure 1: MCMC (Metropolis) v.s. iid sampling

F Find $(1/n)\sum_{i=1}^{n}\theta^{(i)} \approx 0.495$. True value $= 0.5$

Back to the Censored Median Regression Model

- Metropolis-Hastings with:

$$\tilde{p_n}(\theta) = exp(Ln(\theta))\pi(\theta)$$

$$Z = \int_{\Theta} exp(Ln(\theta))\pi(\theta)d\theta$$

- The authors find that the resulting LTEs outperform the most common alternative (Buchinsky, 1991 [1])
- LTEs: get both valid point estimate and confidence intervals

LTE applied to non-linear IV-QR

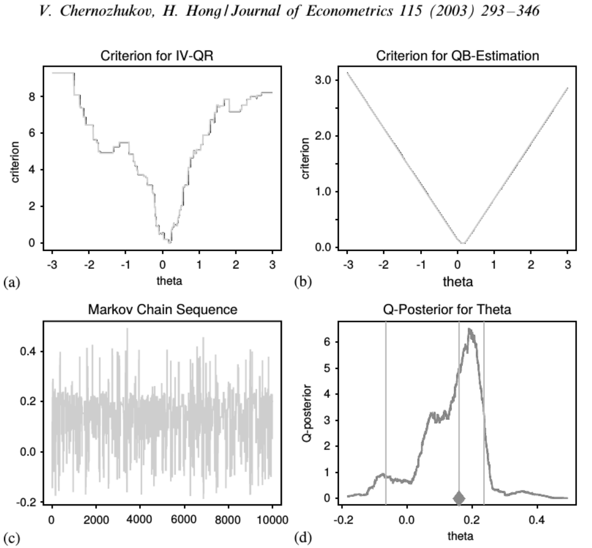

Figure 2: Non-linear IV-QR

Theoretical properties of LTEs - Definition

Define:

$$Q_n(\zeta) = \int_\Theta \rho_n(\theta - \zeta) p_n(\theta) d\theta$$

with:

$$p_n(\theta) = \frac{exp(Ln(\theta))\pi(\theta)}{\int_\Theta exp(Ln(\theta))\pi(\theta) d\theta}$$

and $\rho_n(.)$ a given loss function (see next slide)

Definition LTE

The class of LTEs minimizes the function $Q_n(\zeta)$ for various choices of ρ_n:

$$\hat{\theta} = arginf_{\zeta \in \Theta}[Q_n(\zeta)] \tag{1}$$

Loss functions

Penalty associated with making an incorrect decision:

- $\rho_n(u) = |\sqrt{n}u|^2$
  Solution of (1): quasi-posterior mean
- $\rho_n(u) = \sqrt{n}\sum_{j=1}^{d}|u_j|$
  Solution of (1): quasi-posterior median
- $\rho_n(u) = \sqrt{n}\sum_{j=1}^{d}(\tau_j - 1(u_j \leq 0))u_j$, with $\tau_j \in (0,1)$ (see, Koenker and Bassett(1978) [5])
  Solution of (1): quasi-posterior marginal τ_jth quantile

Assumptions required

Assumption 1: parameter

The true parameter θ_0 belongs to the interior of a compact convex subset Θ of the Euclidean space R^d.

Assumption 2: loss and weighting functions

The loss function $\rho_n : R^d \to R_+$ respects some convexity properties (see paper) and the weighting function $\pi : \Theta \to R_+$ is a continuous density function.

Assumptions required

Assumption 3: identifiability

$\forall \delta > 0, \exists \epsilon > 0$ such that

$$liminf_{n \to \infty} P_* (\sup_{|\theta - \theta_0| \geq \delta} (1/n)(L_n(\theta) - L_n(\theta_0)) \leq -\epsilon) = 1$$

Assumption 4: expansion

In a neighborhood of θ_0, one can represent $L_n(\theta)$ as a Taylor expansion around the true value θ_0 + impose regularity conditions on the expansion (see paper).

Assumptions required

Intuition for assumption 3

Assumption 3 holds if $L_n(\theta)/n$ converges to a criterion function $M(\theta)$, which is maximized uniquely at θ_0.

Theoretical properties of LTEs - Asymptotics

- Under assumptions 1-4: quasi-posterior $p_n(.)$ concentrates at the speed $1/\sqrt{n}$ around the true parameter θ_0
- quasi-posterior quantiles provide asymptotically valid confidence intervals

- Models for which the regularity conditions apply:
  - Censored quantile regression
  - Instrumental quantile regression
  - GMM
  - M-estimators
  - etc.

Conclusion

- One can apply Bayesian techniques to non-Bayesian problems (under some regularity conditions)
- The use of MCMC makes LTEs computationally attractive
- Similar set of tools are being applied in Macro (e.g. DSGE [4])

M Buchinsky.
Methodological issues in quantile regression, chapter 1 of the theory and practice of quantile regression.
PH dissertation, Harvard University, 1991.

V. Chernozhukov and H. Hong.
An mcmc approach to classical estimation.
Journal of Econometrics, 115(2):293–346, 2003.

W. K. Hastings.
Monte carlo sampling methods using markov chains and their applications.
Biometrika, 57(1):97–109, 1970.

E. Herbst and F. Schorfheide.
Bayesian inference for dsge models.
Manuscript. Board of Governors of the Federal Reserve System, Washington, DC, 2014.

R. Koenker and Bassett Jr. G.
Regression quantiles.

Econometrica: journal of the Econometric Society, pages 33–50, 1978.

N. Metroplis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller.
Equation of state calculations by fast computing machines.
J. chem. Phys., 21(6):1986–1092, 1953.

J. L Powell.
Least absolute deviations estimation for the censored regression model.
Journal of Econometrics, 25(3):303–325, 1984.