# Instrumental Variable Quantile Regression

Thomas Wiemann
*University of Chicago*

TA Discussion # 8
Econ 31740

March 7, 2022

# Outline

1. Running Empirical Example: Chernozhukov and Hansen (2004)
2. The IVQR Model of Chernozhukov and Hansen (2005)
   - ▷ Model Overview
   - ▷ Point Identification
   - ▷ Estimation Approaches
3. Suggestions for Further Reading

Today's discussion is based in large part on Chernozhukov et al. (2017), but with additional discussion of the application of instrumental variable quantile regression in Chernozhukov and Hansen (2004).

# Running Empirical Example

401(k) plans were introduced in the 70-80s to increase retirement savings.

- 401(k)s are a tax-deferred savings option w/ employer contribution.

Participation in 401(k) is not plausibly random.

- Heterogeneous saving preferences reflect in both 401(k) participation and accumulated savings.

Rich literature on the effect of 401(k) plans on wealth in the 90s-2000s:

- Poterba et al. (1995) assume 401(k) participation is random conditional on income (and other covariates);
- Abadie (2003) assumes 401(k) *eligibility* is random conditional on incomes (and other covariates), and considers it as an instrument for 401(k) participation;

Downside of this early literature is the focus on measures of central tendency: the mean or median. From a policy perspective, we may be particularly interested in, e.g., the lower tail of the savings distribution.

# Running Empirical Example (Contd.)

Chernozhukov and Hansen (2004) use instrumental variable quantile regression (IVQR) developed by Chernozhukov and Hansen (2005) to study the effect of 401(k) participation on the wealth *distribution*.

Following Abadie (2003), 401(k) eligibility as assumed to be a valid instrument on 401(k) participation conditional on income (and other covariates).

To fix ideas, consider:

- $Y \equiv$ net financial wealth;
- $D \equiv$ 401(k) participation;
- $X \equiv$ vector of income, age, family size, education, martial status, two-earner status, defined benefit pension status, IRA participation, homeownership;
- $Z \equiv$ 401(k) eligibility.

The sample is a collection of $(y_i, d_i, x_i, z_i) \sim (Y, D, X, Z)$ of size $n = 9915$ from the 1991 SIPP, assumed to be iid.

# Model Overview

Let $(Y, D, X)$ denote the outcome, treatment, controls, and instrument, respectively. Consider the setting

- ▷ $\operatorname{supp} Y \subset \mathbb{R}$, $\operatorname{supp} X \subset \mathbb{R}^{d_x}$;
- ▷ $\operatorname{supp} D = \{0, 1\}$.

Let $Y(d)$ denote the potential outcome with $D = d$ fixed. Note that only one potential outcome is observed for each unit: $Y \equiv Y(D)$.

Key object of interest is the quantile function of the potential outcomes conditional on $X = x$:

$$\mathrm{Q}_{Y(d)}(\tau|x) = q(\tau, d, x), \quad \forall \tau \in (0, 1). \tag{1}$$

Given (1), consider the conditional quantile treatment effect $(\mathrm{CQTE})$:

$$\mathrm{CQTE}_\tau(x) \equiv q(\tau, 1, x) - q(\tau, 0, x). \tag{2}$$

The $\mathrm{CQTE}_\tau(x)$ gives the effect of the treatment on the $\tau$th conditional quantiles of the potential outcomes.

# Model Overview (Contd.)

A useful representation of the potential outcomes $\{Y(0), Y(1)\}$ conditional on $X = x$ is obtained from the Skorohod representation:

$$Y(d) = q(U_d, d, x), \quad \text{where } U_d \sim \mathcal{U}(0,1). \tag{3}$$

$U_d$ is the structural error term that results in heterogeneity in potential outcomes across units with the same observed characteristics $X = x$.

- ▷ May think of $U_d$ as representing unobserved characteristics, e.g., propensity to save.

Chernozhukov and Hansen (2005) refer to $U_d$ as the rank variable, because it determines the ranking of obsessionally equivalent units in the distribution of potential outcomes.

# The Fundamental Problem of Causal Inference

The challenge of identifying the $\mathrm{CQTE}$ is the familiar fundamental problem of causal inference.

The $\tau$th quantile of $Y(1)$ is defined by the restriction

$$\begin{aligned}\tau =& P\left(Y(1) \leq Q_{Y(1)}(\tau|X)|X=x\right)\\ =& P\left(Y(1) \leq Q_{Y(1)}(\tau|X)|D=1, X=x\right) P(D=1|X=x)\\ &+ P\left(Y(1) \leq Q_{Y(1)}(\tau|X)|D=0, X=x\right) P(D=0|X=x)\\ =& P\left(Y \leq Q_{Y}(\tau, X)|D=1, X=x\right) P(D=1|X=x)+\\ &+ P\left(Y(1) \leq Q_{Y(1)}(\tau|X)|D=0, X=x\right) P(D=0|X=x),\end{aligned} \tag{4}$$

where $P\left(Y(1) \leq Q_{Y(1)}(\tau|X)\right)$ is not identified from the data w/o additional assumptions. (The argument for $Y(0)$ follows analogously.)

# The Fundamental Problem of Causal Inference (Contd.)

As a consequence of (4), conventional quantile regression of $Y$ on $(D, X)$, which relies on the restriction

$$\tau = P\left(Y \leq Q_Y(\tau|D, X)|D, X\right) \tag{5}$$

is not appropriate for learning about the $\mathrm{CQTE}$ in settings where $Y(0), Y(1) \not\perp D|X$.

Chernozhukov and Hansen (2005) develop the IVQR model that provides conditions under which the $\mathrm{CQTE}$ is identified under endogeneity.

In addition to $(Y, D, X)$ considered so far, consider an instrumental variable $Z$. Chernozhukov and Hansen (2005) provide results for binary, discrete, and continuous $D$ and $Z$. The below primarily focuses on the setting when $\operatorname{supp} D = \operatorname{supp} Z = \{0, 1\}$.

- ▷ Relevant for 401(k) setting of Chernozhukov and Hansen (2004).

# The IVQR Model

The main assumptions of Chernozhukov and Hansen (2005) are:

**A1.** Potential Outcomes: Conditional on $X = x$, for each $d$, $Y(d) = q(U_d, d, x)$, where $q(\tau, d, x)$ is strictly increasing in $\tau$ and $U_d \sim \mathcal{U}(0, 1)$.

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

**A3.** Selection: $D \equiv \delta(Z, X, V)$ for some unknown function $\delta$ 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 $Y \equiv q(U_D, D, X)$, $D$, $X$, and $Z$.

# The IVQR Model (Contd.)

**Theorem 1 (Theorem 1 of Chernozhukov and Hansen (2005))**

*Suppose conditions **A1-A5** hold. Then,* $\forall \tau \in (0,1)$, *a.s.*

$$P\left(Y \leq q(\tau, D, X) | X, Z\right) = P\left(Y < q(\tau, D, X) | X, Z\right) = \tau, \qquad (6)$$

*and* $U_D \sim \mathcal{U}(0,1)$ *conditional on* $Z$ *and* $X$.

Theorem 1 is key because it provides a means of identifying the $\mathrm{CQTE}$ in a general heterogeneous effects model.

It is also constructive in that it suggests using the sample analogue of (6) for estimation of the conditional quantile functions of the potential outcome distributions.

# Discussion of the IVQR Model

Consider **A1-A5** in the context of Chernozhukov and Hansen (2004).

By the Skorohod representation, we have

$$Y(d) = q(U_d, d, X), \quad U_d \sim \mathcal{U}(0,1), \tag{7}$$

where $Y(d)$ is the net financial wealth of the individual in 401(k) participation state $d \in \{0,1\}$, $U_d$ is an unobserved variable (e.g., propensity to save), and $q$ is the conditional quantile function of $Y(d)$.

Assumption **A1** requires that $q$ is *strictly* increasing in $\tau$.

- $Y$ may not have a point mass. (What about zero-savings?)

Assumption **A2** requires that $\{U_d\} \perp\!\!\!\perp Z|X$.

- Implies that we may estimate the QTE of $Z$ on $Y$ via conventional quantile regression.
- Here: motivated by reasoning of Abadie (2003), which highlights that eligibility is determined by the employer and may thus be plausibly exogeneous conditional on employer characteristics such as paid wages (or: income of the employee).

Assumption **A3** requires that $D = \delta(Z, X, V)$ for some unknown function $\delta$ and random vector $V$.

We may motivate this by letting individuals select 401(k) participation to maximize expected utility:

$$\begin{aligned} D &= \arg\max_{d \in \{0,1\}} E\left[W\left(Y(d), d\right)|X, Z, V\right] \\ &= \arg\max_{d \in \{0,1\}} E\left[W\left(q(U_d, d, x), d\right)|X, Z, V\right], \end{aligned} \tag{8}$$

where $W$ is the unobserved utility function and $V$ is an unobserved information component that depends on $U_d$ and contains other information pertinent for treatment selection.

# Discussion of the IVQR Model (Contd.)

Assumption **A4** (a) implies that saving preferences across participation states $d \in \{0, 1\}$ are homogeneous. That is, $U_0 = U_1$.

- ▷ An untreated individual at the $\tau$th quantile of $Y(0)$ would be at the $\tau$th quantile of $Y(1)$ under treatment.

A relaxed version of rank invariance is rank similarity. Assumption **A4** (b) requires $U_0 \overset{d}{=} U_1$ (i.e., identically distributed conditional on $(V, X, Z)$).

- ▷ Only requires rank invariance *in expectation*.
- ▷ $\{U_d\}$ are the same ex-ante, but realized values can be different.

Chernozhukov and Hansen (2004) highlight the following:

- ▷ "[...] *matching practices of employers could jeopardize the validity of the similarity assumption. This is because 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.*"

# Discussion of the IVQR Model (Contd.)

To see the issue, suppose supp $V = \{\text{lmr}, \text{hmr}\}$, corresponding to a low match rate and high match rate, respectively. Omitting dependence on $(X, Z)$, Assumption **A4** requires

$$\begin{aligned} & E[U_0|V = \text{lmr}] = E[U_1|V = \text{lmr}] \\ \text{and} \quad & E[U_0|V = \text{hmr}] = E[U_1|V = \text{hmr}]. \end{aligned} \tag{9}$$

It seems reasonable to assume $E[U_0|V = \text{lmr}] = E[U_0|V = \text{hmr}]$, but then we must also have $E[U_1|V = \text{lmr}] = E[U_1|V = \text{hmr}]$ which may be less plausible.

- ▷ If employees follow rules of thumb (e.g., target savings), then $U_d$ is insensitive to match rates.

Finally: Assumption **A5** follows immediately from the sample.

# Point Identification

For ease of notation, the following suppresses conditioning on $X = x$.

By Theorem 1, there exists at least one function $q(d) \equiv q(d, x, \tau) : P(Y \leq q(D)|Z) = \tau$. The question of point identification is whether there exists only a single such function.

Notice $q(d) = (1 - d)q(0) + dq(1)$ so that $q$ can be represented by a vector of its values: $q = (q(0), q(1))^\top$. Then, for vectors $y = (y_0, y_1)^\top$ and the moment equations

$$\Pi(y) \equiv \begin{bmatrix} P(Y \leq y_D|Z = 0) - \tau \\ P(Y \leq y_D|Z = 1) - \tau \end{bmatrix}, \tag{10}$$

where $y_D = (1 - D)y_0 + Dy_1$, the identification question is whether $y = q$ uniquely solves $\Pi(y) = 0$.

# Point Identification (Contd.)

Rank conditions on the Jacobian $\Pi(y)$ result in point identification.

Define the Jacobian $\Pi'(y)$ with respect to $y = (y_0, y_1)^\top$ as

$$\Pi'(y) \equiv \begin{bmatrix} f_{Y,D}(y_0, 0|Z=0) & f_{Y,D}(y_1, 1|Z=0) \\ f_{Y,D}(y_0, 0|Z=1) & f_{Y,D}(y_1, 1|Z=1) \end{bmatrix}, \tag{11}$$

where $f_{Y,D}$ denotes the joint pdf of $(Y, D)$.

The condition that $\Pi'(y)$ is full rank requires that the impact instrument $Z$ on the joint distribution of $(Y, D)$ is sufficiently strong.

- ▷ $Z$ cannot be (conditionally) independent of $D$.

# Point Identification (Contd.)

Note that full rank of $\Pi'(y)$ is equivalent to $\det \Pi'(y) \neq 0$, which implies

$$\frac{f_{Y,D}(y_1, 1|Z=1)}{f_{Y,D}(y_0, 0|Z=1)} > \frac{f_{Y,D}(y_1, 1|Z=0)}{f_{Y,D}(y_0, 0|Z=0)}, \tag{12}$$

or the same condition but with $>$ replaced with $<$.

- ▷ Referred to as the *monotone likelihood ratio condition*.
- ▷ Stronger than the conventional rank condition that requires $Z$ to be correlated with $D$.

Condition (12) is trivially satisfied in the 401(k) setting of Chernozhukov and Hansen (2004):

- ▷ $f_{Y,D}(y_1, 1|Z=0) = 0$ because it is impossible for an individual to participate in a 401(k) (i.e., $D=1$) if she is not eligible for a 401(k) (i.e., $Z=0$).

# Point Identification (Contd.)

The following result relates identification of the quantiles of the potential outcomes with the full rank of $\Pi'(y)$ for $y \in \mathcal{L}$, with $\mathcal{L}$ being a set of potential solutions to the moment condition $\Pi(y) = 0$ (defined on the next slide).

**Theorem 2 (Theorem 2 of Chernozhukov and Hansen (2005))**

*Suppose conditions **A1**-**A5** hold, and* $\operatorname{supp} D = \operatorname{supp} Z = \{0, 1\}$*. Assume that for the set* $\mathcal{L}$ *and* $s_d$ *(see next slide), (i)* $\Pi'(y)$ *is continuous* $\forall y \in \mathcal{L}$*, and (ii)* $q(d) \in s_d, \forall d \in \{0, 1\}$*. Then the* $\tau$*-quantiles of potential outcomes,* $q = (q(0), q(1))^\top$*, are point identified if* $\operatorname{rank} \Pi'(y)$ *is full rank* $\forall y \in \mathcal{L}$*.*

# Point Identification (Contd.)

The sets $\mathcal{L}$ and $s_d$ referred to in the statement of Theorem 2 are defined as follows:

Fix some small constants $\delta > 0$ and $\underline{f} > 0$. Define $\mathcal{L}$ as the closed rectangle containing all vectors $y = (y_0, y_1)^\top$ that satisfy

(a) $\forall z \in \{0, 1\}$, $P(Y < y_D | Z = z) \in [\tau - \delta, \tau + \delta]$, and

(b) $\forall d \in \{0, 1\}$, $y_d \in s_d$ where

$$s_d \equiv \{\lambda : f_Y(\lambda | d, z) \geq \underline{f}, \forall z : P(D = d | Z = z) > 0\}.$$

The first condition (a) defines $\mathcal{L}$ as the set of potential solutions to $\Pi(y) = 0$.

The second condition (b) requires the solutions to $\Pi(y) = 0$ to be in the conditional support of $Y$.

# Estimation

We now turn to estimating the conditional quantile function $q(\tau, d, X)$ in the point identified setting. Following Section 9.3 of Chernozhukov et al. (2017), we focus on linear-in-parameters structural quantile models

$$q(\tau, d, x) = d^\top \alpha_0(\tau) + x^\top \beta_0(\tau), \tag{13}$$

for some single quantile of interest $\tau \in (0, 1)$.

- ▷ $(\alpha_0(\tau), \beta_0(\tau))$ are unknown and fixed.
- ▷ $\alpha_0(\tau)$ is the causal effect of $D$ on the $\tau$-th quantile of $Y(d)$ given $X = x$. Similarly for $\beta_0(\tau)$.

We consider estimation of $\alpha_0(\tau)$ in the presence of a (low-dimensional) nuisance parameter $\beta_0(\tau)$.

In the setting where $d \in \{0, 1\}$, we have $\mathrm{CQTE}_\tau(x) = \mathrm{QTE}_\tau = \alpha_0(\tau)$.

# Estimation via GMM

For ease of notation, the following omits dependence on $\tau$.

A generalized methods of moments (GMM) estimator for $(\alpha_0, \beta_0)$ can be constructed from equation (6) in Theorem 1.

In particular, an implication of (6) combined with the (13) is

$$E\left[(\tau - \mathbb{1}\{Y - D^\top \alpha_0 - X^\top \beta_0 \leq 0\})\Psi\right], \tag{14}$$

where $\Psi \equiv \Psi(X, Z)$ is a vector of functions of the instruments.

- ▷ For example, $\Psi(X, Z) = (X^\top, Z^\top)^\top$.
- ▷ Minimal necessary condition $\dim \Psi \geq \dim \alpha_0 + \dim \beta_0$.

# Estimation via GMM (Contd.)

Let $\theta \equiv (\alpha, \beta)$, $M \equiv (Y, D, X, Z)$, and define

$$g_\tau(M, \theta) \equiv (\tau - \mathbb{1}\{Y - D^\top \alpha_0 - X^\top \beta_0 \leq 0\})\Psi. \tag{15}$$

The sample analogue of (14) is then given by

$$\hat{g}_n(\theta) \equiv n^{-1} \sum_{i=1}^{n} g_\tau(m_i, \theta), \tag{16}$$

where $m_i \equiv (y_i, d_i, x_i, z_i)$. Then estimate $\theta_0 \equiv (\alpha_0, \beta_0)$ by GMM via

$$\hat{\theta} = \arg\min_{\theta \in \Theta} n \hat{g}_n(\theta)^\top \Omega_n \hat{g}_n(\theta), \tag{17}$$

where $\Omega_n$ is the GMM weighting matrix typically set to

$$\Omega_n = \left( \tau(1 - \tau) n^{-1} \sum_{i=1}^{n} \Psi_i \Psi_i^\top \right)^{-1}. \tag{18}$$

# Estimation via Quasi-Bayes

A key difficulty in GMM estimation is that the optimization problem in (17) is non-smooth and non-convex in general.

- ▷ Concerning as $\dim \beta$ can be quite large (while $\dim \alpha$ typically small).

Chernozhukov and Hong (2003) consider a quasi-Bayesian approach that allows to leverage MCMC sampling.

Consider forming the GMM-based "quasi-likelihood"

$$L_n(\theta) \equiv \exp\left(-n\hat{g}_n(\theta)^\top \Omega_n \hat{g}_n(\theta)\right). \quad (19)$$

When coupled with an appropriate prior $\theta(\theta)$, this results in a "quasi-posterior" for $\theta$:

$$\pi_n(\theta) \propto L_n(\theta)\pi(\theta) \quad (20)$$

# Estimation via Quasi-Bayes (Contd.)

Chernozhukov and Hong (2003) show that Bayes estimators under the $L_1$ and $L_2$-loss, which correspond to the posterior median and posterior mean, respectively, are consistent for $\theta_0$.

E.g., given $\pi_n(\theta)$, an estimate of $\theta_0$ is

$$\hat{\theta} = \int \theta d\pi_n(\theta). \tag{21}$$

The authors also show that valid frequentist confidence intervals can be constructed by taking appropriate quantiles of the quasi-posterior.

# Estimation via Inverse Quantile Regression

An alternative approach to estimation of $\alpha_0$ is proposed in Chernozhukov and Hansen (2006). This approach makes explicit that $\beta_0$ are nuisance parameters.

Inverse quantile regression (IQR) is based on the insight that (6) combined with the (13) implies

$$Q_{Y-D^\top \alpha_0}(\tau|X,Z) = X^\top \beta_0 + Z^\top \gamma_0, \quad \text{with } \gamma_0 \equiv 0. \tag{22}$$

▷ $Z$ does not have any explanatory value conditional on $(X, D)$.

IQR *via grid search* then proceeds as follows:

(1) Hypothesize a value $a$ for $\alpha_0$.

(2) Estimate $(\beta(a), \gamma(a))$ via quantile regression of $Y - D^\top a$ on $(X, Z)$. Let $(\hat{\beta}(a), \hat{\gamma}(a))$ denote corresponding estimates.

(3) Define $W_n(a) \equiv n\hat{\gamma}(a)^\top \hat{\Omega}_n(a)\hat{\gamma}(a)$, where $\hat{\Omega}_n(a)$ is the estimated covariance matrix of $\sqrt{n}(\hat{\gamma}(a) - \gamma(a))$.

(4) Repeat steps (1)-(3) over a grid of values $\mathcal{A}$ for $a$ and select $\hat{\alpha} = \arg\min_{a\in\mathcal{A}} W_n(a)$.

# Estimation via Inverse Quantile Regression (Contd.)

A key benefit of IQR is that the non-convex optimization problem is over only the parameters $\alpha$, which are typically low-dimensional.

The IQR minimization problem $\min_{a \in \mathcal{A}} W_n(a)$ can also be solved with procedures other than grid search, of course. If $\dim \alpha_0$ is very small (e.g., 1), grid search seems like a reasonable and easy-to-implement approach.

Chernozhukov and Hansen (2006) analyze the asymptotic properties of $\hat{\alpha}$ under point identification and verify their asymptotic normality.

# The Effects of 401(k) Participation

We now revisit the empirical example.

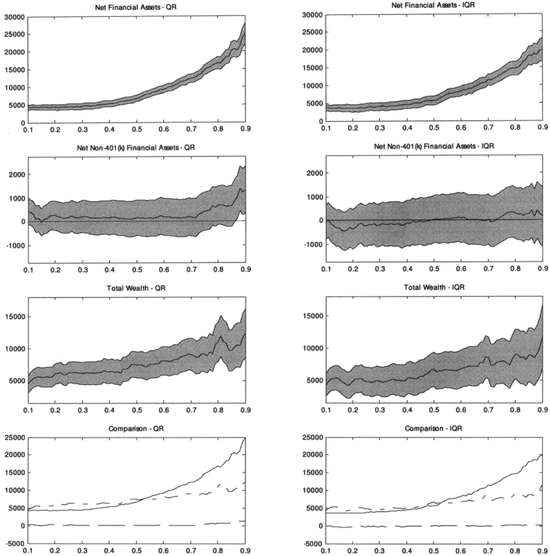
Recall the setting of Chernozhukov and Hansen (2004):

- ▷ $Y \equiv$ net financial wealth;
- ▷ $D \equiv$ 401(k) participation;
- ▷ $X \equiv$ vector of income, age, family size, education, martial status, two-earner status, defined benefit pension status, IRA participation, homeownership;
- ▷ $Z \equiv$ 401(k) eligibility.

The sample is a collection of $(y_i, d_i, x_i, z_i) \sim (Y, D, X, Z)$ of size $n = 9915$ from the 1991 SIPP, assumed to be iid.

Figure 1 on the next slide is a key empirical result of the paper.

# Figure 1 of Chernozhukov and Hansen (2004)




The sample size is 9915. The left column contains standard quantile regression estimates, and the right column contains instrumental quantile regression. Each panel is labeled with the dependent variable used in estimation of the presented results. The bottom panel in each column compares the point estimates for each wealth measure. The solid line corresponds to net financial assets, the dashed line to net non-401(k) financial assets, and the dash-dot line to total wealth. The vertical axis measures the dollar increase in the wealth measure due to 401(k) participation. The quantile of the conditional wealth distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [0.10, 0.90]$ at 0.01-unit intervals.

# Suggestions for Further Reading

Instrumental variable quantile regression is an active area of research with many interesting extensions and applications.

A few suggestions for further reading:

- Section 9.4 of Chernozhukov et al. (2017) on IVQR w/ high-dim $X$;
- Melly and Wüthrich (2017) on *local* quantile treatment effects;
- Pouliot (2021) on MIP formulation of IQR.

# References


Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263.

Chernozhukov, V. and Hansen, C. (2004). The effects of 401 (k) participation on the wealth distribution: an instrumental quantile regression analysis. *Review of Economics and statistics*, 86(3):735–751.

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

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525.

Chernozhukov, V., Hansen, C., and Wüthrich, K. (2017). Instrumental variable quantile regression. In Koenker, R., Chernozhukov, V., He, X., and Peng, L., editors, *Handbook of Quantile Regression*, chapter 9, pages 119–143. CRC Press.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Melly, B. and Wüthrich, K. (2017). High-dimensional quantile regression. In Koenker, R., Chernozhukov, V., He, X., and Peng, L., editors, *Handbook of Quantile Regression*, chapter 10, pages 145–164. CRC Press.

Poterba, J. M., Venti, S. F., and Wise, D. A. (1995). Do 401 (k) contributions crowd out other personal saving? *Journal of Public Economics*, 58(1):1–32.

Pouliot, G. A. (2021). Instrumental variables quantile regression with multivariate endogenous variable. working paper.