Topics in Econometrics

Instrumental Variable Quantile Regression

Instructor: Ma, Jun

Renmin University of China

June 2, 2022

- This class: Chernozhukov, V., Hansen, C., 2004. The effects of 401(k) participation on the wealth distribution: An instrumental quantile regression analysis. *Review of Economics and Statistics* (CH04)
- Matlab code for IVQR: https://voices.uchicago.edu/christianhansen/code-and-data/

CH04 Background

- In the early 1980s, the United States introduced several tax-deferred savings options in an effort to increase individual saving for retirement.
- The two options which have generated the most interest are individual retirement accounts (IRAs) and 401(k) plans.
- Tax-deferred IRAs and 401(k) plans are similar in that both allow the individual to deduct contributions from taxable income and allow tax-free accrual of interest on assets held within the plan.
- The key differences between the two savings options are that employers provide 401(k) plans, and employers may also match a certain percentage of an employee's contribution.
- Because 401(k) plans are provided by employers, only workers in firms offering plans are eligible for participation, whereas participation in IRAs is open to everyone.

CH04 Research Question

- Objective: estimate the causal effect of participation in 401(k) on savings/financial assets/wealth.
- Main issue: "preference for savings" is unobserved.
- Individuals with the highest unobserved preference for saving would be most likely to choose to participate in tax-advantaged retirement savings plans and would also have higher savings in other assets than individuals with lower preference.
- Conventional estimates that do not allow for saver heterogeneity and selection of the participation state will be biased-upward, tending to overstate the actual savings effects of 401(k) and IRA participation.

Addressing the Endogeneity Issue

- They argue that 401(k) eligibility can be taken as exogenous given income. The argument is motivated by the fact that eligibility is determined by the employer, and so may be taken as exogenous conditional on covariates.
- CH04 present IV regression results. Note that the standard linear IV model assumes homogeneous effects.
- The IVQR model (Chernozhukov and Hansen, 2005) allows for heterogenous treatment effects.
- The quantile treatment effects (QTE) are identified in the IVQR model.
- QTE provides insight into difference between distributions of potential outcomes under different treatment status, other than mean difference.

Quantile Methods in Economics

- The oldest quantile method, median regression, was introduced by Boscovitch in 1760. It is believed to have a even longer history than the OLS (Gauss, 1795).
- The quantile regression, which generalized median regression, was introduced to economics by the influential work Koenker and Bassett (1978).
- Quantile methods gained popularity among applied econometricians since 1990's.

Unconditional Quantiles

- A continuous random variable Y, its cumulative distribution function (CDF): $F_Y(y) = \Pr[Y \leq y]$.
- The quantile of Y ($Q_Y(\tau)$, $\tau \in (0,1)$) is just the generalized inverse of its CDF: $Q_Y(\tau) = \inf\{y : F_Y(y) \geq \tau\}$ (if F_Y is strictly increasing, the right hand side can be simplified to the usual inverse: $Q_Y(\tau) = F_Y^{-1}(\tau)$). $\tau = 1/2$: the median.
- Define the "check function":

$$\rho_\tau(u) = \begin{cases} \tau \cdot u & \text{if } u \geq 0 \\ -(1-\tau) \cdot u & \text{if } u < 0. \end{cases}$$

  Then, $Q_Y(\tau)$ solves the minimization problem :

$$\mathrm{E}[\rho_\tau(Y - r)] \geq \mathrm{E}[\rho_\tau(Y - Q_Y(\tau))], \forall r.$$

- An estimator (sample quantile) of $Q_Y(\tau)$ is the minimizer of $\sum_{i=1}^{n} \rho_\tau(Y_i - r)$ with respect to r.

Conditional Quantiles

- We similarly define the conditional quantiles. Dependent variable Y, possibly multi-dimensional independent variables $X \in \mathbb{R}^k$ and their conditional CDF $F_{Y|X}(y \mid x)$.
- The conditional quantile of Y $\left(Q_{Y|X}(\tau \mid x), \tau \in (0,1)\right)$ is

$$Q_{Y|X}(\tau \mid x) = \inf\left\{y : F_{Y|X}(y|x) \geq \tau\right\}.$$

- $Q_{Y|X}(\tau \mid X)$ solves:

$$\mathrm{E}\left[\rho_\tau\left(Y - r(X)\right)\right] \geq \mathrm{E}\left[\rho_\tau\left(Y - Q_{Y|X}(\tau \mid X)\right)\right], \forall r(\cdot).$$

- Suppose that $Q_{Y|X}(\tau \mid X)$ is a linear function of X: $Q_{Y|X}(\tau \mid X) = X^\top \beta_\tau$. Then it must hold that:

$$\mathrm{E}\left[\rho_\tau\left(Y - X^\top b\right)\right] \geq \mathrm{E}\left[\rho_\tau\left(Y - X^\top \beta_\tau\right)\right], \forall b.$$

Quantile Regression

- Under the linear model $Q_{Y|X}(\tau \mid X) = X^{\top}\beta_{\tau}$, the quantile regression estimator $\widehat{\beta}_{\tau}$ is defined to be the minimizer of $\sum_{i=1}^{n} \rho_{\tau}\left(Y_i - X_i^{\top} b\right)$ with respect to b.
- Note that the objective function of this minimization problem is not differentiable (and also multi-dimensional) and so commonly used algorithms (e.g., Newton-Raphson) can not be applied.
- We formulate it as a linear programming problem and solve it by the simplex method:

$$\min_{(b,u,v) \in \mathbb{R}^k \times \mathbb{R}_+^{2n}} \tau \cdot \iota^{\top} u + (1-\tau) \cdot \iota^{\top} v \text{ s.t. } Xb + u - v = Y,$$

  where $\iota = (1, ..., 1)^{\top} \in \mathbb{R}^n$, $X = (X_1, ..., X_n)^{\top} \in \mathbb{R}^{n \times k}$ and $Y = (Y_1, ..., Y_n)^{\top} \in \mathbb{R}^n$.
- Simplex method works only when n is relatively small. Use interior point method when n is very large.

Motivation for QR

- Angrist and Pischke (Chapter 7)
- The motivation for the use of QR to look at the wage distribution comes from labor economists' interest in the question of how inequality varies conditional on covariates like education and experience.
- Table 7.1.1 reports schooling coefficients from QRs estimated using the 1980, 1990, and 2000 censuses.
- The models used to construct these estimates control for race and a quadratic function of potential labor market experience.
- In contrast to the simple pattern in 1980 and 1990 census data, QR estimates from the 2000 census differ markedly across quantiles.
- By 2000, inequality began to increase with education as well: a pattern of increasing schooling coefficients across quantities means the wage distribution spreads out as education increases.

Motivation for QR

TABLE 7.1.1
Quantile regression coefficients for schooling in the 1980, 1990, and 2000 censuses

		Desc. Stats.		Quantile Regression Estimates					OLS Estimates	
Census	Obs.	Mean	SD	0.1	0.25	0.5	0.75	0.9	Coeff.	Root MSE
1980	65,023	6.4	.67	.074	.074	.068	.070	.079	.072	.63
				(.002)	(.001)	(.001)	(.001)	(.001)	(.001)	
1990	86,785	6.5	.69	.112	.110	.106	.111	.137	.114	.64
				(.003)	(.001)	(.001)	(.001)	(.003)	(.001)	
2000	97,397	6.5	.75	.092	.105	.111	.120	.157	.114	.69
				(.002)	(.001)	(.001)	(.001)	(.004)	(.001)	

Notes: Adapted from Angrist, Chernozhukov, and Fernandez-Val (2006). The table reports quantile regression estimates of the returns to schooling in a model for log wages, with OLS estimates shown at the right for comparison. The sample includes U.S.-born white and black men aged 40–49. The sample size and the mean and standard deviation of log wages in each census extract are shown at the left. Standard errors are reported in parentheses. All models control for race and potential experience. Sampling weights were used for the 2000 census estimates.

Asymptotic Properties of QR

- Consistency: $\widehat{\beta}_\tau \to_p \beta_\tau$.
- Asymptotic normality: $\sqrt{n}\left(\widehat{\beta}_\tau - \beta_\tau\right) \to_d \mathrm{N}\left(0, \mathrm{H}^{-1}\mathrm{V}\mathrm{H}^{-1}\right)$,
  $\mathrm{V} = \tau\left(1-\tau\right)\mathrm{E}\left[XX^\top\right]$ and
  $\mathrm{H} = \mathrm{E}\left[f_{Y|X}\left(X^\top\beta_\tau \mid X\right)XX^\top\right]$.
- A consistent estimator of H (Powell, 1984):

  $$\widehat{\mathrm{H}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{Y_i - X_i^\top\widehat{\beta}_\tau}{h}\right)X_iX_i^\top.$$

  Notice that a user-specified bandwidth h is required.
- Alternatively, bootstrap percentile confidence intervals for β_τ.

IVQR Model

- The IVQR model is developed within the potential outcome framework. Potential continuous outcomes, which vary among units, are indexed against (possibly more than 2) potential treatment states $d \in \mathcal{D}$ and denoted Y_d.
- The potential outcomes $\{Y_d\}_{d\in\mathcal{D}}$ are latent. Given the observed treatment D, the observed outcome is $Y = Y_D$.
- The objective is to learn about features of the distributions of potential outcomes $\{Y_d\}_{d\in\mathcal{D}}$. Of primary interest are the $\tau-$th quantiles of potential outcomes, conditional on $X = x$: $Q_{Y_d|X}(\tau \mid x)$, $\tau \in (0,1)$.
- Assume: $Y_d = q(U_d, d, X)$, where $q(\cdot, d, x)$ is strictly increasing and $U_d \sim \mathrm{U}(0,1)$. Then, $\forall \tau \in (0,1)$,

$$\Pr\left[Y_d \leq q(\tau, d, x) \mid X = x\right] = \tau \Longrightarrow Q_{Y_d|X}(\tau \mid x) = q(\tau, d, x).$$

- Quantile treatment effect (QTE, impact of D on quantiles of potential outcomes):

$$q(\tau, d_1, x) - q(\tau, d_0, x) = Q_{Y_1|X}(\tau \mid x) - Q_{Y_0|X}(\tau \mid x).$$

- D is dependent on $\{U_d\}_{d\in\mathcal{D}}$, inducing endogeneity.
- $Q_{Y|D,X}(\tau \mid d, x)$ is generally not equal to $Q_{Y_d|X}(\tau \mid x)$.
- The structural errors $\{U_d\}_{d\in\mathcal{D}}$ are responsible for heterogeneity of potential outcomes $Y_d = q(U_d, d, X)$ and individual treatment effects $Y_{d_1} - Y_{d_0} = q(U_{d_1}, d_1, X) - q(U_{d_0}, d_0, X)$ among individuals with the same observed characteristics X.
- This error term determines the relative ranking of observationally equivalent individuals in the distribution of potential outcomes given the individuals' observed characteristics, and thus we refer to $\{U_d\}_{d\in\mathcal{D}}$ as the rank variables.
- One may think of $\{U_d\}_{d\in\mathcal{D}}$ as representing some unobserved characteristics.

Assumptions and Implications

1. $Y_d = q(U_d, d, X)$, where $q(\cdot, d, x)$ is strictly increasing and left continuous and $U_d \sim \text{Uniform}(0, 1)$
2. Conditional on X and $\forall d \in \mathcal{D}$, U_d is independent of the instrumental variables Z.
3. $D = \delta(Z, X, V)$ for some unknown function δ and random error V.
4. Rank similarity: conditional on (X, Z, V), $\{U_d\}_{d \in \mathcal{D}}$ are identically distributed.

Under these assumptions, the main implication is:

$$\Pr[Y \leq q(\tau, D, X) \mid X, Z] = \tau, \ \forall \tau \in (0, 1).$$

- We do not require that the instruments Z are independent of the error V in the selection equation.
- V may depend on $\{U_d\}_{d \in \mathcal{D}}$ and include other unobserved variables that affect treatment status.

Nonparametric Identification

- For fixed τ, the identification problem is whether or not $q(\tau,\cdot,\cdot)$ is the only function that satisfies the restriction $\Pr[Y \leq q(\tau,D,X) \mid X,Z] = \tau$.
- Assume no covariate, $D \in \{0,1\}$ and $Z \in \{0,1\}$. $q(\tau,\cdot)$ can be represented by a vector $(q(0), q(1))$ and $q(\tau,D) = D \cdot q(1) + (1-D) \cdot q(0)$.
- We know $\Pr[Y \leq D \cdot q(1) + (1-D) \cdot q(0) \mid Z = z] = \tau$, for $z \in \{0,1\}$.
- Define

$$
\begin{aligned}
\Pi_0(y_0,y_1) &= \Pr[Y \leq D \cdot y_1 + (1-D) \cdot y_0 \mid Z = 0] - \tau \ . \\
\Pi_1(y_0,y_1) &= \Pr[Y \leq D \cdot y_1 + (1-D) \cdot y_0 \mid Z = 1] - \tau \ . \\
\Pi(y_0,y_1) &= \begin{bmatrix} \Pi_0(y_0,y_1) \\ \Pi_1(y_0,y_1) \end{bmatrix}.
\end{aligned}
$$

- The identification problem is whether or not $\Pi(y_0,y_1) = (0,0)^\top$ has a unique solution (at $(q(0), q(1))$).

- The Jacobian matrix:

$$
\begin{aligned}
J\left(y_0, y_1\right) &= \left[\begin{array}{cc} \frac{\partial \Pi_0\left(y_0, y_1\right)}{\partial y_0} & \frac{\partial \Pi_0\left(y_0, y_1\right)}{\partial y_1} \\ \frac{\partial \Pi_1\left(y_0, y_1\right)}{\partial y_0} & \frac{\partial \Pi_1\left(y_0, y_1\right)}{\partial y_1} \end{array}\right] \\
&= \left[\begin{array}{cc} f_{Y \mid D, Z}\left(y_0 \mid 0,0\right) p_{00} & f_{Y \mid D, Z}\left(y_1 \mid 1,0\right) p_{10} \\ f_{Y \mid D, Z}\left(y_0 \mid 0,1\right) p_{01} & f_{Y \mid D, Z}\left(y_1 \mid 1,1\right) p_{11} \end{array}\right],
\end{aligned}
$$

  where $p_{dz} = \Pr\left[D = d \mid Z = z\right]$.
- A necessary condition for identification is that $J\left(q\left(0\right), q\left(1\right)\right)$ has full rank ($\det\left(J\left(q\left(0\right), q\left(1\right)\right)\right) \neq 0$).
- See Chernozhukov and Hansen (2005) for sufficient conditions for nonparametric identification in the general IVQR model.

Model Assumptions in CH04

- An empirical model of savings decisions may be embedded in this framework.
- The wealth Y_d in the participation status $d \in \{0,1\}$ is generated by $Y_d = q(d, X, U_d)$, where $U_d \sim \mathrm{Uniform}(0,1)$ is understood as the rank in the preference-for-savings distribution.
- The individual selects the 401(k) participation status to maximize the expected utility:

$$\begin{aligned} D &= \underset{d\in\{0,1\}}{\operatorname{argmax}} \mathrm{E}\left[W_d\left(Y_d\right) \mid X, Z, V\right] \\ &= \underset{d\in\{0,1\}}{\operatorname{argmax}} \mathrm{E}\left[W_d\left(q\left(d, X, U_d\right)\right) \mid X, Z, V\right], \end{aligned}$$

  where W_d is the (random) utility function that may depend on random variables that are unobserved to the individual, in status d, V is an information component that is observed to the individual but unobserved to the researcher.
- V may depend on $\{U_0, U_1\}$ and include other unobserved variables.

Discussion on Rank Similarity

- A stronger assumption is "rank invariance": $U_0 = U_1 = U$. In this case, individuals' preference for savings does not change in different states.
- This assumption is implausible in the current context. Employers' match rates vary in the participation status. U_1 seems to be random conditionally on U_0.
- The rank similarity condition relaxes the exact invariance of ranks by allowing noisy, unsystematic variations of U_d across $d \in \{0, 1\}$, conditionally on (V, X, Z).
- This relaxation allows for variation in the ranks across the treatment states, requiring only a "rank invariance in expectation".
- It states that given the information in (V, X, Z) employed to make the selection of treatment D, the expectations of any function of the rank U_d does not vary across the treatment states.

- Though we feel that similarity may be a reasonable assumption in many contexts, imposing similarity is not innocuous.
- In the context of 401(k) participation, matching practices of employers could jeopardize the validity of the similarity assumption.
- Individuals in firms with high match rates may be expected to have a higher rank in the asset distribution than workers in firms with less generous match rates.
- The distribution of U_d may be different across the treatment state, since in the participation state, one more variable (match rate) affects U_1.
- Similarity may still hold in the presence of the employer match if the rank U_1 is insensitive to the match rate.
- CH04 uses another empirical model (Abadie, Angrist and Imbens, 2002) that does not assume rank similarity for robustness check.

Estimation and Inference

- We focus on linear-in-parameters structural quantile models at a single quantile of interest τ:

$$q(\tau, d, x) = d^\top \alpha_0(\tau) + x^\top \beta_0(\tau).$$

- $\alpha_0(\tau)$ captures the causal effect of the endogenous variables D on the τ-th quantile of the conditional distribution of potential outcomes Y_d given $X = x$.

Generalized Methods of Moments

- Unconditional moment conditions:
$$\mathrm{E}\left[\left(\tau - 1\left(Y - D^\top \alpha_0 - X^\top \beta_0 \leq 0\right)\right)\Psi\right] = 0$$
where $\Psi = \Psi(X, Z)$ is a vector of functions of the instruments and endogenous variables (e.g., $\Psi = \left(X^\top, Z^\top\right)^\top$).
- For $\theta = (\alpha, \beta)$, $V = (Y, D, X, Z)$ and $g_\tau(V, \theta) = \left(\tau - 1\left(Y - D^\top \alpha - X^\top \beta \leq 0\right)\right)\Psi$
- one may then form the sample analog of the right-hand-side of the equation
$$\widehat{g}(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_\tau(V_i, \theta)$$
and estimate $\theta_0 = (\alpha_0, \beta_0)$ by GMM:
$$\widehat{\theta} = \left(\widehat{\alpha}, \widehat{\beta}\right) = \underset{\theta}{\operatorname{argmin}}\, \widehat{g}(\theta)^\top \widehat{\Omega}\widehat{g}(\theta),$$
where $\widehat{\Omega}$ is the GMM weighting matrix: $\widehat{\Omega} = \left(\tau(1-\tau)\, n^{-1}\sum_{i=1}^{n} \Psi_i \Psi_i^\top\right)^{-1}$.

Quasi-Bayesian Approach

- The chief difficulty in implementing standard GMM estimation is that the objective function being minimized is non-smooth and non-convex.
- One option is to take the quasi-Bayesian approach of Chernozhukov and Hong (2003).
- The quasi-likelihood: $L_N(\theta) = \exp\left(-n \cdot \widehat{g}(\theta)^\top \widehat{\Omega}\widehat{g}(\theta)/2\right)$, when coupled with a prior density $\pi(\theta)$ over model parameters θ, defines a "posterior" density:

$$\pi_N(\theta) = \frac{L_N(\theta)\,\pi(\theta)}{\int L_N(\theta)\,\pi(\theta)\,\mathrm{d}\theta}.$$

- The quasi posterior mean $\widehat{\theta} = \int \theta \pi_N(\theta)\,\mathrm{d}\theta$ is consistent for model parameters with the same asymptotic distribution as the standard GMM estimator.
- Chernozhukov and Hong (2003) also demonstrate that confidence intervals may be obtained by taking quasi-posterior quantiles.

Inverse Quantile Regression

- Chernozhukov and Hansen (2006, 2008).
- The method is based on:

$$\Pr\left[Y \leq \left(D^\top \alpha_0 + X^\top \beta_0\right) \mid X, Z\right] = \tau$$
$$\Longrightarrow Q_{Y - D^\top \alpha_0}\left(\tau \mid X, Z\right) = X^\top \beta_0 + Z^\top \gamma_0 \text{ with } \gamma_0 = 0.$$

- For any hypothesized value a for α_0, estimate coefficients $\beta\left(a\right)$ and $\gamma\left(a\right)$ from the model

$$Q_{Y - D^\top a}\left(\tau \mid X, Z\right) = X^\top \beta\left(a\right) + Z^\top \gamma\left(a\right)$$

  by running ordinary linear QR of $Q_{Y-D^\top a}$ onto X and Z. Note that this can be solved by linear programming.
- Let $\widehat{\beta}\left(a\right)$ and $\widehat{\gamma}\left(a\right)$ denote the resulting estimators. Let $\widehat{\Omega}\left(a\right)$ denote the estimated asymptotic covariance matrix of $\sqrt{n}\left(\widehat{\gamma}\left(a\right) - \gamma\left(a\right)\right)$.

- The IQR estimator of α_0:

$$\widehat{\alpha} = \underset{a}{\operatorname{argmin}}\, \widehat{\gamma}\left(a\right)^{\top} \widehat{\Omega}^{-1} \widehat{\gamma}\left(a\right).$$

Note that a is very often low-dimensional, in empirical applications.

Unconditional QTE

- Unconditional QTE can be obtained from the conditional quantile functions in three steps.
- Step 1:

$$F_{Y_d}(y \mid x) = \int_0^1 1\left(d^\top \alpha_0(\tau) + x^\top \beta_0(\tau) \leq y\right) \mathrm{d}\tau.$$

- Step 2:

$$F_{Y_d}(y) = \int F_{Y_d}(y \mid x) \, \mathrm{d}F_X(x).$$

- Step 3:

$$\text{Unconditional } \tau - \text{QTE} = F_{Y_{d_1}}^{-1}(\tau) - F_{Y_{d_0}}^{-1}(\tau).$$

Extreme Quantiles

- Suppose D is one-dimensional $D \in \mathbb{R}$. Consider the case where we are interested in $\alpha_0(\tau)$ for very small τ. The coefficient of interest is an extreme quantile (τ is close to zero or one).
- Estimation of extreme quantiles is of interest in many empirical settings, e.g., in the study of the determinants of birthweight, where much interest is given to very low quantiles of birthweight.
- One may estimate $\alpha_0(\tau)$ by the standard methods. The resulting estimator usually suffers from low precision due to sparsity near extreme quantiles.

(Extreme) Quantile Extrapolation

- A very simple and easy-to-implement approach due to Firpo, Galvao, Pinto, Poirier and Sanroman (2021).
- We can postulate a flexible parametric model: $\alpha_0(\tau) = \sum_{j=1}^{J} \theta_j \tau^j$ $((\theta_1, ..., \theta_J) \in \mathbb{R}^J)$.
- θ can be estimated using a large number of quantiles further from zero, resulting in increased precision compared to standard IVQR.
- These benefits occur under correct specification of $\alpha_0(\tau) = \sum_{j=1}^{J} \theta_j \tau^j$.
- Choose a grid: $\tau_1 < \tau_2 < \cdots < \tau_L$ $(L > J)$ where τ_1, τ_L are far away from 0 and 1. Take any IVQR estimator $\widehat{\alpha}_0(\tau)$.
- Solve:

$$\left(\widehat{\theta}_1, ..., \widehat{\theta}_J\right) = \underset{\theta_1,...,\theta_J}{\operatorname{argmin}} \sum_{l=1}^{L} \left(\widehat{\alpha}_0(\tau_l) - \sum_{j=1}^{J} \theta_j \tau_l^j\right)^2.$$

For τ_0 that is very close to 0 or 1, use $\sum_{j=1}^{J} \widehat{\theta}_j \tau_0^j$ and bootstrap the standard errors.

CH04 Data

- 9915 households from from wave 4 of the 1990 Survey of Income and Program Participation (SIPP).
- These data include a variable for whether a person works for a firm that offers a 401(k) plan. Households in which a member works for such a firm are classified as eligible for a 401(k).
- Households with a positive 401(k) balance are classified as participants, and eligible households with a zero balance are considered nonparticipants.
- CH04 focuses on total wealth, net financial assets, and net non-401(k) financial assets.
- Total wealth is net financial assets plus housing equity and the value of business, property, and motor vehicle.
- Covariates: age, income, family size, education, marital status, two-earner status, defined benefit (DB) pension status, IRA participation status, and homeownership status.

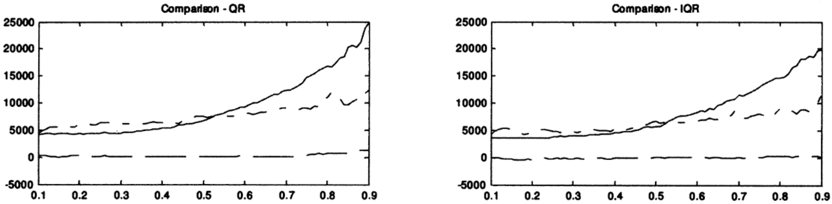
OLS and 2SLS

TABLE 3.—OLS AND 2SLS ESTIMATES OF EFFECT OF 401(K) PARTICIPATION

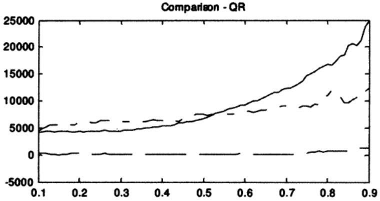
Sample	N	First Stage	Net Financial Assets		Net Non-401(k) Financial Assets		Total Wealth	
			OLS	2SLS	OLS	2SLS	OLS	2SLS
A. Full Sample								
Full Sample	9915	0.697	14,250	13,087	778	−355	10,694	9,259
		(0.006)	(1,551)	(1,922)	(1,477)	(1,855)	(2,388)	(3,035)

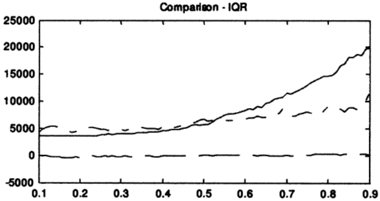
- In the full sample, the 2SLS estimates are uniformly smaller than the OLS estimates, confirming the intuition that the OLS estimates should be upward biased.
- The magnitude of effects on wealth and financial assets are similar, though slightly larger for net financial assets, suggesting little substitution between 401(k) assets and other forms of wealth. On the other hand, 401(k) participation has relatively little effect on net non-401(k) financial assets.
- These results suggest that the majority of the increase in net financial assets may be attributed to new saving due to 401 (k) plans and not to substitution from other forms of wealth.

QR and IVQR



- Another interesting feature of the results is that the effect of participation on net financial assets is highly nonconstant, appearing to increase monotonically in the quantile index.
- This result suggests that, conditional on income and other observables, people who rank higher in the conditional wealth distribution are affected far more than those ranking lower in the conditional distribution. In addition, the effect is strongly positive across the entire distribution.
- In particular, if people were simply substituting financial assets held in 401(k)s for other forms of financial assets, the effect of 401(k) participation on net financial assets would be zero.





- These findings suggest that the increase in net financial assets observed in the lower tail of the conditional assets distribution can be interpreted as an increase in wealth, while the increase in the upper tail of the distribution is being mitigated by substitution with some other (nonfinancial) component of wealth.

Tests on IVQR Process

TABLE 4.—TESTS ON THE INSTRUMENTAL QUANTILE REGRESSION PROCESS IN THE FULL SAMPLE

	Net Financial Assets		Net Non-401(k) Financial Assets		Total Wealth	
Null Hypothesis	Statistic	$c_{.95}$	Statistic	$c_{.95}$	Statistic	$c_{.95}$
No effect	12.875	3.009	0.921	2.882	4.538	3.003
Constant effect	9.093	3.321	0.843	3.452	1.850	3.213
Exogeneity	3.851	3.209	2.287	3.056	1.899	3.086

- The tests strongly reject the hypothesis that the effect of 401(k) participation on net financial assets is constant and confirm that it is significantly different from 0.
- The hypothesis of exogeneity of treatment is rejected for net financial assets.
- The tests fail to reject both the hypothesis of a constant treatment effect and the hypothesis of exogeneity for total wealth and net non-401(k) financial asset.

CH04 Conclusion

- The results suggest that there is substitution between assets held in 401(k)s and other components of wealth in the upper tail of the wealth distribution, but that most financial assets held in 401(k)s in the lower tail of the distribution represent new savings.
- This has important policy implications, as people in the low tail of the net financial asset distribution are also likely to be the people with the lowest retirement savings.

Abadie, Angrist and Imbens (2002) Model

- Binary D and binary Z.
- Model assumptions: $Y = q(D, X, U)$ (unrestricted, U can be of any dimension) but $D = 1(V \leq \delta(Z, X))$ (unobserved error in the selection equation is scalar). Assume $\delta(1, x) > \delta(0, x)$.
- $D_{xz} = 1(V \leq \delta(z, x))$. The "complier group" (with $X = x$): $D_{x1} = 1, D_{x0} = 0$ ($D_{x1} > D_{x0}$).
- $Q_{Y_d|X,D_{x1}>D_{x0}}(\tau \mid x, D_{x1} > D_{x0}) = F^{-1}_{Y_d|X,D_{x1}>D_{x0}}(\tau \mid x, D_{x1} > D_{x0})$ is identified.
- QTE for the complier group is identified and can be estimated by reweighted QR.
- Drop the rank similarity assumption, but the cost is that only QTE corresponding to a sub-population is identifiable.
- The AAI model and the CH model are complementary and non-nested.