Endogeneity in Modal Regression

Tao Wang

Department of Economics, University of California, Riverside

Introduction

Popular Location/Center Measures and Regression Methods

Given $f(X)$ and $f(Y \mid X)$

- Mean: $\mathbb{E}\{X\}$

  $\longrightarrow$ Mean regression: $\mathbb{E}(Y \mid X)$
- Median/Quantile: Median/Quantile$\{X\}$

  $\longrightarrow$ Median/Quantile regression: $Q_{Y|X}(\tau)$
- **Mode**: Mode$\{X\}$

$\overset{?}{\Longrightarrow}$ How about regression using **mode**–$Mode(Y \mid X)$?

NOT TOO MUCH RESEARCH!

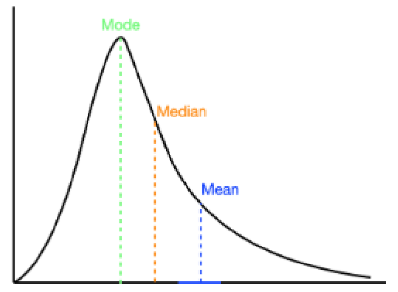
$\overset{?}{\Longrightarrow}$ How about regression with **endogeneity** using **mode**?

No.

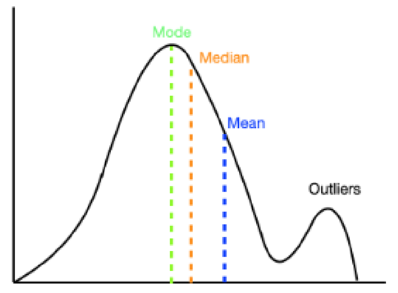
Features of Modal Regression

Modal Regression: $Y = X^T\beta + U$ and $Mode(U \mid X) = 0 \rightarrow$ $Mode(Y \mid X) = X^T\beta$

1. No moment restriction, i.e., Cauchy distribution
2. Better for skewed data, i.e., mode return and Bayesian estimation
3. Applicable to clustered/inhomogeneous data
4. Shorter prediction intervals
5. Suitable for truncated data
6. Robust to outliers and heavy-tailed distributions



(I) ASYMMETRIC



(II) SYMMETRIC WITH OUTLIERS

Definition of Endogeneity

- **Endogeneity** is prevalent in economics and statistics, i.e., simultaneous causality (education or prices), sample selection, and omitted variables.
- *Interpret the endogeneity in modal regression as the nonzero value of the conditional mode of error term given covariates,*

$$Mode(U \mid X) \neq 0.$$

- Given instrumental variables Z such that $X = h(Z, V)$, endogeneity in modal regression implies that

$$V \text{ is stochastically dependent on } U.$$

- Endogeneity renders the modal regression **inconsistent** for estimating the causal (structural) effects of covariates on the **mode** of outcomes.
- Control function approach (Newey et al., 1999; Su and Ullah, 2008).

Parametric Triangular System

- For observations $\{Y_i, X_i, Z_i\}_{i=1}^n$ from random vector (Y, X, Z),

$$\begin{cases} Y_i = X_i\beta + Z_{1,i}^T\gamma + U_i \text{ (structural equation)}, \\ X_i = \alpha + Z_i^T\pi + V_i \text{ (reduced form equation)}, \end{cases}$$

  where $Mode(V_i \mid Z_i) = 0$ (a.s.) and $Mode(U_i \mid X_i, Z_i) \neq 0$ (a.s.).
- For **identification**, no constant is in the structural equation, and the standard rank condition is satisfied ($dim(Z_{2,i}) \geq 1$, $Z_i = (Z_{1,i}^T, Z_{2,i}^T)^T$).

REMARK

Release the strict parametric assumption,

$$\begin{cases} Y_i = g(X_i, Z_{1,i}) + U_i \textit{ (structural equation)}, \\ X_i = h(Z_i) + V_i \textit{ (reduced form equation)}, \end{cases}$$

where $g(\cdot)$ and $h(\cdot)$ are real-valued (non-constant) functions.

- With restriction of a **mode independence** of U_i on Z_i conditional on V_i,

$$\begin{cases} Mode(Y_i \mid X_i, Z_i, V_i) = X_i\beta + Z_{1,i}^T\gamma + Mode(U_i \mid X_i, Z_i, V_i) \\ \qquad = X_i\beta + Z_{1,i}^T\gamma + Mode(U_i \mid \alpha + Z_i^T\pi + V_i, Z_i, V_i) \\ \qquad = X_i\beta + Z_{1,i}^T\gamma + Mode(U_i \mid V_i, Z_i) \\ \qquad = X_i\beta + Z_{1,i}^T\gamma + Mode(U_i \mid V_i), \\ \boxed{Mode(X_i \mid Z_i) = \alpha + Z_i^T\pi.} \end{cases}$$

- Define $Mode(U_i \mid V_i) = m(V_i)$ as a real-valued unknown function.

$$\begin{cases} Y_i = X_i\beta + Z_{1,i}^T\gamma + m(V_i) + \underbrace{U_i - m(V_i)}_{\text{new error term}} \\ \boxed{Mode(Y_i \mid X_i, Z_{1,i}, V_i) = X_i\beta + Z_{1,i}^T\gamma + m(V_i)} \end{cases}$$

$\hookrightarrow$ SEMIPARAMETRIC PARTIALLY LINEAR MODAL REGRESSION

Motivation 1

- Consider a two-period economy with two assets, one risk and one risk-free. Under the **modal maximization decision** and time-separability,

$$\max_{\theta, \theta^f} Mode_t(U(C_t) + \beta U(C_{t+1})) = U(C_t) + \beta U(Mode_t(C_{t+1})),$$

  since $Mode(U(C)) = U(Mode(C))$ (INVARIANCE).
- The budget constraint is $\begin{cases} C_t = W_t - P_t\theta - P_t^f\theta^f, \\ C_{t+1} = X_{t+1}\theta + X_{t+1}^f\theta^f. \end{cases}$
- Define $U(C) = C^{1-\gamma}/(1-\gamma)$. The **Modal Euler Equation** is

$$Mode\left(\frac{C_{t+1}}{C_t} \mid \Omega_t\right) = (\beta X_{t+1}^f)^{1/\gamma},$$

$$\boxed{Mode_t(ln(X_{t+1}^f)) = -\gamma ln(\beta) + Mode\left(ln\left(\frac{C_{t+1}}{C_t}\right) \mid \Omega_t\right).}$$

$\hookrightarrow$ **Modal Regression but with Endogeneity!**

Estimation Procedure

- First Step is the construction of estimated residuals $\{\hat{V}_i\}_{i=1}^n$.

$$Q_n(\alpha, \pi) = \frac{1}{nh}\sum_{i=1}^n \phi\left(\frac{X_i - \alpha - Z_i^T\pi}{h}\right),$$

  where $\phi(\cdot)$ is chosen as a Gaussian kernel.
- **Identification** depends on whether the population moment conditions are satisfied uniquely,

$$\mathbb{E}\left(\frac{Z_i}{h^3}\phi\left(\frac{X_i - \alpha - Z_i^T\pi}{h}\right)\left(X_i - \alpha - Z_i^T\pi\right)|_{\alpha=\alpha_0, \pi=\pi_0}\right) = 0,$$

$$\mathbb{E}\left(\frac{1}{h^3}\phi\left(\frac{X_i - \alpha - Z_i^T\pi}{h}\right)\left(X_i - \alpha - Z_i^T\pi\right)|_{\alpha=\alpha_0, \pi=\pi_0}\right) = 0.$$

LEMMA

If the partial derivative matrix of the above moment condition with respect to α and π is full rank, local identification is achieved.

- Second Step focuses a semiparametric partially linear modal regression.

$$Mode(Y_i \mid X_i, Z_{1,i}, \hat{V}_i) = X_i\beta + Z_{1,i}^T\gamma + m(\hat{V}_i) + o_p(1)$$

- **First Stage** applies local linear technique to approximate $m(\hat{V}_i)$.

$$Q_n(\beta, \gamma, \alpha_1, \alpha_2) = \frac{1}{nh_1h_2}\sum_{i=1}^n \phi\left(\frac{Y_i - X_i\beta - Z_{1,i}^T\gamma - \alpha_1 - \alpha_2(\hat{V}_i - v)}{h_1}\right) K\left(\frac{\hat{V}_i - v}{h_2}\right)$$

- **Second Stage** improves the **convergence rates** of the estimators of the parametric components using all data.

$$Q_n(\beta, \gamma) = \frac{1}{nh_3}\sum_{i=1}^n \phi\left(\frac{Y_i - \tilde{m}(\hat{V}_i) - X_i\beta - Z_{1,i}^T\gamma}{h_3}\right)$$

- **Third Stage** improves the **efficiency** of the nonparametric part.

$$Q_n(\alpha_1, \alpha_2) = \frac{1}{nh_4h_5}\sum_{i=1}^n \phi\left(\frac{Y_i - X_i\hat{\beta} - Z_i^T\hat{\gamma} - \alpha_1 - \alpha_2(\hat{V}_i - v)}{h_4}\right) K\left(\frac{\hat{V}_i - v}{h_5}\right)$$

Motivation 2

Dependent variable is log wages; endogenous variable is years of schooling (ed76); instrumental variable is living near a four-year college (Card, 2001).

$$\begin{cases} ln(Wage) = \alpha * ed76 + W^T\theta + U, \ Mode(U \mid ed76) \neq 0 \\ ed76 = \beta * Z + W^T\theta + V, \ Mode(U \mid V) \neq 0. \end{cases}$$

TABLE: Estimates of Return to Schooling

Variables	Two-Step Modal	Naive Linear Modal	Mean-2SLS
ed76	0.1331*** (0.0010)	0.0772*** (0.0005)	0.1315** (0.0548)
	Quantile (0.3)	Quantile (0.5)	Quantile (0.7)
ed76	0.1652*** (0.0561)	0.1351*** (0.0790)	0.0945** (0.0391)

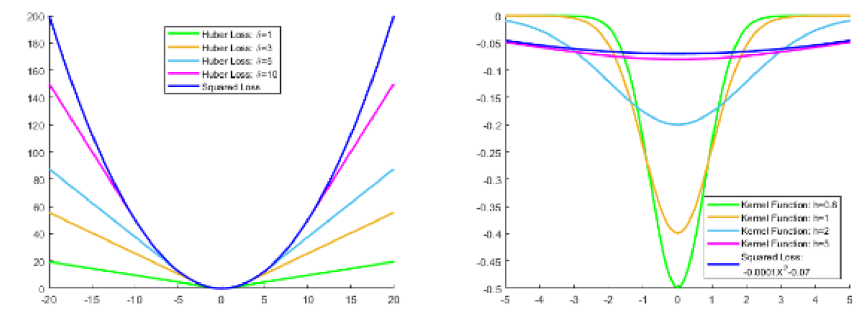
Note: The standard error is calculated from Bootstrap based on mode value.

Modal-Based Control Function

- The fundamental principle: with symmetric data, *modal regression line is identical to mean regression line.*
- (Local linear) mean estimator is sensitive to **outliers** and does not perform well when the data have **heavy-tailed** distributions.
- The existing robust techniques, including robust Huber's estimation, can achieve robustness by *sacrificing some of the efficiency.*
- With the focus on modal regression for symmetric data, $Mode(V_i \mid Z_i) = \mathbb{E}(V_i \mid Z_i) = 0$, $Mode(U_i \mid X_i, Z_i) = \mathbb{E}(U_i \mid X_i, Z_i) \neq 0$,

$$\begin{cases} Mode(Y_i \mid X_i, Z_i, V_i) = \mathbb{E}(Y_i \mid X_i, Z_i, V_i) = X_i\beta + Z_{1,i}^T\gamma + Mode(U_i \mid V_i), \\ Mode(X_i \mid Z_i) = \mathbb{E}(X_i \mid Z_i) = \alpha + Z_i^T\pi. \end{cases}$$

- Utilize the **same** kernel-based objective functions but with **constant** bandwidths associated with error terms.
- When there are outliers or the error distribution has heavy tails, the proposed modal-based estimation performs **better**.
- **As asymptotically efficient as** the mean estimation when there are no outliers and the error is normally distributed.



With a Gaussian kernel, $1 - exp(-\varepsilon_i^2/2h^2) \approx \varepsilon_i^2/(2h^2)$

IV Selection in First Step

- For the first step, it may have a **large** set of instrumental variables to be used in practice and face the dimensionality curse of many instruments.
- Provide a *modal adaptive lasso* method to cull the **weak** instrumental variables to get more robust results.
- Penalized modal regression with an adaptive lasso is

$$Q(\theta) = \frac{1}{nh}\sum_{i=1}^n \phi\left(\frac{X_i - Z_i^{*T}\theta}{h}\right) + \lambda_n\|\hat{w} \circ \theta\|,$$

  where λ_n is a nonnegative regularization parameter, $\hat{w} \circ \theta = \sum_{j=1}^{d_Z+1} \hat{w}_j|\theta_j|$, and $\hat{w}_j = 1/|\hat{\theta}_j|^\gamma$ with $0 < \gamma < 2$. $|\theta_j|/|\hat{\theta}_j|$ converges to $I(\theta_j \neq 0)$ in probability.
- Select λ_n by a **Consistent** BIC-type procedure

$$\lambda_{n,opt} = \arg\min_{\lambda_n} BIC(\lambda_n) = -\frac{1}{nh}\sum_{i=1}^n \phi\left(\frac{X_i - Z_i^{*T}\hat{\theta}^P}{h}\right) + \frac{\log(nh^3)}{nh^3} df_{\lambda_n},$$

  where df_{λ_n} is the degrees of freedom of the fitted model.

Contact

tao.wang@email.ucr.edu.