Conformal Inference for Heterogeneous Policy Effect

Jiayi Cheng Peilin Yang *

August 24, 2022

1 Continuous Treatment Variable

Background Evaluating of policy effect is the vital problem in economics, but it is always intertwined with heterogeneous distribution of agents with a spectrum of state variables such as asset, age, education etc. Differentiating from traditional causal inference problem, policy in real world always comes with different strength. A typical example is tax rebate, people gain different quantity according to their age and salary. An ensuing question is even though the distribution data is given ex ante, the estimation of extreme value part is inaccurate, which contaminates the evaluation of policy. Therefore, promotion of off-the-shelf tools of estimating is always necessary. In what follows, we will give an example to illustrate a typical problem we face in macroeconomics.

Example This is an example of heterogeneous response model, which has been explored by several important works in economics (Johnson et al. [2006], Parker et al. [2013], Misra and Surico [2014]). In year 2001 and 2008, the US government has issued a series of policy to resolve urgent crisis during that period. As the mailing of rebate is a random behavior considering the sequence is predetermined by the penultimate digit of social security number (SSN), the arrival of tax rebate can be viewed as exogenous variable vis-à-vis personal characteristic. If the effect of tax rebate stimulus is homogenous, we have following specification:

$$\Delta C_{it+1} = \sum_s \beta_{0s} \times M_s + \beta_1' X_{it} + \beta_2 R_{it+1} + u_{it+1}$$

where ΔC_{it+1} is the response of consumption expenditure of individual i at time $t+1$, M_s is designed to absorb the monthly effect of consumption or other common factors; X_{it} controls heterogeneous effects; R_{it+1} is key variables we are interested in, the tax rebate; u_{it+1} is unobservable factor.

However, the real case is individual on different asset level will have different response to the policy. In light of this, we adjust the specification: The potential outcome distribution on different quantile is

$$Q_{\Delta C_{it+1}|R_{it+1}, M_s, X_{it}}(\tau), \quad \tau \in (0,1)$$

*Stanford University, Graduate School of Business; jiayic@stanford.edu; peiliny@stanford.edu.

we specify a linear conditional quantile

$$q\left(R_{it+1}, X_{it}, M_s, \tau\right) = Q_{\Delta C_{i,t+1|}}(\tau) = \sum_s \alpha_{0s}(\tau) M_s + \alpha_1(\tau)' X_{it} + \alpha_2(\tau) R_{it+1}$$

Figure 1 shows the results of quantile estimation. On different percentile τ, the marginal effect of tax rebate varies. The difference compared to Lei and Candès [2021] (equation 2) is how to estimate the continuous treatment effect instead of $T = \{0, 1\}$.

$$\hat{C}(x) = \left[\hat{q}_{\alpha_{\mathrm{lo}}}\left(x; \mathcal{Z}_{\mathrm{tr}}\right) - \eta(x), \hat{q}_{\alpha_{\mathrm{hi}}}\left(x; \mathcal{Z}_{\mathrm{tr}}\right) + \eta(x)\right] \tag{1}$$

If possible, it may be a function of treatment and testing point x. We want to explore if we can modify current conformal inference method to accommodate continuous treatment effect.

$$\hat{C}(x, T) = \left[\hat{q}_{\alpha_{\mathrm{lo}}}\left(x, T; \mathcal{Z}_{\mathrm{tr}}\right) - \eta(x, T), \hat{q}_{\alpha_{\mathrm{hi}}}\left(x, T; \mathcal{Z}_{\mathrm{tr}}\right) + \eta(x, T)\right] \tag{2}$$

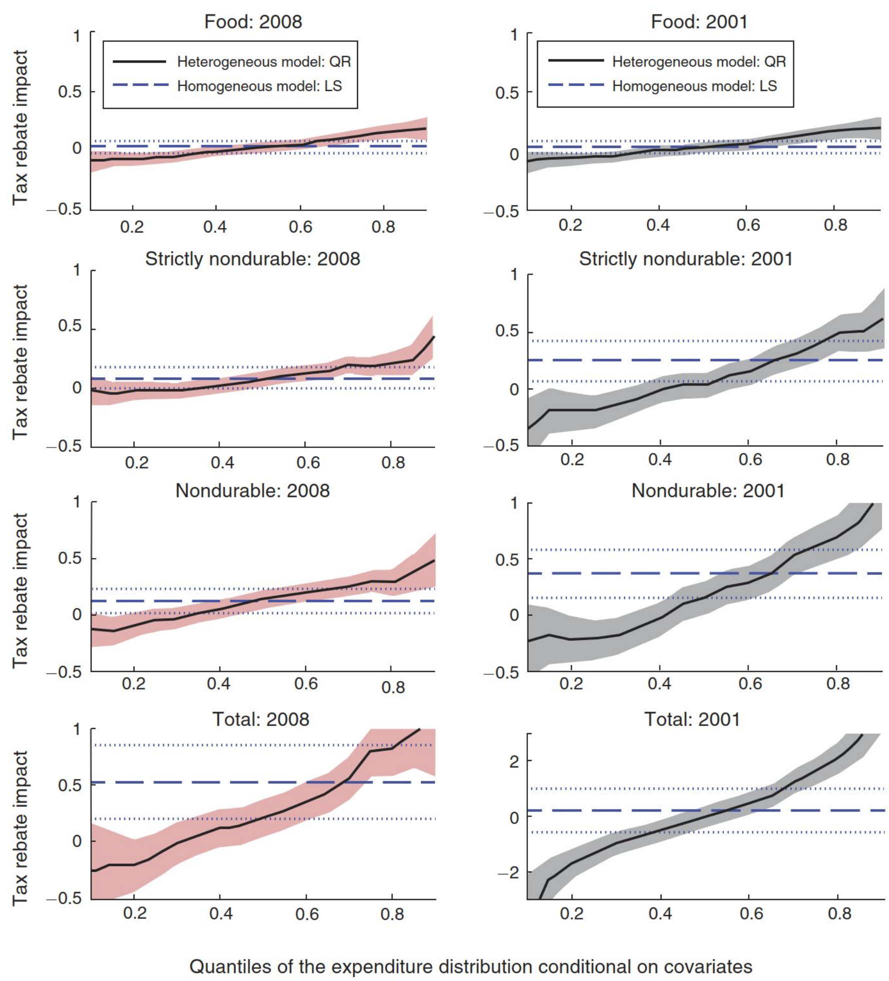

Figure 1: Quantile Estimation $Q_{\Delta C_{it+1} | R_{it+1}, M_s, X_{it}}(\tau)$

2 Conformal Inference When Treatment is Endogenous

Background and Example As stated in Chernozhukov and Hansen [2005], in observational studies, treatment variables are sometimes endogenous, such as education and prices. For binary treatment, there is an example in McCrary [2008] shows endogeneity or manipulation in treatment. The DGP is as follows:

Two periods, 1 and 2; in period 1, worker i get salary R_{1i}, where

$$R_{1i} = \begin{cases} \alpha_i, & full - time, \\ f_i\alpha_i, & part - time. \end{cases} \tag{3}$$

$\alpha_i \sim N(\mu, \sigma^2)$ and $f_i \sim U(0,1)$. In Period 2, worker i earns R_{2i}, where

$$R_{2i} = \alpha_i + \beta_i \mathbb{1}\{R_{1i} < c\}$$

If β is large in mean, then workers might decide to work part time in period 1 purposely, thus treatment is endogenous.

Instrument variables are usually used to deal with endogeneity. Chernozhukov and Hansen [2005] propose a IV model of Quantile Treatment Effect. We want to explore if we can use conformal quantile regression to update these models.

In Chernozhukov and Hansen [2005], Z is an instrument variable which affects treatment D but independent of potential outcome Y_d. Condition on $X = x$, they propose latent outcome $Y_d = q(d, x, U_d)$, where $U_d \sim U(0,1)$. U_d represents unobserved characteristics. Restriction is $P(Y \leq q(D, X, \tau)|X, Z) = \tau$. There main assumptions are as follows:

A1. Potential Outcomes: Conditional on $X = x$, for each d, $Y_d = q\left(d, x, U_d\right)$, where $q(d, x, \tau)$ is strictly increasing in τ and $U_d \sim U(0,1)$.

A2. Independence: Conditional on $X = x$, $\{U_d\}$ are independent of Z.

A3. Selection: $D \equiv \delta(Z, X, V)$ for some unknown function δ and random vector V.

A4. Rank Invariance or Rank Similarity: Conditional on $X = x$, $Z = z$ (a) $\{U_d\}$ are equal to each other; or, more generally, (b) $\{U_d\}$ are identically distributed, conditional on V.

A5. Observed Variables: Observed variables consist of $Y \equiv q\left(D, X, U_D\right)$, D, X, and Z

In this setting, their main goal is still to estimate quantile treatment effects $q(1, x, \tau) - q(0, x, \tau)$, or $\int_0^1 (\frac{\partial}{\partial d} q(d, x, \tau)) d\tau$. We want to follow Lei and Candès [2021] to develop prediction intervals for models with IV.

3 Conformalized Quantile Autoregression

Another important application of quantile regression is to estimation long-run effect of policy i.e. time series. According to Koenker and Xiao [2006], we define following quantile vector autoregression (QVAR) process. Let $\{U_t\}$ be a sequence of iid standard uniform random variables, and consider the p th-order autoregressive process,

$$y_t = \theta_0\left(U_t\right) + \theta_1\left(U_t\right) y_{t-1} + \cdots + \theta_p\left(U_t\right) y_{t-p} \tag{4}$$

where the θ_j 's are unknown functions $[0,1] \rightarrow \mathbb{R}$ that we want to estimate. Provided that the right side of equation (4) is monotone increasing in U_t, it follows that the τ th conditional quantile function of y_t can be written as

$$
\begin{aligned}
& Q_{y_t}\left(\tau \mid y_{t-1}, \ldots, y_{t-p}\right) \\
& =\theta_0(\tau)+\theta_1(\tau) y_{t-1}+\cdots+\theta_p(\tau) y_{t-p}
\end{aligned}
$$

Figure 2 is an example of quantile autoregression. This is an impulse response path of monetary shock. On different percentile, the impulse response varies. Traditional VAR identification strategy relies on strong assumptions of normal distribution of white noise, however, which is not true in *most* cases. We want to explore if we can introduce conformal method into traditional VAR regression to avoid problems because of finite sample and strong assumption of distribution.

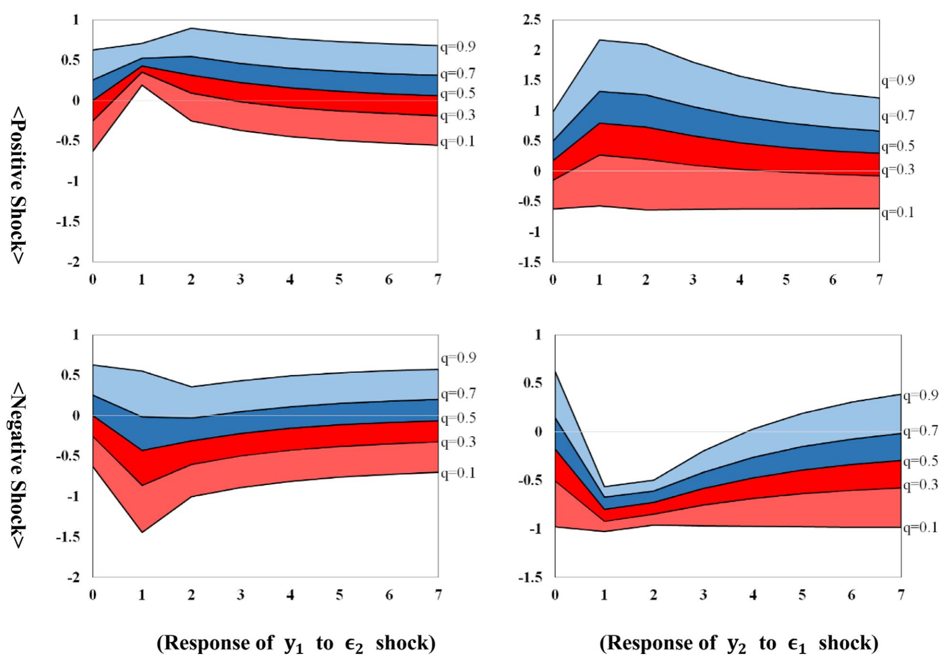

Figure 2: Quantile VAR Estimation

References

Victor Chernozhukov and Christian Hansen. An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261, 2005.

David S. Johnson, Jonathan A. Parker, and Nicholas S. Souleles. Household expenditure and the income tax rebates of 2001. *American Economic Review*, 96(5):1589–1610, December 2006.

Roger Koenker and Zhijie Xiao. Quantile autoregression. *Journal of the American statistical association*, 101(475):980–990, 2006.

Lihua Lei and Emmanuel J Candès. Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2021.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714, 2008.

Kanishka Misra and Paolo Surico. Consumption, income changes, and heterogeneity: Evidence from two fiscal stimulus programs. *American Economic Journal: Macroeconomics*, 6(4):84–106, October 2014.

Jonathan A. Parker, Nicholas S. Souleles, David S. Johnson, and Robert McClelland. Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, 103(6):2530–53, October 2013.