Quantile Regression Methods For Recursive Structural Equation Models

Lingjie Ma and Roger Koenker
University of Illinois at Urbana-Champaign

La Roche-en-Ardennes: September, 2005

Endogoneity in Quantile Regression Models?

- 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
- Chesher (2002, 2003) considers triangular models with continuous endogonous variables.

An Application

How do changes in class size affect the academic performance of Dutch primary school students?

- Do small classes improve performance of all students?
- By the same amount?
- Irrespective of initial class size?
- For language and math equally?
- Are there interactions with other covariates?
- Should class size be treated as endogonous?

PRIMA Data

- Dutch PRIMA school survey: 1994-1995

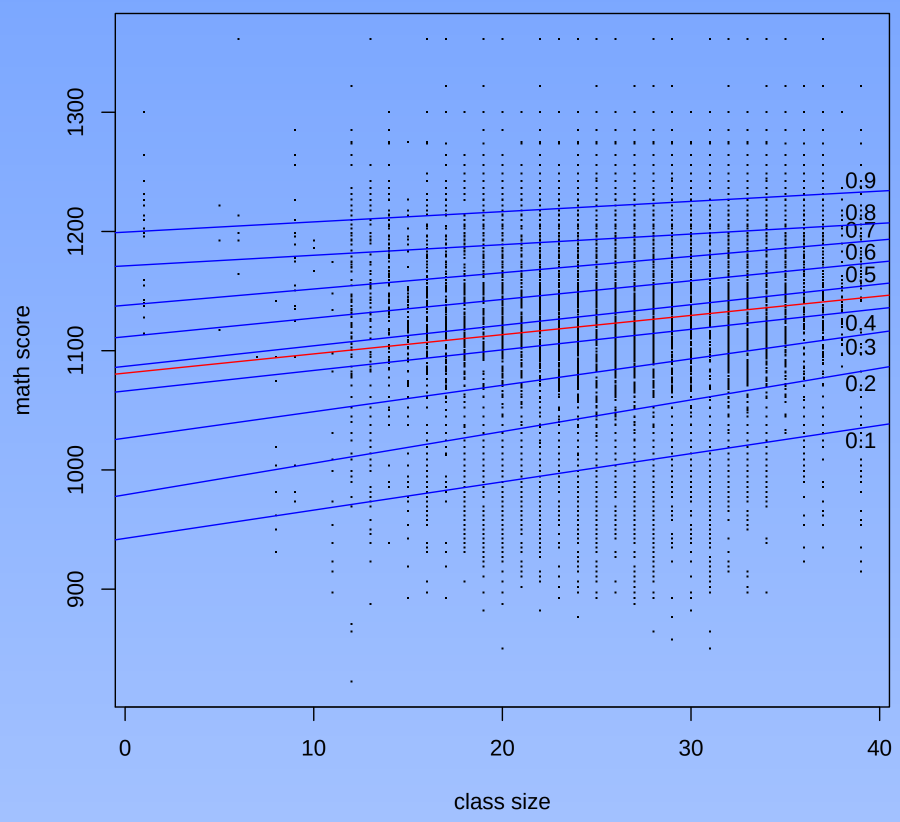
- Academic performance measured by:
  - ⋆ language score
  - ⋆ math score

- Covariates:
  - ⋆ Pupils: IQ, gender, SES, peer effects, risk
  - ⋆ Class: class size, teachers' experience
  - ⋆ School: denomination (public/parochial)

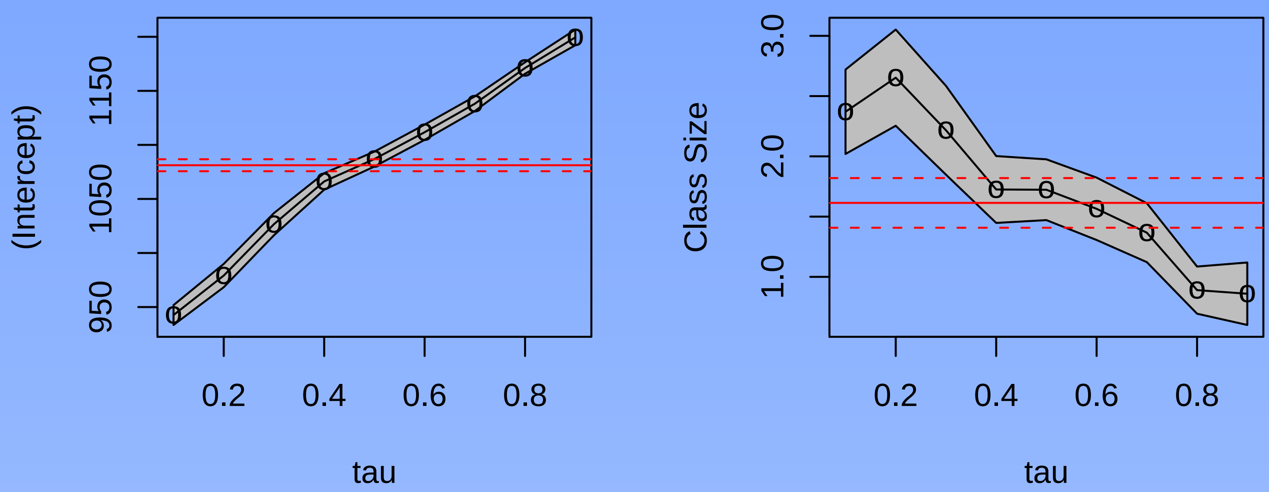
	Min	Max	Mean	Std. Dev.
Language Score	841.80	1261.20	1073.26	51.56
Math Score	822.70	1361.30	1123.49	83.94
Pupil's Gender (Female=1)	0	1	0.50	0.50
IQ	4.00	37.00	25.53	4.95
Socio-Economic Status	0	1	0.53	0.50
Risk	1.00	5.00	2.20	0.87
Peers (Language)	935.65	1179.10	1073.19	40.99
Peers (Math)	852.67	1271.16	1123.44	69.70
Class Size*	5	39	23.81	6.46
Teacher's Experience *	1	40	19.05	8.06
School Denomination **	0	1	0.72	0.44
Weighted Enrollment **	23	684	250.35	120.42

Table 1: PRIMA Survey Summary Statistics: There are 12,203 observations grades 4, 6, and 8 combined.

Does Class Size Matter for Math?



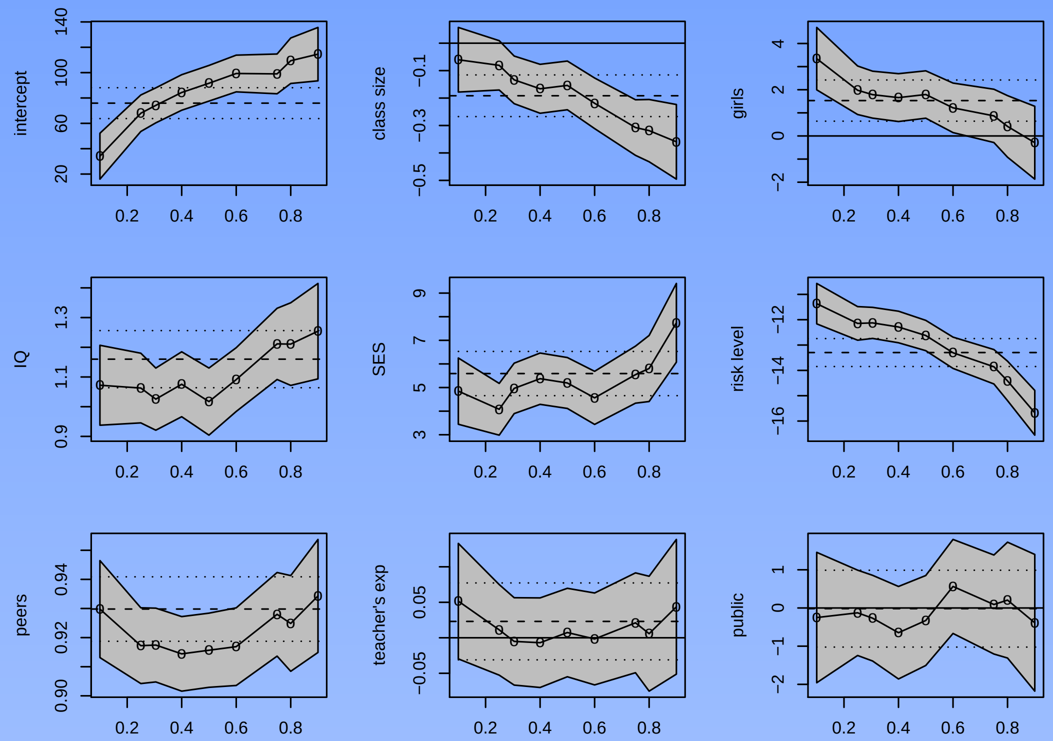
Quantile Regression Coefficient Plots



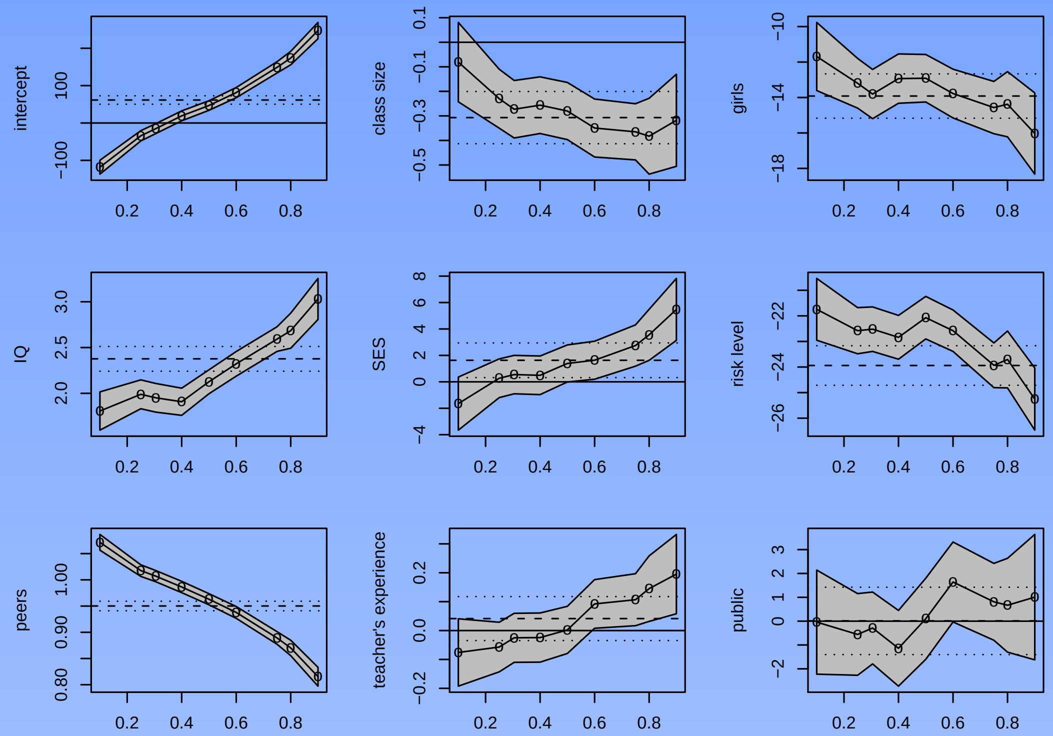
Red lines indicate least squares fit and confidence interval.

Solid line indicates the quantile regression point estimates with gray 90 percent confidence band.

Language Performance: Covariate Effects



Mathematics Performance: Covariate Effects



A Linear Location Shift Recursive Model

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

$$S = z\beta_1 + x^\top \beta_2 + \nu \quad (2)$$

A Linear Location Shift Recursive Model

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
\hat{\theta}(\tau_1) &= \text{argmin}_\theta \sum_{i=1}^{n} \rho_{\tau_1}(Y_i - h_1(S, x, z, \theta(\tau_1))) \\
\hat{\beta}(\tau_2) &= \text{argmin}_\beta \sum_{i=1}^{n} \rho_{\tau_2}(S_i - h_2(z, x, \beta(\tau_2)))
\end{aligned}
$$

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:

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

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

2SLS as a Control Variate Estimator

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

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

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

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.

Monte-Carlo: The Course

Monte-Carlo: The Course

We consider a simple location-scale shift model:

$$\begin{aligned} Y_1 &= \alpha_1 + \alpha_2 x + (\alpha_3 + \delta(\lambda\nu + \epsilon))Y_2 \\ Y_2 &= \beta_1 + \beta_2 x + \beta_3 z + \nu \end{aligned}$$

where x, z, ν_1 and ν_2 are generated as the following:

$$x \sim t_3, \;\; z \sim N(15, 2^2), \;\; \epsilon \sim N(0,1), \;\; \nu \sim N(0, 0.5^2).$$

Parameters: $(\alpha_1, \alpha_2, \alpha_3, \delta, \lambda) = (3, 4, 4, 5, 3)$, The structural quantile treatment effect of Y_2 on Y_1 is

$$\pi(\tau_1, \tau_2) = 4 + 15F_\nu^{-1}(\tau_2) + 5F_\epsilon^{-1}(\tau_1).$$

For the sake of simplicity, we consider only $\tau_1 = \tau_2 = \tau$.

Monte-Carlo: The Cars

- The serious contenders:
  - ⋆ WADQR Weighted Average Derivative Quantile Regression Estimator
  - ⋆ CVQR Control Variate Quantile Regression Estimator

- The also rans:
  - ⋆ 2SQRQ – 2SQR using τ_2 quantile regression in stage one
  - ⋆ 2SQRA – 2SQR using median regression in stage one
  - ⋆ 2SQRS – 2SQR using least squares in stage one
  - ⋆ QR – naive QR

Monte-Carlo: The Results

	Coefficient	Bias	Std. Error	RMSE
$\tau_1 = \tau_2 = \mathbf{0.1}$				
Theoretical Value	-12.019	0.000	0.000	0.000
CVQR	-10.799	1.221	11.715	11.778
WADQR	-10.748	1.271	12.057	12.124
2SQRQ	-7.191	4.829	11.505	12.478
2SQRA	-7.149	4.871	11.473	12.464
2SQRS	-7.152	4.867	11.473	12.463
QR	-2.788	9.231	11.820	14.997
$\tau_1 = \tau_2 = \mathbf{0.3}$				
Theoretical Value	-2.555	0.000	0.000	0.000
CVQR	-1.969	0.586	8.905	8.925
WADQR	-1.876	0.679	9.280	9.305
2SQRQ	-0.345	2.210	9.225	9.486
2SQRA	-0.337	2.218	9.229	9.492
2SQRS	-0.330	2.225	9.226	9.490
QR	4.031	6.586	9.086	11.221
$\tau_1 = \tau_2 = \mathbf{0.5}$				
Theoretical Value	4.000	0.000	0.000	0.000
CVQR	3.715	-0.285	8.656	8.661
WADQR	3.722	-0.278	8.934	8.939
2SQRQ	3.847	-0.153	8.488	8.490
2SQRA	3.847	-0.153	8.488	8.490
2SQRS	3.855	-0.145	8.490	8.492
QR	8.006	4.006	8.313	9.228

Monte-Carlo: The Results (continued)

	Coefficient	Bias	Std. Error	RMSE
$\tau_1 = \tau_2 = \mathbf{0.7}$				
Theoretical Value	10.555	0.000	0.000	0.000
CVQR	9.945	-0.610	8.953	8.974
WADQR	9.968	-0.587	9.506	9.524
2SQRQ	8.417	-2.138	8.895	9.148
2SQRA	8.425	-2.130	8.896	9.148
2SQRS	8.425	-2.130	8.900	9.152
QR	12.626	2.071	8.694	8.937
$\tau_1 = \tau_2 = \mathbf{0.9}$				
Theoretical Value	20.019	0.000	0.000	0.000
CVQR	19.507	-0.513	11.166	11.177
WADQR	19.367	-0.653	12.390	12.407
2SQRQ	14.750	-5.270	11.617	12.756
2SQRA	14.796	-5.223	11.665	12.781
2SQRS	14.787	-5.232	11.656	12.776
QR	19.191	-0.828	11.385	11.415

Meanwhile back in Primary School...

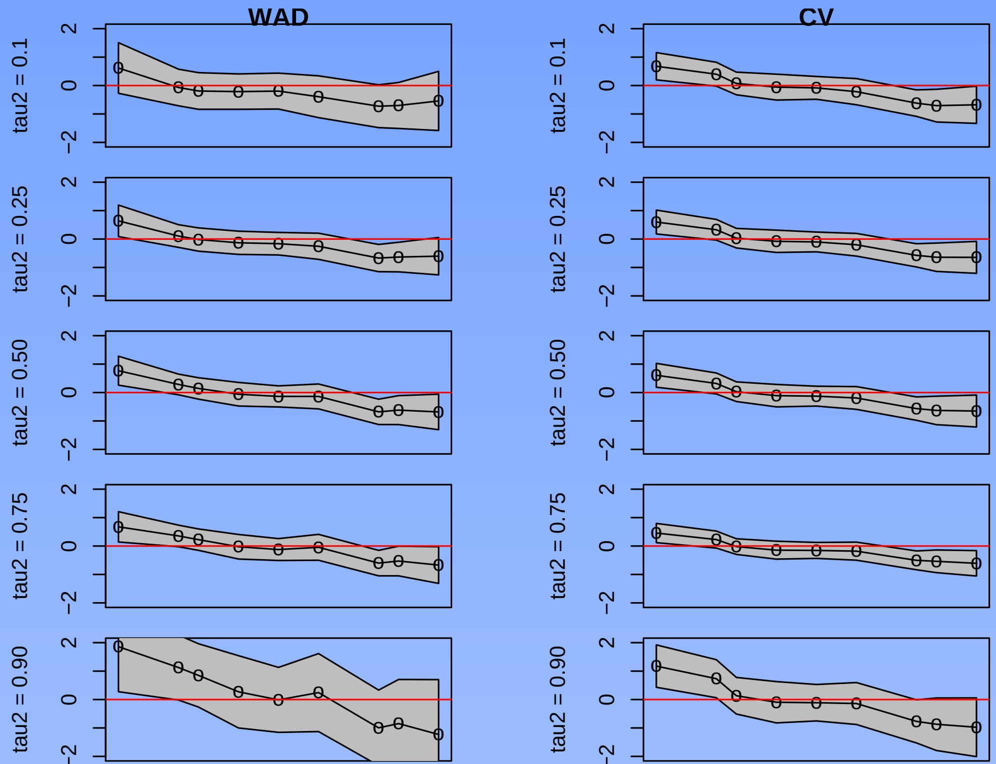
We now reconsider our model of primary school academic performance treating class size as endogonous. Following Levin (2001), we use as our instrumental variable, the Dutch Ministry of Education's "weighted school enrollment",

$$Z_i = 1.03 \max\{(\sum_{j=1}^{n_i} s_{ij} - .09n_i), n_i\},$$

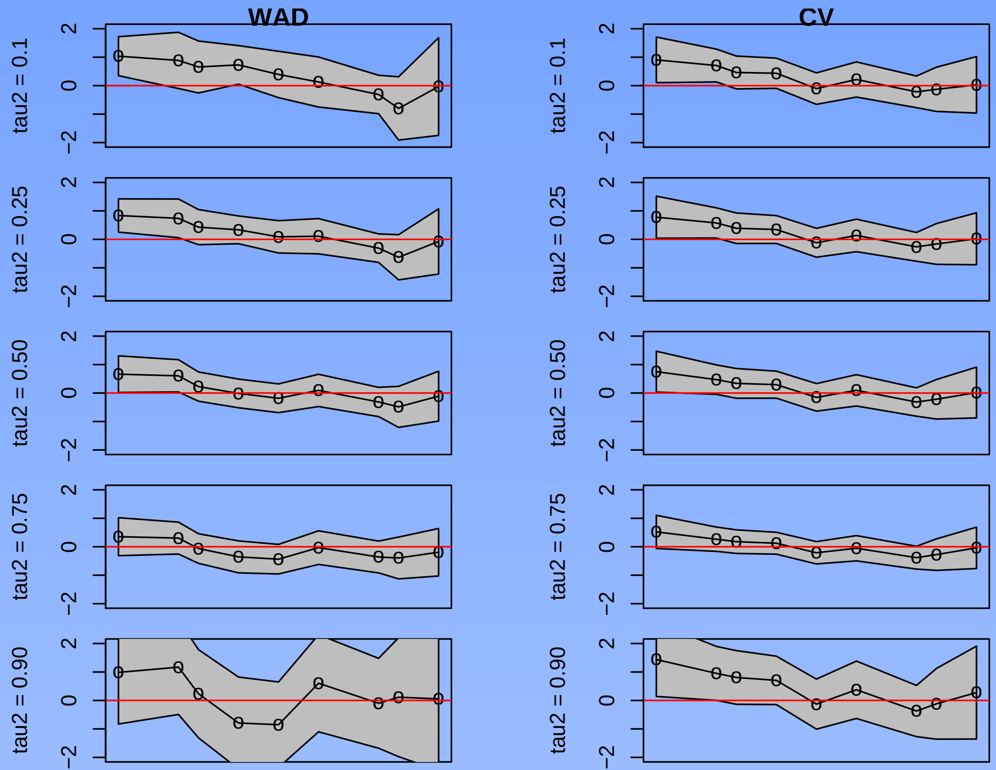
where n_i is total enrollment of school i and s_{ij} is the socio-economic status of student j, scored 1-5, in school i.

This variable clearly influences class size, via funding decisions, but *conditional on our other covariates* is plausibly independent of student performance.

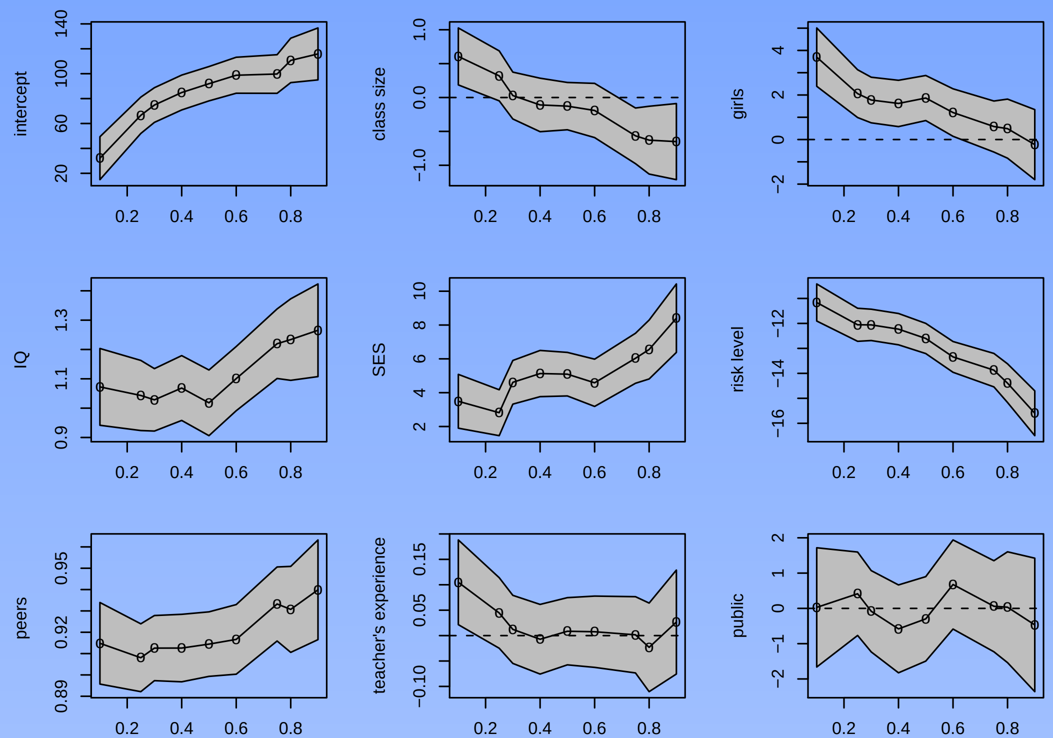
Language Performance: Endogonous Class Size Effect



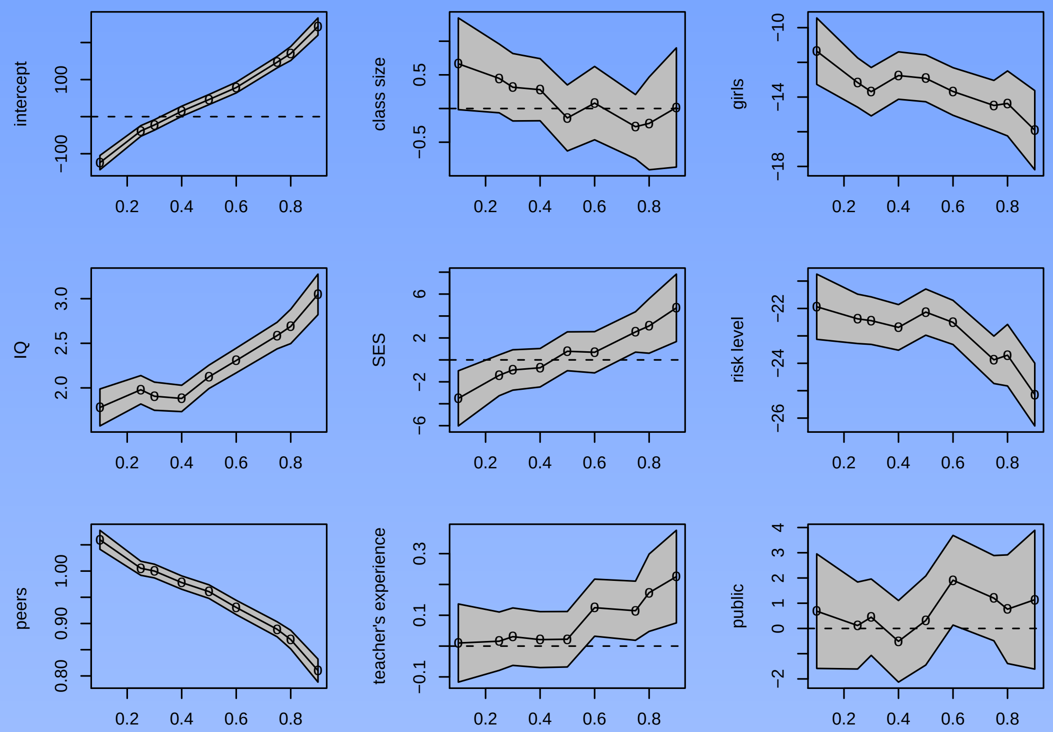
Mathematics Performance: Endogonous Class Size Effect



Other Covariate Effects on Language: Endogonous Class Size



Other Covariate Effects on Math: Endogonous Class Size



Policy Prescriptions

- With class size treated as exogonous:
  - ⋆ Smaller classes improve performance in both language and math.

Policy Prescriptions

- With class size treated as exogonous:
  - ⋆ Smaller classes improve performance in both language and math.

- With class size treated as endogonous:
  - ⋆ For language: weaker students do better with large classes, while better students do marginally better with smaller classes.
  - ⋆ For mathematics: weaker students do slightly better with small classes, and there are no significant class size effects for average and good students.

Policy Prescriptions

- With class size treated as exogonous:
  - ⋆ Smaller classes improve performance in both language and math.

- With class size treated as endogonous:
  - ⋆ For language: weaker students do better with large classes, while better students do marginally better with smaller classes.
  - ⋆ For mathematics: weaker students do slightly better with small classes, and there are no significant class size effects for average and good students.

- Other covariate effects are unaffected by endogoneity treatment of class size.
- Peer effects remain a major empirical challenge.

Conclusions

- Triangular structural models facilitate causal analysis via recursive conditioning.
- 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.