Endogoneity and All That

Roger Koenker

CEMMAP and University of Illinois, Urbana-Champaign

LSE: 17 May 2011

Is there IV for QR?

- Amemiya (1982) and Powell (1983) consider analogues of 2SLS for median regression models
- Chen and Portnoy (1986) consider extensions to quantile regression
- Abadie, Angrist and Imbens (2002) consider models with binary endogonous treatment
- Chernozhukov and Hansen (2003) propose "inverse" quantile regression
- Chesher (2003) considers triangular models with continuous endogonous variables.

Chernozhukov and Hansen QRIV

Motivation: Yet another way to view two stage least squares.

Model: $y = X\beta + Z\alpha + u, \quad W \perp\!\!\!\perp u$

Estimator:

$$\hat{\alpha} = \text{argmin}_{\alpha} \|\hat{\gamma}(\alpha)\|^2_{A = W^\top M_X W}$$

$$\hat{\gamma}(\alpha) = \text{argmin}_{\gamma} \|y - X\beta - Z\alpha - W\gamma\|^2$$

Thm $\quad \hat{\alpha} = (Z^\top P_{M_X W} Z)^{-1} Z^\top P_{M_X W} y$, the 2SLS estimator.

Heuristic: $\hat{\alpha}$ is chosen to make $\|\hat{\gamma}(\alpha)\|$ as small as possible to satisfy (approximately) the exclusion restriction/assumption.

Generalization: The quantile regression version simply replaces $\| \cdot \|^2$ in the definition of $\hat{\gamma}$ by the corresponding QR norm.

A Linear Location Shift Recursive Model

$$\begin{aligned} Y &= S\alpha_1 + x^\top \alpha_2 + \epsilon + \lambda\nu & (1) \\ S &= z\beta_1 + x^\top \beta_2 + \nu & (2) \end{aligned}$$

A Linear Location Shift Recursive Model

$$Y = S\alpha_1 + x^\top \alpha_2 + \epsilon + \lambda\nu \quad (1)$$
$$S = z\beta_1 + x^\top \beta_2 + \nu \quad (2)$$

Suppose: $\epsilon \perp\!\!\!\perp \nu$ and $(\epsilon, \nu) \perp\!\!\!\perp (z, x)$. Substituting for ν from (2) into (1),

$$Q_Y(\tau_1 | S, x, z) = S(\alpha_1 + \lambda) + x^\top(\alpha_2 - \lambda\beta_2) + z(-\lambda\beta_1) + F_\epsilon^{-1}(\tau_1)$$
$$Q_S(\tau_2 | z, x) = z\beta_1 + x^\top \beta_2 + F_\nu^{-1}(\tau_2)$$

$$\begin{aligned}\pi_1(\tau_1, \tau_2) &= \nabla_{S_i} Q_{Y_i}|_{S_i = Q_{S_i}} + \frac{\nabla_{z_i} Q_{Y_i}|_{S_i = Q_{S_i}}}{\nabla_{z_i} Q_{S_i}} \\ &= (\alpha_1 + \lambda) + (-\lambda\beta_1)/\beta_1 \\ &= \alpha_1\end{aligned}$$

A Linear Location-Scale Shift Model

$$\begin{aligned} Y &= S\alpha_1 + x^\top \alpha_2 + S(\epsilon + \lambda\nu) \\ S &= z\beta_1 + x^\top \beta_2 + \nu \\ \pi_1(\tau_1, \tau_2) &= \alpha_1 + F_\epsilon^{-1}(\tau_1) + \lambda F_\nu^{-1}(\tau_2) \end{aligned}$$

A Linear Location-Scale Shift Model

$$\begin{aligned} Y &= S\alpha_1 + x^\top \alpha_2 + S(\epsilon + \lambda \nu) \\ S &= z\beta_1 + x^\top \beta_2 + \nu \\ \pi_1(\tau_1, \tau_2) &= \alpha_1 + F_\epsilon^{-1}(\tau_1) + \lambda F_\nu^{-1}(\tau_2) \end{aligned}$$

$$\begin{aligned} Q_Y(\tau_1 | S, x, z) &= S\theta_1(\tau_1) + x^\top \theta_2 + S^2\theta_3 + Sz\theta_4 + Sx^\top \theta_5 \\ Q_S(\tau_2 | z, x) &= z\beta_1 + x^\top \beta_2 + F_\nu^{-1}(\tau_2) \end{aligned}$$

$$\hat{\pi}_1(\tau_1, \tau_2) = \sum_{i=1}^{n} w_i \left\{ \hat{\theta}_1(\tau_1) + 2\hat{Q}_{S_i}\hat{\theta}_3(\tau_1) + z_i\hat{\theta}_4(\tau_1) + x_i^\top \hat{\theta}_5(\tau_1) + \frac{\hat{Q}_{S_i}\hat{\theta}_4(\tau_1)}{\hat{\beta}_1(\tau_2)} \right\}$$

a weighted average derivative estimator with $\hat{Q}_{S_i} = \hat{Q}_S(\tau_2 | z_i, x_i)$.

The General Recursive Model

$$
\begin{aligned}
Y &= \varphi_1(S, x, \epsilon, \nu;\ \alpha) \\
S &= \varphi_2(z, x, \nu;\ \beta)
\end{aligned}
$$

Suppose: $\epsilon \perp\!\!\!\perp \nu$ and $(\epsilon, \nu) \perp\!\!\!\perp (z, x)$. Solving for ν and substituting we have the conditional quantile functions,

$$
\begin{aligned}
Q_Y(\tau_1 | S, x, z) &= h_1(S, x, z, \theta(\tau_1)) \\
Q_S(\tau_2 | z, x) &= h_2(z, x, \beta(\tau_2))
\end{aligned}
$$

The General Recursive Model

$$
\begin{aligned}
Y &= \varphi_1(S, x, \epsilon, \nu;\ \alpha) \\
S &= \varphi_2(z, x, \nu;\ \beta)
\end{aligned}
$$

Suppose: $\epsilon \perp\!\!\!\perp \nu$ and $(\epsilon, \nu) \perp\!\!\!\perp (z, x)$. Solving for ν and substituting we have the conditional quantile functions,

$$
\begin{aligned}
Q_Y(\tau_1 | S, x, z) &= h_1(S, x, z, \theta(\tau_1)) \\
Q_S(\tau_2 | z, x) &= h_2(z, x, \beta(\tau_2))
\end{aligned}
$$

Extensions to more than two endogonous variables are "straightforward."

The (Chesher) Weighted Average Derivative Estimator

$$\hat{\theta}(\tau_1) = \text{argmin}_\theta \sum_{i=1}^{n} \rho_{\tau_1}(Y_i - h_1(S, x, z, \theta(\tau_1)))$$

$$\hat{\beta}(\tau_2) = \text{argmin}_\beta \sum_{i=1}^{n} \rho_{\tau_2}(S_i - h_2(z, x, \beta(\tau_2)))$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$, giving structural estimators:

The (Chesher) Weighted Average Derivative Estimator

$$\hat{\theta}(\tau_1) = \text{argmin}_\theta \sum_{i=1}^{n} \rho_{\tau_1}(Y_i - h_1(S, x, z, \theta(\tau_1)))$$

$$\hat{\beta}(\tau_2) = \text{argmin}_\beta \sum_{i=1}^{n} \rho_{\tau_2}(S_i - h_2(z, x, \beta(\tau_2)))$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$, giving structural estimators:

$$\hat{\pi}_1(\tau_1, \tau_2) = \sum_{i=1}^{n} w_i \left\{ \nabla_S \hat{h}_{1i}|_{S_i=\hat{h}_{2i}} + \frac{\nabla_z \hat{h}_{1i}|_{S_i=\hat{h}_{2i}}}{\nabla_z \hat{h}_{2i}} \right\},$$

$$\hat{\pi}_2(\tau_1, \tau_2) = \sum_{i=1}^{n} w_i \left\{ \nabla_x \hat{h}_{1i}|_{S_i=\hat{h}_{2i}} - \frac{\nabla_z \hat{h}_{1i}|_{S_i=\hat{h}_{2i}}}{\nabla_z \hat{h}_{2i}} \nabla_x \hat{h}_{2i} \right\},$$

2SLS as a Control Variate Estimator

$$\begin{aligned} Y &= S\alpha_1 + X_1\alpha_2 + u \equiv Z\alpha + u \\ S &= X\beta + V, \quad \text{where } X = [X_1 \vdots X_2] \end{aligned}$$

Set $\hat{V} = S - \hat{S} \equiv M_X Y_1$, and consider the least squares estimator of the model,

$$Y = Z\alpha + \hat{V}\gamma + w$$

2SLS as a Control Variate Estimator

$$\begin{aligned} Y &= S\alpha_1 + X_1\alpha_2 + u \equiv Z\alpha + u \\ S &= X\beta + V, \quad \text{where } X = [X_1 \vdots X_2] \end{aligned}$$

Set $\hat{V} = S - \hat{S} \equiv M_X Y_1$, and consider the least squares estimator of the model,

$$Y = Z\alpha + \hat{V}\gamma + w$$

Claim: $\hat{\alpha}_{CV} \equiv (Z^\top M_{\hat{V}} Z)^{-1} Z^\top M_{\hat{V}} Y = (Z^\top P_X Z)^{-1} Z^\top P_X Y \equiv \hat{\alpha}_{2SLS}$.

Proof of Control Variate Equivalence

$$M_{\hat{V}} = M_{M_X S} = I - M_X S(S^\top M_X S)^{-1} S^\top M_X$$

$$\begin{aligned} S^\top M_{\hat{V}} &= S^\top - S^\top M_X = S^\top P_X \\ X_1^\top M_{\hat{V}} &= X_1^\top - X_1^\top M_X = X_1^\top = X_1^\top P_X \end{aligned}$$

Reward for information leading to a reference prior to Dhrymes (1970). Recent work on the control variate approach by Blundell, Powell, Smith, Newey and others.

Quantile Regression Control Variate Estimation I

Location scale shift model:

$$\begin{aligned} Y &= S(\alpha_1 + \epsilon + \lambda\nu) + x^\top \alpha_2 \\ S &= z\beta_1 + x^\top \beta_2 + \nu. \end{aligned}$$

Using $\hat{\nu}(\tau_2) = S - \hat{Q}_S(\tau_2|z, x)$ as a control variate,

$$\begin{aligned} Y &= w^\top \alpha(\tau_1, \tau_2) + \lambda S(\hat{Q}_S - Q_S) + S(\epsilon - F_\epsilon^{-1}(\tau_1)), \\ \text{where} \quad & w^\top = (S, x^\top, S\hat{\nu}(\tau_2)) \\ & \alpha(\tau_1, \tau_2) = (\alpha_1(\tau_1, \tau_2), \alpha_2, \lambda)^\top \\ & \alpha_1(\tau_1, \tau_2) = \alpha_1 + F_\epsilon^{-1}(\tau_1) + \lambda F_\nu^{-1}(\tau_2). \end{aligned}$$

$$\hat{\alpha}(\tau_1, \tau_2) = \text{argmin}_a \sum_{i=1}^{n} \rho_{\tau_1}(Y_i - w_i^\top a).$$

Quantile Regression Control Variate Estimation II

$$\begin{aligned} Y &= \varphi_1(S, x, \epsilon, \nu;\ \alpha) \\ S &= \varphi_2(z, x, \nu;\ \beta) \end{aligned}$$

Regarding $\nu(\tau_2) = \nu - F_\nu^{-1}(\tau_2)$ as a control variate, we have

$$\begin{aligned} Q_Y(\tau_1 | S, x, \nu(\tau_2)) &= g_1(S, x, \nu(\tau_2), \alpha(\tau_1, \tau_2)) \\ Q_S(\tau_2 | z, x) &= g_2(z, x, \beta(\tau_2)) \end{aligned}$$

$$\hat{\nu}(\tau_2) = \varphi_2^{-1}(S, z, x, \hat{\beta}) - \varphi_2^{-1}(\hat{Q}_s, z, x, \hat{\beta})$$

$$\hat{\alpha}(\tau_1, \tau_2) = \text{argmin}_a \sum_{i=1}^{n} \rho_{\tau_1}(Y_i - g_1(S, x, \hat{\nu}(\tau_2), a)).$$

Asymptopia

Theorem: Under regularity conditions, the weighted average derivative and control variate estimators of the Chesher structural effect have an asymptotic linear (Bahadur) representation, and after efficient reweighting of both estimators, the control variate estimator has smaller covariance matrix than the weighted average derivative estimator.

Asymptopia

Theorem: Under regularity conditions, the weighted average derivative and control variate estimators of the Chesher structural effect have an asymptotic linear (Bahadur) representation, and after efficient reweighting of both estimators, the control variate estimator has smaller covariance matrix than the weighted average derivative estimator.

Remark: The control variate estimator imposes more stringent restrictions on the estimation of the hybrid structural equation and should thus be expected to perform better when the specification is correct. The advantages of the control variate approach are magnified in situations of overidentification.

Asymptotics for WAD

Theorem

The $\hat{\pi}_n(\tau_1, \tau_2)$ *has the asymptotic linear (Bahadur) representation,*

$$\sqrt{n}(\hat{\pi}_n(\tau_1,\tau_2) - \pi(\tau_1,\tau_2)) = W_1 \bar{J}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_{i1} \dot{h}_{i1} \psi_{\tau_1}(Y_{i1} - \xi_{i1})$$

$$+ W_2 \bar{J}_2^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_{i2} \dot{h}_{i2} \psi_{\tau_2}(Y_{i2} - \xi_{i2})$$

$$\Longrightarrow \mathcal{N}(0,\ \omega_{11} W_1 \bar{J}_1^{-1} J_1 \bar{J}_1^{-1} W_1^\top + \omega_{22} W_2 \bar{J}_2^{-1} J_2 \bar{J}_2^{-1} W_2^\top)$$

$$J_j = \lim_{n\to\infty} \frac{1}{n} \sum \sigma_{ij}^2 \dot{h}_{ij} \dot{h}_{ij}^\top,\ \ \bar{J}_j = \lim_{n\to\infty} \frac{1}{n} \sum \sigma_{ij} f_{ij}(\xi_{ij}) \dot{h}_{ij} \dot{h}_{ij}^\top,$$

$$W_1 = \nabla_\theta \pi(\tau_1,\tau_2),\ \ W_2 = \nabla_\beta \pi(\tau_1,\tau_2),$$

$$\dot{h}_{i1} = \nabla_\theta h_{i1},\ \ \dot{h}_{i2} = \nabla_\beta h_{i2},\ \ \omega_{jj} = \tau_j(1-\tau_j).$$

Asymptotics for CV

Theorem

The $\hat{\alpha}_n(\tau_1,\tau_2)$ *has the Bahadur representation,*

$$\sqrt{n}(\hat{\alpha}_n(\tau_1,\tau_2)-\alpha(\tau_1,\tau_2)) = \bar{D}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_{i1}\dot{g}_{i1}\psi_{\tau_1}(Y_{i1}-\xi_{i1})$$

$$+\bar{D}_1^{-1}\bar{D}_{12}\bar{D}_2^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_{i2}\dot{g}_{i2}\psi_{\tau_2}(Y_{i2}-\xi_{i2})$$

$$\Longrightarrow \mathcal{N}(0,\ \omega_{11}\bar{D}_1^{-1}D_1\bar{D}_1^{-1}+\omega_{22}\bar{D}_1^{-1}\bar{D}_{12}\bar{D}_2^{-1}D_2\bar{D}_2^{-1}\bar{D}_{12}^{\top}\bar{D}_1^{-1})$$

$$D_j=\lim_{n\to\infty}n^{-1}\sum\sigma_{ij}^2\dot{g}_{ij}\dot{g}_{ij}^{\top},\ \ \bar{D}_j=\lim_{n\to\infty}n^{-1}\sum\sigma_{ij}f_{ij}(\xi_{ij})\dot{g}_{ij}\dot{g}_{ij}^{\top},$$

$$\bar{D}_{12}=\lim_{n\to\infty}n^{-1}\sum\sigma_{i1}f_{i1}\eta_i\dot{g}_{i1}\dot{g}_{i2}^{\top},$$

$$\dot{g}_{i1}=\nabla_{\alpha}g_{i1},\ \ \dot{g}_{i2}=\nabla_{\beta}g_{i2},\ \ \eta_i=(\partial g_{1i}/\partial\nu_{i2}(\tau_2))(\nabla_{\nu_{i2}}\varphi_{i2})^{-1}.$$

ARE of WAD and CV

- Efficient weights: $\sigma_{ij} = f_{ij}(\xi_{ij})$

$$\sqrt{n}(\hat{\pi}_n(\tau_1, \tau_2) - \pi(\tau_1, \tau_2)) \Rightarrow \mathcal{N}(0, \omega_{11} W_1 J_1^{-1} W_1^\top + \omega_{22} W_2 J_2^{-1} W_2^\top)$$

$$\sqrt{n}(\hat{\alpha}_n(\tau_1, \tau_2) - \alpha(\tau_1, \tau_2)) \Rightarrow \mathcal{N}(0, \omega_{11} D_1^{-1} + \omega_{22} D_1^{-1} D_{12} D_2^{-1} D_{12}^\top D_1^{-1}).$$

The mapping: $\tilde{\pi}_n = L\hat{\alpha}_n, \;\; L\alpha = \pi.$

$$\begin{aligned} W_1 J_1^{-1} W_1^\top &\geqslant LD_1^{-1} L^\top \\ W_2 J_2^{-1} W_2^\top &\geqslant LD_1^{-1} D_{12} D_2^{-1} D_{12}^\top D_1^{-1} L^\top. \end{aligned}$$

Theorem

Under efficient reweighting of both estimators,

$$Avar(\sqrt{n}\tilde{\pi}_n) \leqslant Avar(\sqrt{n}\hat{\pi}_n).$$

Conclusions

- Triangular structural models facilitate causal analysis via recursive conditioning, directed acyclic graph representation.
- Recursive conditional quantile models yield interpretable heterogeneous structural effects.
- Control variate methods offer computationally and statistically efficient strategies for estimating heterogeneous structural effects.
- Weighted average derivative methods offer a less restrictive strategy for estimation that offers potential for model diagnostics and testing.