Local Structural Quantile Effects in a Model with a Nonseparable Control Variable[*]

Sung Jae Jun[†]

First Draft: November 2005

This Version: May 2008

Abstract

I propose an asymptotically normal estimator for a local structural quantile effect in a nonseparable triangular system (see e.g., Imbens and Newey (2002) and Chesher (2003)). In order to obtain a practical estimator while keeping the flexibility of the triangular system, I consider a linear specification that has a control variable included in a nonparametric and nonseparable manner. In simulation experiments, local instruments and identification pitfalls are also discussed: the instruments do not affect a particular quantile of the endogenous variable although they do affect others. It is shown that the proposed method enables inference about local effects with locally relevant instruments at the expense of the parametric $\sqrt{n}$ rate.

Key Words: Triangular Models, Local Instruments, Control Variables, Quantile Regression.

JEL Classification Code: C14, C30

[†]Department of Economics, The Pennsylvania State University, 608 Kern Graduate Building, University Park PA 16802, sjun@psu.edu

1 Introduction

I propose a new semiparametric estimator of a model with nonseparable latent variables. The proposed model is linear in variables with random coefficients, which are unknown functions of the (uniformly normalized) latent variables V and ϵ. In the case with a single endogenous regressor, the model is a triangular system with two equations, where one equation contains V, and the other equation contains both V and ϵ. The parameter of interest $\alpha(\tau_1, \tau_2)$ is the marginal effect of the endogenous variable on the outcome variable evaluated at $V = \tau_1$ and $\epsilon = \tau_2$, where $\tau_1, \tau_2 \in (0, 1)$ are chosen by the researcher. I develop a simple two-step estimator, and establish its asymptotic normal distribution.

This setup is attractive, because it is simple but preserves the flexibility of nonseparable triangular systems (see e.g., Imbens and Newey (2002) and Chesher (2003, 2007)). Unlike fully nonparametric approaches, including many exogenous covariates does not cause the curse of dimensionality. The marginal effect $\alpha(V, \epsilon)$ of the endogenous variable is allowed to depend on both V and ϵ. Instruments are local in the sense that they only need to affect the τ_1 conditional quantile of the endogenous variable when the interest is at $V = \tau_1$. The cost of this degree of flexibility is that the proposed estimator converges at a nonparametric rate.

Identification and estimation of heterogeneous marginal effects have been frequently discussed in the literature. Koenker and Basset (1978) provide the first possibility via quantile regression when there is no endogeneity. Chernozhukov and Hansen (2005, 2006) propose a way of modeling endogeneity in quantile regression models, which can be understood in a generalized method of moments (GMM) framework. Imbens and Newey (2002) and Chesher (2003, 2007) consider nonseparable triangular models to capture heterogeneity in marginal effects. In particular, Chesher (2003) provides a set of conditions under which heterogeneous marginal effects are nonparametrically identified in general triangular models. Although his identification results lead to a straightforward nonparametric estimation method, it will suffer from the dimensionality issue when many exogenous covariates are included. Lee (2007) takes a semiparametric approach, where he uses a partially linear quantile regression method. Ma and Koenker (2006) propose a parametric version of Chesher (2003).

The model proposed in this paper can be compared with Lee (2007) and Ma and Koenker (2006). In contrast to Ma and Koenker (2006), the approach of this paper is semiparametric in the sense that the functional form of the random coefficients of the structural equation need not be specified. Lee (2007) takes a control function approach, where he assumes that the control function is additively separable. This separability assumption prevents $\alpha(\tau_1, \tau_2)$ from being heterogeneous over τ_1, and hence Lee (2007) does not preserve the flexibility of triangular models.

Since the dependence between the endogenous variable and the random coefficients is due to the fact that they both depend on V, this common latent variable plays the role of a control variable. It is in fact used as a control variable in the estimation procedure, because the two-step procedure essectially fixes V at τ_1 by kernel smoothing.

Note that the parameter $\alpha(\tau_1, \tau_2)$ is a local feature. The reason why this local parameter is focused on is that identification of the average quantile effect (i.e., $\int_0^1 \alpha(s, \tau_2)ds$) requires that the instruments affect all the distributional points of the endogenous variable, which is a strong assumption in nonseparable models. In the simulation section, I investigate the possibility of *local instruments* by conducting small experiments with the Angrist and Krueger data: the instruments do not affect a particular conditional quantile of the endogenous variable although they do affect others. The Monte Carlo results confirm that the proposed estimator only requires local relevance of instruments.

2   Motivation: Returns to Education and Instrumental Variables

Consider the following wage equation

$$W = g(s, X, A, U), \tag{1}$$

where W is a wage, s is a level of education, X is a vector of individual characteristics, A is unobserved ability, and U is unobserved market fortune. Suppose that g is increasing in U. Individuals choose their level of education to maximize the difference between expected wages given

their characteristics and ability and costs:

$$S = \arg\max_{s} E\big(g(s, X, A, U)|X, A\big) - C(s, Z, X, A), \tag{2}$$

where C is the cost of getting a particular level of education, which generally depends on other characteristics and unobserved ability. Letting $h(s, X, A) = E\big(g(s, X, A, U)|X, A\big)$ (i.e. the expected wage with s level of education given individual characteristics and ability), the optimal level of education satisfies the first order condition

$$h_s(S, X, A) = C_s(S, Z, X, A), \tag{3}$$

which shows that S is a function of Z, X, and A.[1] Assume that $h_{ss}(S, X, A) - C_{ss}(S, Z, X, A) < 0$ so that the second order condition is also satisfied. It then follows from the implicit function theorem that

$$\frac{\partial S}{\partial A} = \frac{C_{sA}(S, Z, X, A) - h_{sA}(S, X, A)}{h_{ss}(S, X, A) - C_{ss}(S, Z, X, A)} > 0 \tag{4}$$

as long as $h_{sA}(S, X, A) - C_{sA}(S, Z, X, A) > 0$ (i.e. the net marginal benefit of taking extra education is increasing in A). Therefore, this setup suggests the following model for returns to education

$$\begin{cases} W = g(S, X, A, U), \\ S = S(Z, X, A), \end{cases} \tag{5}$$

where g is increasing in U and S is increasing in A. Chesher (2003) pointed out that one of the most important features of this model is that the returns to education $g_s(S, X, A, U)$ generally depends on A and U. He also showed how useful quantile regression and recursive conditioning is to identify $g_s(S, X, a, u)$ at particular values of $A = a$ and $U = u$.

Chesher's identification conditions include a rank condition, which can be written as

$$\frac{\partial S(Z, X, a)}{\partial Z} = \frac{C_{sA}(S, Z, X, a)}{h_{ss}(S, X, a) - C_{ss}(S, Z, X, a)} \neq 0, \tag{6}$$

---

[1]Partial derivatives will be denoted by sub–indices.

where the first equality is by the implicit function theorem. Note however that the sensitivity of S to changes in Z generally depends on the level of unobserved ability. In particular, no economic theory guarantees that Z satisfies condition (6) for every value of a in support of A. In the sense of Chesher (2003), Z is a relevant instrument to identify $g_s(S, X, a, u)$ only for those values of a such that $C_{sA}(S, Z, X, a) \neq 0$. Of course, when $C_{sA}(S, Z, X, a) \neq 0$ for every value a in support of A, then $g_s(S, X, a, u)$ is identified for every a and u in support of A and U, which would provide more complete information about returns to education than any other average notion of returns to education.

Although Chesher (2003) provided a set of conditions to nonparametrically identify $g_s(S, X, a, u)$ for particular values of $A = a$ and $U = u$, implementation of this model unavoidablely requires a certain amount of parametrization. The goal of this paper is to semiparametrically implement model (5) with keeping all the flexibility of this model. The essential flexibility to be preserved includes that (i) the returns to education $g_s(S, X, a, u)$ are allowed to be heterogeneous over both a and u, and (ii) the degree of relevance of Z as an instrument is allowed to be heterogeneous over a.

For the second part of flexibility, the data used by Angrist–Kruger (1991) provide an example.[2] They used birth–quarters as instrumental variables with the idea that the cost of taking extra education varies over different birth–quarters due to the compulsory schooling system of the US. However, since those with lower ability are more likely to be affected by the compulsory schooling system, this idea naturally suggests that birth–quarters will be more relevant for those with lower level of ability. This possiblity will be further discussed in section 5.

The parametrization taken in this paper uses random coefficients e.g. $W = S\tilde{\alpha}(A, U) + X'\tilde{\beta}(A, U)$.[3] To simplify presentation, (A, U) will be represented by $\left(F_A^{-1}(V), F_{U|A}^{-1}(\epsilon | F_A^{-1}(V))\right)$, where V and ϵ are two independent uniform random variables. It then leads to a model with random coefficients,

$$W = S\alpha(V, \epsilon) + X'\beta(V, \epsilon).$$

---

[2]The Angrist and Krueger's data measure education by the years of schooling, which is discrete. This is an issue that is not covered by this paper. Chesher (2005) showed that discrete variations in endogenous variables only deliever partial identification.

[3]As a matter of specification, $W = q_1(S)'\tilde{\alpha}^*(A, U) + q_2(X)'\tilde{\beta}^*(A, U)$ can be easily considered, where $q_1(\cdot)$ and $q_2(\cdot)$ are known functions.

3   The Model

I propose the following model.

$$\begin{cases} Y = D\alpha(V,\epsilon) + X'\beta(V,\epsilon) \\ D = X'\Pi_1(V) + Z'\Pi_2(V), \end{cases} \tag{7}$$

where Y is a scalar outcome variable, D is a scalar endogenous variable, X is a k_x dimensional vector of exogenous covariates, and Z is a k_z dimensional vector of instruments. V and ϵ are both scalars that represent unobserved components. The model is linear in variables with random coefficients. The endogenous variable is correlated with the random coefficients of the structural equation. Since the correlation is due to the fact that V appears twice, the common latent variable V plays the role of a control variable. Now, the following assumptions are made.

A1(Independence) V and ϵ are independent, and they are normalized to be uniformly distributed.X and Z are also assumed to be independent of ϵ and V.

A2(Monotonicity) $\tau \mapsto D\alpha(V,\tau) + X'\beta(V,\tau)$ is strictly increasing in $\tau \in (0,1)$.

Similarly, $\tau \mapsto X'\Pi_1(\tau) + Z'\Pi_2(\tau)$ is strictly increasing in $\tau \in (0,1)$.

This model covers the specifications by Lee (2007), Ma and Koenker (2006), and many others. It allows for two different kinds of heterogeneity, one kind via V and the other via ϵ. In the example of returns to schooling, Chesher (2003) calls one heterogeneity from unobserved ability and the other heterogeneity from unobserved market fortune. Although it is more restrictive than the fully nonparametric version of Imbens and Newey (2002) and Chesher (2003), it still preserves many interesting features of triangular systems and it has practical advantage. A completely nonparametric model suffers from the curse of dimensionality, when many exogenous variables are included. Although the proposed model is linear in variables, the control variable V is incorporated in a nonparametric and nonseparable manner, and it preserves all flexiblity of nonseparable models.

The parameter of interest is $\alpha(\tau_1, \tau_2)$ for some fixed $\tau_1, \tau_2 \in (0,1)$, which are chosen by the researcher.[4] It is the marginal effect of the endogenous variable on the outcome variable evaluated

---

[4]In the following discussion, τ_1 and τ_2 will be regarded as fixed values.

at $V = \tau_1, \epsilon = \tau_2$. For example, it can be interpreted as the returns to schooling for the person who has ability of τ_1, and has market fortune of τ_2. Chesher (2003) provides a set of general conditions for nonparametric identification of this parameter, which includes continuous variation of instruments. Instruments are not, however, so rich in practice. Since this is a linear model, discrete instruments can still deliver identification, as long as the following full rank assumption is satisfied.

A3(Full Rank) $\Pi_2(\tau_1)$ has full column rank for the chosen $\tau_1 \in (0, 1)$.

Proposition 1 *Suppose that **A1**, **A2**, and **A3** are satisfied. Let $W = [X' : Z']'$. If $E(WW')$ has full rank, then $\alpha(\tau_1, \tau_2), \beta(\tau_1, \tau_2)$ are identified.*

Proof: See the appendix.

Although proposition 1 is a special case of Chesher (2003, 2007), it exploits the linear quantile regression specification of the model. Therefore, in contrast to Chesher (2003, 2007), instruments need not be continuous and discrete instruments are also allowed.

The nature of instrument relevance is local in the sense that the full rank of $\Pi_2(\tau_1)$ is required only for a particular $\tau_1 \in (0, 1)$. For example, consider $D = Z\Phi^{-1}(V) + F^{-1}(V)$, where Z is non-negative almost surely, $\Phi^{-1}(s)$ is the inverse of the standard normal distribution, and $F^{-1}(s)$ is the inverse of some distribution function. In this case, the instrument is not informative for the conditional median of D, although it is relevant for the other conditional quantiles. When $\alpha(\tau_1, \tau_2)$ can vary over different τ_1, the identification pitfall at $\tau_1 = 0.5$ can be a problem in estimating the average quantile effect $\int_0^1 \alpha(v, \tau_2)dv$. Instead of estimating the average quantile effect, the local parameter $\alpha(\tau_1, \tau_2)$ will be focused on.

Ma and Koenker (2006) discuss estimation of $\alpha(\tau_1, \tau_2)$ and $\beta(\tau_1, \tau_2)$ under additional parametric assumptions. Alternatively, Lee (2007) considers the case where $\alpha(\tau_1, \tau_2)$ and $\beta(\tau_1, \tau_2)$ do not depend on τ_1 except for the intercept and shows that $\alpha(\cdot, \tau_2)$ can be estimated at the regular $\sqrt{n}$ rate. This approach loses the important feature of the triangular system, because it imposes separability of the control variable. In this paper, the control variable is allowed to be arbitrarily involved in the structural equation, which implies that the marginal effect of the endogenous variable

can be arbitrarily heterogeneous over different values of $V = \tau_1$. In contrast to Chesher (2003, 2007), the linear structural equation delivers identification when the instruments are discrete. This model is substantially simpler than fully nonparametric and nonseparable models, but yet flexibility of triangular models is well preserved.

For estimation, a local quantile regression approach is taken by using the kernel smoothing idea. I will write $Q_{A|B=b}(\tau)$ for the conditional τ quantile of A given $B = b$. When I do not specify any particular value of b, I will simply write $Q_{A|B}(\tau)$. Recall that

$$\Pr\big(Y \leq D\alpha(\tau_1, \tau_2) + X'\beta(\tau_1, \tau_2)|D = Q_{D|X,Z}(\tau_1), X, Z\big) = \tau_2,$$

because $D = Q_{D|X=x,Z=z}(\tau_1), X = x, Z = z$ is equivalent to $V = \tau_1, X = x, Z = z$. The idea is that $D\alpha(\tau_1, \tau_2) + X'\beta(\tau_1, \tau_2)$ is the conditional quantile of Y given X, Z for the people whose D is equal to $Q_{D|X,Z}(\tau_1)$. First, let $\rho_\tau(s) = |s| + (2\tau - 1)s$ be the check function for $\tau \in (0, 1)$.

Step 1. Estimate the τ_1 conditional quantile of D given $W_i = [X_i' : Z_i']'$ by $\hat{Q}_{D|W_i}(\tau_1) = W_i'\hat{\Pi}(\tau_1)$, where

$$\hat{\Pi}(\tau_1) = \arg \min_{\Pi} \sum_{i=1}^{n} \rho_{\tau_1}(D_i - W_i'\Pi).$$

Step 2. Estimate the parameters of interest by local smoothing quantile regression. Let $\hat{S}_n(\tau_1) = \{1 \leq i \leq n : |D_i - W_i'\hat{\Pi}(\tau_1)| \leq \frac{h_n}{2}\}$, where h_n is a bandwidth choice that shrinks to 0. Then estimate the parameters of interest by

$$\left[\hat{\alpha}(\tau_1, \tau_2) : \hat{\beta}(\tau_1, \tau_2)'\right]' = \arg \min_{a,b} \sum_{i \in \hat{S}_n(\tau_1)} \rho_{\tau_2}(Y_i - D_i a - X_i'b).$$

The second step is another quantile regression with a subsample of those whose first step residuals are sufficiently small. This is kernel-smoothing with a uniform kernel $1\{|s| \leq \frac{h_n}{2}\}$, where $1\{\cdot\}$ denotes an indicator function. I focus on the uniform kernel, because regularity conditions for asymptotics become easier to find. As long as there are enough observations with $|D_i - \hat{Q}_{D|X_i,Z_i}(\tau_1)|$ small, this approach is intuitively appealing. The resulting estimators will be shown to be consistent and asymptotically normal under some additional conditions.

4   Asymptotics

This section discusses the asymptotic properties of the proposed two-step estimators. Let $R_i(\tau_1) = D_i - W_i'\Pi(\tau_1)$ and $\epsilon_i(\tau_1, \tau_2) = Y_i - S_i'\theta(\tau_1, \tau_2)$, where $W_i = [X_i', \ Z_i']'$, $S_i = [D_i, \ X_i']'$, $\Pi_0 = \Pi(\tau_1) = [\Pi_1(\tau_1)', \ \Pi_2(\tau_1)]' \in \Xi$, and $\theta_0 = \theta(\tau_1, \tau_2) = [\alpha(\tau_1, \tau_2), \ \beta(\tau_1, \tau_2)']' \in \Theta$. The data $\{(Y_i, D_i, X_i', Z_i')\}_{i=1}^n$ are iid, and the observation index i will be suppressed when generic random variables are considered. I will write $f_A, f_{A|B}$ to denote the marginal density of A and the conditional density of A given B, respectively. I make the following assumptions.

Assumption A *There is a compact neighborhood $\mathcal{N}$ around 0 such that (i) for every $\delta \in \mathcal{N}$, $R(\tau_1) + W'\delta$ has a marginal density bounded away from 0 at 0 which is twice continuously differentiable, and (ii) $\sup_{\delta \in \mathcal{N}} |f_{R(\tau_1)+W'\delta}(t)| \leq b_0(t)$, $\sup_{\delta \in \mathcal{N}} |f'_{R(\tau_1)+W'\delta}(t)| \leq b_1(t)$ and $\sup_{\delta \in \mathcal{N}} |f''_{R(\tau_1)+W'\delta}(t)| \leq b_2(t)$ for some continuous functions $b_0(t)$, $b_1(t)$ and $b_2(t)$.*

Assumption B *$\epsilon(\tau_1, \tau_2)$ has a conditional density at 0 given $R(\tau_1)$ and S.*

Assumptions A and B require that Y and D are continuous random variables but they do not require continuous instruments. Assumption A says that $R(\tau_1) = D - W'\Pi_0$ has a *marginal* density bounded away from 0 at 0 and so do its small perturbations $R(\tau_1) + W'\delta$. The assumptions that Y and D are continuously distributed are crucial but the other assumptions are for regularity. Note also that assumption A ensures that there are sufficiently many observations in $\{1 \leq i \leq n : |R_i(\tau_1)| \leq \delta\}$ for any $\delta > 0$ as n increases.

Note that two sets of conditional moment restrictions are available:

$$\begin{cases} \Pr(D_i \leq W_i'\Pi_0 | W_i) = \tau_1 \\ \Pr(Y_i \leq S_i'\theta_0 | S_i, R_i(\tau_1) = 0) = \tau_2, \end{cases} \tag{8}$$

which implies

$$\begin{cases} E\big(W_i(1\{D_i \leq W_i'\Pi_0\} - \tau_1)\big) = 0 \\ E\big(S_i(1\{Y_i \leq S_i'\theta_0\} - \tau_2) | R_i(\tau_1) = 0\big) f_{R(\tau_1)}(0) = 0. \end{cases} \tag{9}$$

Since D_i is a function of X_i, Z_i conditional on $R_i(\tau_1) = 0$, $E\big(S_i(1\{Y_i \leq S_i'\theta_0\} - \tau_2)|R_i(\tau_1) = 0\big) = 0$ obtains by integrating over the distribution of X_i, Z_i.[5] Although these are not the only moment conditions implied by (8), I do not address the issues of efficiency in this paper.

The proposed estimators can be understood in a generalized method of moments (GMM) framework based on the moment conditions (9) (see e.g., Pakes and Pollard (1989), Newey and McFadden (1994)). In particular, the (quasi) first order conditions of the proposed estimator can be written as

$$
\begin{cases}
\frac{1}{n} \sum_{i=1}^{n} W_i(1\{D_i \leq W_i'\hat{\Pi}(\tau_1)\} - \tau_1) = o_p(\frac{1}{\sqrt{n}}) \\
\frac{1}{nh_n} \sum_{i=1}^{n} S_i(1\{Y_i \leq S_i'\hat{\theta}(\tau_1, \tau_2)\} - \tau_2)1\{|D_i - W_i'\hat{\Pi}(\tau_1)| \leq \frac{h_n}{2}\} = o_p(\frac{1}{\sqrt{nh_n}}),
\end{cases}
\tag{10}
$$

where h_n is a bandwidth choice with $h_n \downarrow 0$, $nh_n \to \infty$. It is in fact easy to show that

$$
M_n(\Pi, \theta) = \begin{bmatrix} M_{1n}(\Pi) \\ M_{2n}(\Pi, \theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} W_i(1\{D_i \leq W_i'\Pi\} - \tau_1) \\ \frac{1}{nh_n} \sum_{i=1}^{n} S_i(1\{Y_i \leq S_i'\theta\} - \tau_2)1\{|D_i - W_i'\Pi| \leq \frac{h_n}{2}\} \end{bmatrix}
$$

$$
\xrightarrow{p} M(\Pi, \theta) = \begin{bmatrix} M_1(\Pi) \\ M_2(\Pi, \theta) \end{bmatrix} = \begin{bmatrix} E\big(W_i(1\{D_i \leq W_i'\Pi\} - \tau_1)\big) \\ E\big(S_i(1\{Y_i \leq S_i'\theta\} - \tau_2)|D - W'\Pi = 0\big)f_{D-W'\Pi}(0) \end{bmatrix}
$$

for each Π and θ. The only complication is nondifferentiability of M_n due to the indicators, and the nonparametric convergence rate of the bottom part. Note also that we only need to assume that $D - W'\Pi$ has a marginal density when Π belongs to a neighborhood of Π_0, because Π_0 is consistently estimated from $M_1(\Pi_0) = 0$.

Assumption C *The parameters Π_0 and θ_0 uniquely solve $M(\Pi, \theta) = 0$, and they are in the interior of the compact parameter space $\Xi \times \Theta$.*

Assumption D *$E\big(M_n(\Pi, \theta)\big)$ and $M(\Pi, \theta)$ are continuous in $(\Pi, \theta) \in \Xi \times \Theta$. It is also assumed that $M(\Pi, \theta)$ is twice differentiable at (Π_0, θ_0).*

---

[5]When $R(\tau_1)$ has a density conditional on Y_i and S_i, this conditional expectation can be written as $E\big(S_i(1\{Y_i \leq S_i'\theta_0\} - \tau_2)f_{R(\tau_1)|Y,S}(0|Y_i, S_i)\big) = 0$. See the appendix.

Assumption E $E\big(W_i W_i' f_{R(\tau_1)|W}(0|W_i)\big)$ *and* $E\big(S_i S_i' f_{\epsilon(\tau_1,\tau_1)|R(\tau_1),S}(0|R_i(\tau_1),S_i)|R_i(\tau_1)=0\big)$ *have full column ranks.*

Assumption F $E\big(\|W_i\| \sup_s f_{R(\tau_1)|W}(s|W_i)\big) < \infty.$

Assumption G $\varphi(\theta,\delta,r) = E\big(S_i(1\{Y_i \leq S_i'\theta\} - \tau_2)|R_i(\tau_1) - W_i'\delta = r\big)$ *is twice continuously differentiable with respect to* r. *Moreover,* $\frac{\partial \varphi(\theta,\delta,r)}{\partial r}$ *and* $\frac{\partial^2 \varphi(\theta,\delta,r)}{\partial r^2}$ *are continuous in* $\theta \in \Theta$ *and* $\delta \in \mathcal{N}.$

Assumption H *Let* w_{ij} *and* s_{it} *be generic elements of* W_i *and* S_i. *For some* $\delta > 0$, $E(|w_{ij}|^{2+\delta}) < \infty$, *and* $E\big(|s_{it}|^{2+\delta}|R(\tau_1)=0\big) < \infty.$

Assumption C ensures global identification of (Π_0,θ_0), which is satisfied when the conditional density of $R(\tau_1)$ given W and the conditional density of $\epsilon(\tau_1,\tau_2)$ given W and $R(\tau_1) = 0$ are bounded away from 0 at 0; see the appendix.

Assumption D is standard. Differentiability of $M(\Pi,\theta)$ is guaranteed e.g. when $T(D_i,W_i) = E\big(S_i(1\{\epsilon(\tau_1,\tau_2) \leq 0\}-\tau_2)|D_i,W_i\big)$ is diffrentiable with respect to D_i. Assumption E is the standard rank condition but it is local. In particular, note that the full rank assumption fails to hold when $\Pi_2(\tau_1)$ is equal to 0.[6] The rank condition is therefore that the instruments are relevant for the τ_1 conditional quantile of D given W. Assumption H is needed for a central limit theorem. Since kernel smoothing is used, bias is a problem which I address by undersmoothing.

Assumption I $h_n \propto n^{-\kappa}$ *with* $\frac{1}{5} < \kappa < \frac{1}{2(1+\delta)}$ *for some* $\delta > 0.$

Lemma 1 *Let* $\Xi_0 = \Pi(\tau_1) + \mathcal{N}$. *Under assumptions A-B,* $\Upsilon_{n,h_n}\big(E(M_n(\Pi,\theta)) - M(\Pi,\theta)\big) =$
$$\begin{bmatrix} 0 \\ O(\sqrt{nh_n}h_n^2) \end{bmatrix} \textit{ uniformly over } \Xi_0 \times \Theta, \textit{ where } \Upsilon_{n,h_n} = \begin{bmatrix} \sqrt{n}I_{k_x+k_z} & 0 \\ 0 & \sqrt{nh_n}I_{1+k_x} \end{bmatrix}. \textit{ In particular,}$$
assumption I ensures $\Upsilon_{n,h_n}\big(E(M_n(\Pi,\theta)) - M(\Pi,\theta)\big) = o(1)$ *uniformly over* $\Xi_0 \times \Theta.$

Proof: See the appendix.

Now, I state the main theorem.

---

[6]Note that given $R(\tau_1) = 0$, $S' = [D : X'] = [X'\Pi_1(\tau_1) + Z'\Pi_2(\tau_1) : X'].$

Theorem 1 *Under assumptions A-I,*

$$
\begin{bmatrix}
\sqrt{n}\big(\big[\hat{\Pi}_1(\tau_1)',\ \hat{\Pi}_2(\tau_1)'\big]' - [\Pi_1(\tau_1)',\ \Pi_2(\tau_2)']'\big) \\
\sqrt{nh_n}\big(\big[\hat{\alpha}(\tau_1,\tau_2),\ \hat{\beta}(\tau_1,\tau_2)'\big]' - [\alpha(\tau_1,\tau_2),\ \beta(\tau_1,\tau_2)']'\big)
\end{bmatrix}
$$

$$
= -\begin{bmatrix}
\Gamma_{11}(\tau_1)^{-1} & 0 \\
0 & \Gamma_{22}(\tau_1,\tau_2)^{-1}
\end{bmatrix}\Upsilon_{n,h_n} M_n(\Pi_0,\theta_0) + o_p(1),
$$

where $\Gamma_{11}(\tau_1) = E\big(W_i W_i' f_{R(\tau_1)|W}(0|W_i)\big)$ *and* $\Gamma_{22}(\tau_1,\tau_2) = E\big(S_i S_i' f_{\epsilon(\tau_1,\tau_2)|R(\tau_1),S}(0|R_i(\tau_1),S_i)|R_i(\tau_1) = 0\big) f_{R(\tau_1)}(0)$. *In particular,*

$$
\sqrt{nh_n}\big(\big[\hat{\alpha}(\tau_1,\tau_2),\ \hat{\beta}(\tau_1,\tau_2)'\big]' - [\alpha(\tau_1,\tau_2),\ \beta(\tau_1,\tau_2)']'\big)
$$

$$
\xrightarrow{d} N\big(0,\Gamma_{22}(\tau_1,\tau_2)^{-1}V_{22}(\tau_1,\tau_2)\Gamma_{22}(\tau_1,\tau_2)^{-1}\big),
$$

as $n \to \infty$, *where* $V_{22}(\tau_1,\tau_2) = \tau_2(1-\tau_2)E(S_i S_i'|R_i(\tau_1) = 0)f_{R(\tau_1)}(0)$.

Proof: See the appendix.

The center of zero in the asymptotic distribution comes from undersmoothing as most of nonparametric estimators. Since the first–stage estimator has the faster parametric rate of convergence, it does not affect the asymptotic variance of the second–stage estimator. The linear expansion of $\sqrt{nh_n}\big(\hat{\theta}(\tau_1,\tau_2) - \theta_0(\tau_1,\tau_2)\big)$ also suggests joint normality of the estimators with different sets of quantiles.

Corollary 1 *Suppose that the conditions for theorem 1 are satisfied for* $\{(\tau_1,\tau_2),\ (\tilde{\tau}_1,\tilde{\tau}_2)\}$. *Then,*

$$
\begin{bmatrix}
\sqrt{nh_n}\big(\big[\hat{\alpha}(\tau_1,\tau_2),\ \hat{\beta}(\tau_1,\tau_2)'\big]' - [\alpha(\tau_1,\tau_2),\ \beta(\tau_1,\tau_2)']'\big) \\
\sqrt{nh_n}\big(\big[\hat{\alpha}(\tilde{\tau}_1,\tilde{\tau}_2),\ \hat{\beta}(\tilde{\tau}_1,\tilde{\tau}_2)'\big]' - [\alpha(\tilde{\tau}_1,\tilde{\tau}_2),\ \beta(\tilde{\tau}_1,\tilde{\tau}_2)']'\big)
\end{bmatrix}
$$

$$
\xrightarrow{d} N\left(0, \begin{bmatrix}
\Gamma_{22}(\tau_1,\tau_2)^{-1}V_{22}(\tau_1,\tau_2)\Gamma_{22}(\tau_1,\tau_2)^{-1} & COV(\tau_1,\tilde{\tau}_1,\tau_2,\tilde{\tau}_2) \\
COV(\tau_1,\tilde{\tau}_1,\tau_2,\tilde{\tau}_2)' & \Gamma_{22}(\tilde{\tau}_1,\tilde{\tau}_2)^{-1}V_{22}(\tilde{\tau}_1,\tilde{\tau}_2)\Gamma_{22}(\tilde{\tau}_1,\tilde{\tau}_2)^{-1}
\end{bmatrix}\right),
$$

where $COV(\tau_1,\tau_1,\tau_2,\tilde{\tau}_2) = (\min(\tau_2,\tilde{\tau}_2) - \tau_2\tilde{\tau}_2)\Gamma_{22}(\tau_1,\tau_2)^{-1}E(S_i S_i'|R_i(\tau_1) = 0)\Gamma_{22}(\tau_1,\tilde{\tau}_2)^{-1}$ *and*

$COV(\tau_1, \tilde{\tau}_1, \tau_2, \tilde{\tau}_2) = 0$ *if* $\tau_1 \neq \tilde{\tau}_1$.

Proof: When $\tau_1 = \tilde{\tau}_1$, it directly follows from the asymptotic expansion of theorem 1. When $\tau_1 \neq \tilde{\tau}_1$, note that $1\{|D_i - W_i'\Pi(\tau_1)| \leq \frac{h_n}{2}\}1\{|D_i - W_i'\Pi(\tilde{\tau}_1)| \leq \frac{h_n}{2}\} = 0$ for sufficiently large n. $\square$

Estimation of $\Gamma_{22}(\tau_1, \tau_2)$ and $V_{22}(\tau_1, \tau_2)$ can be done by various nonparametric methods (see e.g., Koenker (2005), Powell (1986)), although it could be more difficult to estimate $\Gamma_{22}(\tau_1, \tau_2)$. The following proposition shows one possibility.

Proposition 2 *Let $k(\cdot)$ be a kernel such that*

(i) $\sup_v |k(v)| < \infty$, $\int |k(v)| dv < \infty$, $\int k(v)^2 dv < \infty$, $\int k(v) dv = 1$, and $\int |k(v)||v| dv < \infty$.

(ii) $k(\cdot)$ is twice differentiable, and $\mathcal{S} = \{v \in \mathbb{R} : |k'(v)| > 0, \ k''(v) = 0\}$ is finite.

Suppose that $\epsilon_i(\tau_1, \tau_2)$ and $R_i(\tau_1)$ have a joint density conditional on S_i such that $\sup_{t,s} |\mathcal{D}_j f_{\epsilon(\tau_1,\tau_2),R(\tau_1)}(t,s|S_i)| \leq \phi(S_i)$ for some $\phi(S_i)$ with $E\big(||S_i||\phi(S_i)\big) < \infty$, where $\mathcal{D}_j$ denotes the derivative with respect to the jth argument. Let $b_{1n} \downarrow 0$, $b_{2n} \downarrow 0$ such that if W_i has bounded support, $\sqrt{n}b_{1n} \to \infty$ and $\sqrt{nh_n}b_{2n} \to \infty$, otherwise $\sqrt{n}b_{1n}^2 \to \infty$ and $\sqrt{nh_n}b_{2n}^2 \to \infty$. Then,

$$\hat{V}_{22}(\tau_1, \tau_2) = \frac{1}{nb_{1n}} \sum_{i=1}^n S_i S_i' k\big(\frac{-\hat{R}_i(\tau_1)}{b_{1n}}\big)\tau_2(1 - \tau_2) \xrightarrow{p} V_{22}(\tau_1, \tau_2),$$

$$\hat{\Gamma}_{22}(\tau_1, \tau_2) = \frac{1}{nb_{2n}^2} \sum_{i=1}^n S_i S_i' k\big(\frac{-\hat{R}_i(\tau_1)}{b_{2n}}\big)k\big(\frac{-\hat{\epsilon}_i(\tau_1, \tau_2)}{b_{2n}}\big) \xrightarrow{p} \Gamma_{22}(\tau_1, \tau_2).$$

Proof: See the appendix.

Existence of the joint conditional density of $\epsilon(\tau_1, \tau_2)$ and $R(\tau_1)$ given S requires that there is at least one cotinuous instrument with non–zero coefficient. This assumption is made for the conveience of the proof, but it seems possible to be relaxed as long as $R(\tau_1)$ has a marginal density. Condition (i) is standard in kernel estimation. Condition (ii) assumes smooth differentiable kernels, which is satisfied by most of commonly used kernels. When W_i does not have bounded support but the bandwidth choices b_{1n} and b_{2n} satisfy the stronger requirements, we do not need condition (ii). As $k(\cdot)$ has higher order derivatives, the requirements for b_{1n} and b_{2n} with unbounded support of W_i become closer to the requirements with bounded support of W_i. The requirement of $\sqrt{nh_n}b_{2n} \to \infty$ shows that estimation of $\Gamma(\tau_1, \tau_2)$ can be quite difficult in practice. Variance estimation of the

proposed estimators need further investigation.

5   Experiments using the Angrist and Krueger Data

In this section, I conduct small experiments with the Angrist-Krueger data (see e.g., Angrist and Krueger (1991), Angrist, Imbens, and Krueger (1999)). Returns to schooling is a leading example that shows how useful triangular models are. One of the most interesting features of triangular models is that it can provide a more complete picture about (random and heterogeneous) marginal effects by investigating two different quantile points. However, even when instruments are not relevant for all different distributional points, analysing a particular quantile of the endogenous variable is still possible; the method proposed in this paper requires that the instruments be relevant for a particular conditional quantile of the endogenous variable. This is a substantially weaker assumption than the traditional rank condition required in simultaneous equations models, where the instruments must affect the conditional mean of the endogenous variable. The Angrist-Krueger data is well-known to suffer from the issue of weak instruments while the sample size is extremely large; $n = 329,509$. The experiments of this section investigate the possibility of local relevance of those *weak* instruments.

One limitation of triangular models employed in this paper is that the assumption of strong monotonicity is required for point-identification (see e.g., Chesher (2003, 2005, 2007)). Since the education variable of the Angrist-Krueger data is observed as the years of schooling, it is a discrete variable, and it clearly does not satisfy the monotonicity assumption. However, since the purpose of this section is not in a rigorous empirical analysis but in investigating potential possibilities and providing an example, I simply pretend that the education variable is continuous by adding a small amount of random noise.[7]

I considered *contaminated* education variables which were obtained by adding a small amount of random noise generated from $N(0, \sigma^2)$ with $\sigma \in \{0.01, 0.02, 0.03, 0.04, 0.05\}$ to the original education variable of the Angsrist-Krueger data. I then ran quantile regression of the simulated education

---

[7]Florens, Heckman, Meghir, and Vytlacil (2008) also used the example of returns to education to motivate their analysis, where they also considered continuous endogenous treatments.

variables on 10 dummy variables indicating the years of birth and 30 instruments of the birth-quarters interacted with the birth-years.

Table I shows the p-values of the coeffcients of those 30 instruments for variaous quantiles. Interestingly, those coefficients are quite significant for most quantiles, but not for e.g., 85% quantile. I tried different values of the standard deviations of the random noises, but the coefficients of those instruments for the 85% conditional quantile were insignificant at 5% levels in most cases. In addition, although it is not reported in table I, the Wald statistic testing all the 30 instruments are irrelevant was largest for the 20% conditional quantile in all five cases.

The fact that the instruments are most relevant for $\tau_1 = 0.20$ needs some comments. Jun (2008) analyzed the Angrist–Krueger data within the framework of the instrumental quantile model (the IV quantile model hereafter) proposed by Chernozhukov and Hansen (2006). In particular, Jun (2008) cast the IV quantile model in a general GMM framework with a moment condition

$$E\Big([X'\ Z']'\big(1\{Y \leq D\alpha(\tau) + X'\beta(\tau)\} - \tau\big)\Big) = 0, \tag{11}$$

where Y is the log wage, D is the years of schooling, X is 10 dummies indicating the birth years, and Z contains the 30 instruments.[8] He then constructed 95% confidence intervals for $\alpha(\tau)$ with various choices of τ without assuming that they are identified. I here recall that his confidence intervals were extremely wide for small values of τ while those for upper quantiles were relatively tighter.[9] The fact that the instruments are weaker for lower quantiles in model (11) shows how different the IV quantile model is from the triangular model considered in this paper. The triangular model is a mutiple equations model, where different (conditional) quantiles of D and different (conditional) quantiles of Y are separately treated and they can be separately interpreted. However, the IV quantile model of Cherhozhukov and Hansen (2006) is a single GMM model, where quantiles of D are irrelevant but quantiles of (counterfactual) outcomes are compared (see Chernozhukov and

---

[8]Understanding the IV quantile model in the GMM framework is a common view. For example, Chernozhukov and Hong (2003) considered the same moment condition and they studied the Laplace–type estimators that are defined by using the quasi–Bayesian posteriors.

[9]For example, the interval for $\alpha(0.20)$ was roughly between 0.075 and 0.625 while the interval for $\alpha(0.85)$ was between 0.025 and 0.125.

Hansen (2006) for more precise interpretation). Note also that the IV quantile model (11) does not require that D be a continuous variable, because conditional quantiles of D are irrelevant for anaylses.

Figure I shows the point estimates of $\alpha(\tau_1, \tau_2)$ with $\sigma \in \{0.01, \ 0.02\}$. Although $\hat{\alpha}(\tau_1, \tau_2)$ would provide more complete information about the marginal effects if there were no issue of weak instruments, it is clear that it is not the case; since the first–stage Wald tests for $\tau_1 = 0.85$ do not even reject the null of irrelevant instruments, there is a clear problem due to weak instruments even in these localized estimates. Note however that some values of τ_1 still strongly reject the null of irrelevant instruments in the first stage. Figure I also illustrates the point estimates of $\alpha(0.20, \tau_2)$ and their 95% confidence intervals, where $\tau_1 = 0.20$ is such that the first–stage Wald statistic is largest among other values of τ_1.

Although these are experimental results based on *contaminated* education, they provide another possibility that is comparable to Chernozhukov and Hansen (2006) and Jun (2008). Although the IV quantile model does allow a discrete endogenous variable, it is a single GMM model, and it is not straightforward to assess instruments. The (triangular) model of Chesher (2003) does not allow a discrete endogenous variable, but it provide further localized inference and assessing instruments is more straightforward; in this paper, I considered a semiparametric implementation of Chesher (2003). Several methods have been proposed in the literature to make triangular models more practical. However, they are either fully parametric or separable in control variables (see e.g. Ma and Koenker (2006) and Lee (2007)). The method proposed in this paper provides a simple yet sufficiently flexible way of utilizing triagular models.

6   Simulations: Local Instruments and Identification Pitfalls

In this section, I consider three different data generating processes for Monte Carlo experiments. In the followings, $N^{-1}(s : \mu, \sigma^2)$ denotes the inverse of the distribution function of $N(\mu, \sigma^2)$. The included exogenous variables, X_1 and X_2 are independently generated from $N(0, 1)$ and $N(2, 4)$, respectively for each design. $(\beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3)$ is set to be $(1, 1, 1, 1, 5)$. The parameter of interest

is $\alpha(\tau_1, \tau_2)$, the marginal effect of D on Y for those whose ϵ and V are at τ_2 and τ_1, respectively. The true value for each design is given by $\alpha(\tau_1, \tau_2) = 1 - 3\tau_1 + \tau_1 + \exp(\tau_1)$. In the following experiments, $\tau_2 = 0.5$ is used as the various values of τ_1 are tried.

DGP1 $\begin{cases} Y = D(1 - 3V + V^2 + \exp(V)) + X_1\beta_1 + X_2\beta_2 + N^{-1}(\epsilon : 0, 1) \\ \\ D = Z_1\pi_1 + Z_2\pi_2 + X_1\gamma_1 + X_2\gamma_2 + \gamma_3 + N^{-1}(V : 0, 1), \end{cases}$

where $Z_1 \sim N(0, 1)$, $Z_2 \sim N(0, 1)$ and $(\pi_1, \pi_2) = (1, 1)$.

DGP2 $\begin{cases} Y = D(1 - 3V + V^2 + \exp(V)) + X_1\beta_1 + X_2\beta2 + N^{-1}(\epsilon : 0, 1) \\ \\ D = Z_1\pi + Z_2 N^{-1}(V : 1, 4) + X_1\gamma_1 + X_2\gamma_2 + \gamma_3 + N^{-1}(V : 0, 1), \end{cases}$

where $Z_1 \sim N(0, 1)$, $Z_2 \sim |N(0, 1)|$ and $\pi = 0.01$.

DGP3 $\begin{cases} Y = D(1 - 3V + V^2 + \exp(V)) + X_1\beta_1 + X_2\beta_2 + N^{-1}(\epsilon : 0, 1) \\ \\ D = Z_3 N^{-1}(V : 1, 4) + X_1\gamma_1 + X_2\gamma_2 + \gamma_3 + N^{-1}(V : 0, 1), \end{cases}$

where $Z_3 \sim Bernoulli(\frac{1}{2})$.

DGP1 is the case where two instruments are globally valid with homosckedastistic independent error. DGP2 has two instruments; one is *weak*, and the other is *local*. Since $N^{-1}(0.3 : 1, 4) \approx 0.048$, DGP2 suffers from *weak* identification when τ_1 is chosen to be around 30%. However, when τ_1 is far from 30%, Z_2 is a relevant instrument, and the parameter of interest is identified. Lastly, DGP3 is the case where there is one binary instrument that can be weak for a particular quantile.

For the sample size, I used $n \in \{3,000, \ 5,000, \ 8,000\}$. For the bandwidth, $h_n = IQR \cdot n^{-\frac{1}{3}}$ was used except for Figures II-1 and II-2, where IQR is the interquantile range of the first–stage quantile regression residuals. For Monte Carlo, $1,000$ replications were made.

Tables II-1 through II-3 show the Monte Carlo results from DGP1, DGP2, and DGP3, respectively. For DGP1, RMSEs are small for all the quantile effects. DGP2 and DGP3 are more interesting. Since the instruments are only local in the sense that they are relevant for some quantiles but not for others, RMSEs remain big when τ_1 is close to 30%. In spite of the identification difficulty around $\tau_1 = 30\%$, the instruments are still informative for other quantile effects and RMSEs are reasonably small.

Figures II-1 and II-2 show the sensitivity of RMSEs to different bandwidth choices. Using $n = 5,000$ and $h_n = \kappa \cdot n^{-1/3}$, $\kappa \in \{0.5, 1.0, 1.5, \cdots, 4.0, 4.5, 5.0, 8.0, 10.0, 15.0\}$ were considered, which led to $h_n \in [0.03, \ 0.88]$. As the figures show, RMSEs are quite stable in all cases within wide

ranges of bandwidth.

Figure III illustrates the rejection rates of the t-statistics testing the truth with the nominal size 5% and $n = 8000$. To estimate the standard errors, the crude kernel estimators suggested in proposition 2 were used using the standard normal density as a kernel. In contrast to the other quantile effects, the rejection rates around $\tau_1 = 30\%$ in DGP2, DGP3 are far from 5%.

The size distortions of DGP2 and DGP3 around $\tau_1 = 30\%$ are not surprising in view of the identification difficulty: the instruments are not informative for the quantile effects around $\tau_1 = 30\%$. This is interesting because it suggests that individual heterogeneity can cause instruments to be only locally relevant and yet the local instruments can still be exploited by focusing on a particular local effect. To be more concrete, the densities of the t-statistics are estimated from Monte Carlo and some of them are shown in Figure IV. Not surprisingly, when $\tau_1 = 30\%$ in DGP2 and DGP3, the densities of the t-statistics are far from the normal density.

Lastly, Figure V shows examples of 95% confidence intervals computed by the normal approximation. I simply generated three sets of artificial data from DGP1, DGP2, and DGP3. Then, 95% condidence intervals were computed for various values of τ_1 with fixing $\tau_2 = 0.5$. Again, the confidence bands from DGP2, DGP3 do not cover the truth when τ_1 is around 30%. However, they do cover the truth when τ_1 is far from 30%.

7   Concluding Remarks

I proposed a two-step estimator in a nonseparable triangular system. In order to obtain a practical estimator while preserving the flexibility of the triangular system, I considered a linear-in-variables specification that has a control variable included in a nonparametric and nonseparable manner. This is interesting because (1) in contrast to Buchinsky (1994), Lee (2007), Chernozhukov and Hansen (2005, 2006) and Jun (2008), the marginal effects are allowed to be heterogeneous over two different quantile points, and (2) inference can be focused on a particular quantile of the endogenous variable when the instruments are relevant for the particular quantile of the endegenous variable (see also Ma and Koenker (2006)).

The integrated quantile effect (or the average quantile effect: $\int_0^1 \alpha(v, \tau_2)dv$) looks interesting too. It can be used as a summary of all the local parameters, and it seems to be possible to recover the regular $\sqrt{n}$ convergence rate.[10] However, when the instruments are "weak" for a particular quantile, the inference on the average quantile effect can be quite misleading. Instead, the method proposed in this paper focuses on a particular quantile of the endogenous variable at the expense of the parametric $\sqrt{n}$ rate. Although there are many methods to estimate certain average effects without estimating the entire local function, they are based on a set of alternative assumptions. For example, two–stage least squares estimators requires that the instruments affect the conditional mean of the endogenous variable, which is quite different from the local quantile relevance condition. See also Florens, Heckman, Meghir, and Vytlacil (2008). As the experiments with the Angrist and Krueger data show, instruments can be locally relevant for some quantiles without being so strong in the conventional mean sense. Therefore, the proposed method can also be considered an alternative approach to the weak instrument problem.

There are several limitations. First, the monotonicity assumption requires that the endogenous variable be a continuous random variable. This can be restrictive in practice. Chesher (2005) discusses a set identification result with a discrete endogenous variable, which is worth further study. Second, the estimation of the variance can be tough in practice. Comparing several methods of variance estimation such as kernel, k-nearest-neighbor, and bootstrap is worth studying. Testing the location of the local parameter without assuming its identification is another interesting question. It is also left for the future research.

---

[10]In standard linear quantile regression models, integrating regression quantiles to obtain a location parameter has been considered by e.g. Portnoy and Koenker (1989).

References

Angrist, J. and A. Krueger, 1991, "Does Compulsory Schooling Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106, 979–1014.

Angrist, J., G. Imbens, and A. Krueger, 1999, "Jacknife Instrumental Variables Estimation," *Journal of Applied Econometrics* 14, 57–67.

Buchinsky, M., 1994, "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62, 405–458.

Buchinsky, M., 1998, "Recent Advances in Qunatile Regression Models: A Practical Quideline for Empirical Research," *Journal of Human Resources* 33, 88–126.

Chernozhukov, V., and C. Hansen, 2005, "An IV Model of Quantile Treatment Effects," *Econometrica* 73, 245–261.

Chernozhukov, V., and C. Hansen, 2006, "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics* 132, 491–525.

Chernozhukov, V., and H. Hong, 2003, "An MCMC Approach to Classical Estimation," *Journal of Econometrics* 115, 293–346.

Chesher, A., 2003, "Identification in Nonseparable Models," *Econometrica* 71, 1405–1441.

Chesher, A., 2005, "Nonparametric Identification under Discrete Variation," *Econometrica* 73, 1525–1550.

Chesher, A., 2007, "Identification of Nonadditive Structural Functions," in: *Advances in Economics and Econometrics Volume III*, Cambridge University Press.

Davidson, J. 1994, *Stochastic Limit Theory*, Oxford University Press.

Florens, J.P., J.J. Heckman, C. Meghir, and E. Vytlacil, 2008, "Identification of Treatments Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica* Forthcoming.

Imbens, G. and W. Newey, 2002, "Identification and Estimation of Triangular Simultaneous Equation Models without Additivity," *NBER Working Paper*.

Jun, S., 2008, "Weak Identification Robust Tests in an Instrumental Quantile Model," *Journal of Econometrics* 144, 118–138.

Koenker, R., 2005, *Quantile Regression*, Econometric Society Monographs No. 38, Cambridge University Press.

Koenker, R. and G. Basset, 1978, Regression Quantiles, *Econometrica* 46, 33–50.

Lee, Sokbae, 2007, "Endogeneity in Quantile Regression Models: A Control Function Approach," *Journal of Econometrics* 141, 1131–1158.

Ma, L. and R. Koenker, 2006, "Quantile Regression Methods for Recursive Structural Equation Models," *Journal of Econometrics* 134, 471–506.

Newey, W. and D. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing," in: *Handbook of Econometrics*, Vol.4, Ch.36, pp. 2113–2148.

Pagan, A. and A. Ullah, 1999, *Nonparametric Econometrics*, Cambridge University Press.

Pakes, A. and D. Pollard, 1989, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica* 57, 1027–1057.

Portnoy, S. and R. Koenker, 1989, "Adaptive L–Estimation for Linear Models," *Annals of Statistics* 17, 362–381.

Powell, J., 1986, "Censored Regression Quantiles," *Journal of Econometrics* 32, 143–155.

van der Vaart, A.W. and J.A. Wellner, 1996, *Weak Convergence and Empirical Processes*, Springer-Verlag New York, Inc..

van der Vaart, A.W., 1998, *Asymptotic Statistics*, Cambridge University Press.

A   Proof of Proposition 1

Let $\Pi_0 = [\Pi_1(\tau_1)' : \Pi_2(\tau_1)']'$. First, note that $Q_{D|W}(\tau_1) = W'\Pi_0$. Multiplying W to both sides and taking expectations yields $E(WW')\Pi_0 = E(WQ_{D|W}(\tau_1))$. Therefore, Π_0 is identified by $E(WW')^{-1}E(WQ_{D|W}(\tau_1))$. Now, note that

$$\Pr(Y \le Q_{D|W}(\tau_1)\alpha(\tau_1, \tau_2) + X'\beta(\tau_1, \tau_2)|D = Q_{D|W}(\tau_1), W)$$
$$= \Pr(Y \le D\alpha(\tau_1, \tau_2) + X'\beta(\tau_1, \tau_2)|D = Q_{D|W}(\tau_1), W)$$
$$= \Pr(Y \le D\alpha(V, \tau_2) + X'\beta(V, \tau_2)|V = \tau_1, W) = \Pr(\epsilon \le \tau_2|V = \tau_1, W) = \Pr(\epsilon \le \tau_2) = \tau_2,$$

because $D = Q_{D|W=w}(\tau_1), W = w$ is equivalent to $V = \tau_1, W = w$ due to the strong monotonicity.[11] Note here that independence between V and ϵ is used only for the last equality, which local independence is in fact sufficient for. Now, note that

$$Q_{Y|D=Q_{D|W}(\tau_1),W}(\tau_2) = Q_{D|W}(\tau_1)\alpha(\tau_1, \tau_2) + X'\beta(\tau_1, \tau_2)$$
$$= X'(\Pi_1(\tau_1)\alpha(\tau_1, \tau_2) + \beta(\tau_1, \tau_2)) + Z'\Pi_2(\tau_1)\alpha(\tau_1, \tau_2).$$

Multiplying W to both sides and taking expectations shows that
$[\Pi_1(\tau_1)'\alpha(\tau_1, \tau_2) + \beta(\tau_1, \tau_2)', \ \Pi_2(\tau_1)'\alpha(\tau_1, \tau_2)]'$ is identified by $E(WW')^{-1}E(WQ_{Y|D=Q_{D|W}(\tau_1),W}(\tau_2))$.
Therefore, **A3** results in identification of $\alpha(\tau_1, \tau_2)$ and $\beta(\tau_1, \tau_2)$. $\square$

B   The Moment Condition and Global Identification

Since $E(\rho_\tau(D_i - W_i'\Pi))$ is weakly convex, solutions to $E(W_i(1\{D_i \le W_i'\Pi\} - \tau_1)) = 0$ are minimizers of $E(\rho_{\tau_1}(D_i - W_i'\Pi))$. Therefore, in order to show that Π_0 uniquely solves $E(W_i(1\{D_i \le W_i'\Pi\} - \tau_1)) = 0$, it suffices to show that Π_0 is the unique minimizer. Since $E(\rho_\tau(D_i - W_i'\Pi))$ is weakly convex, its minimum is unique when its Hessian at the minimum is positive definite. Now, note that the Hessian at Π_0 is given by $E(W_iW_i'f_{R(\tau_1)|W}(0|W_i))$, which is clearly positive definite when

---

[11] For the sake of simplicity, the qualifier of "almost surely" will be suppressed throughout the paper when it is clear from the context.

$f_{R(\tau_1)|W}(0|W_i)$ is bounded away from 0.

Given $R(\tau_1) = 0$, D is only a function of W. Therefore, θ_0 will solve

$$E\big(S_i(1\{Y_i \leq S_i'\theta\} - \tau_2)|R_i(\tau_1) = 0\big) = 0. \tag{12}$$

Following the same reasoning as before, θ_0 will be the unique solution, if it uniquely minimizes $E\big(\rho_{\tau_2}(Y_i - S_i'\theta)|R_i(\tau_1) = 0\big)$. Now, note that its Hessian at θ_0 will be given by $E\big(S_i S_i' f_{\epsilon(\tau_1,\tau_2)}(0|W_i, R_i(\tau_1) = 0)|R_i(\tau_1) = 0\big)$. Therefore, if the conditional density of $\epsilon(\tau_1, \tau_2)$ at 0 given $W_i, R_i(\tau_1) = 0$ is bounded away from 0, then θ_0 will be the unique solution to equation (12).

□

Remark: Note that the moment condition (12) can be expressed in a more common but equivalent form when $R(\tau_1)$ has a density given Y and S; this is the case when we have continuous instruments with non–zero coefficient at τ_1. Note that

$$E\big(S_i(1\{Y_i \leq S_i'\theta\} - \tau_2)|R_i(\tau_1) = 0\big)$$
$$= \int s(1\{y \leq s'\theta\} - \tau_2) f_{Y,S|R(\tau_1)}(y, s|0) d\mu(y, s) = \frac{1}{f_{R(\tau_1)}(0)} \int s(1\{y \leq s'\theta\} - \tau_2) f_{Y,S,R(\tau_1)}(y, s, 0) d\mu(y, s)$$
$$= \frac{1}{f_{R(\tau_1)}(0)} \int s(1\{y \leq s'\theta\} - \tau_2) f_{R(\tau_1)|Y,S}(0|y, s) f_{Y,S}(y, s) d\mu(y, s)$$
$$= \frac{1}{f_{R(\tau_1)}(0)} E\big(S_i(1\{Y_i \leq S_i'\theta\} - \tau_2) f_{R(\tau_1)|Y,S}(0|Y_i, S_i)\big),$$

where μ is the underlying measure.

C   Proof of Lemma 1

Note that

$$\Upsilon_{n,h_n}\Big(E\big(M_n(\Pi, \theta)\big) - M(\Pi, \theta)\Big) = \begin{bmatrix} 0 \\ B_n(\Pi, \theta) \end{bmatrix},$$

where

$$B_n(\Pi, \theta) = \sqrt{nh_n}\Big(\frac{1}{h_n}E\big(S_i(1\{Y_i \leq S_i\theta\} - \tau_2)1\{|D_i - W_i\Pi| \leq \frac{h_n}{2}\}\big)$$
$$- E\big(S_i(1\{Y_i \leq S_i\theta\} - \tau_2)|R_i(\tau_1) - W_i'\delta = 0\big)f_{R_i(\tau_1)-W'\delta}(0)\Big)$$

for $\delta = \Pi - \Pi_0 \in \mathcal{N}$. Let $\varphi(\theta, \delta, r) = E\big(S_i(1\{Y_i \leq S_i\theta\} - \tau_2)|R_i(\tau_1) - W_i'\delta = r\big)$, and it follows from the law of iterated expectations and the change–of–variables that

$$B_n(\Pi, \theta) = \sqrt{nh_n}\int_{-1/2}^{1/2}\varphi(\theta, \delta, th_n)f_{R(\tau_1)-W'\delta}(th_n) - \varphi(\theta, \delta, 0)f_{R(\tau_1)-W'\delta}(0)dt$$
$$= \frac{\sqrt{nh_n}h_n^2}{2}\int_{-1/2}^{1/2}\frac{\partial^2\varphi(\theta, \delta, s)f_{R(\tau_1)-W'\delta}(s)}{\partial s^2}\Big|_{s=\bar{s}}t^2dt,$$

where $\bar{s}$ is between 0 and th_n. Therefore, for every $\theta \in \Theta$ and $\Pi \in \Pi_0 + \mathcal{N}$,

$$||B_n(\Pi, \theta)|| \leq \sqrt{nh_n}h_n^2\Big(\sup_{-1/2\leq r\leq 1/2}||\frac{\partial^2\varphi(\theta, \delta, r)}{\partial r^2}||b_0(r) + 2\sup_{-1/2\leq r\leq 1/2}||\frac{\partial\varphi(\theta, \delta, r)}{\partial r}||b_1(r)$$
$$+ \sup_{-1/2\leq r\leq 1/2}||\varphi(\theta, \delta, r)||b_2(r)\Big)\int_{-1/2}^{1/2}t^2dt. \quad (13)$$

It then follows that $\sup_{\Pi\in\Pi_0+\mathcal{N},\ \theta\in\Theta}||B_n(\Pi, \theta)|| = O(\sqrt{nh_n}h_n^2)$ from compactness of $\Theta \times \mathcal{N}$ and continuity of the right–hand side of (13). $\square$

D   Proof of Theorem 1

In this section, I provide the proof of theorem 1. For the sake of convenience, I splitted the theorem into the consistency part (lemmas 2 and 8) and the normality part (proposition 3). In the following discussion, I will write $M_1(\Pi)$ and $M_2(\Pi, \theta)$ to denote the partitions of $M(\Pi, \theta)$; their sample analogs will be denoted by $M_{1n}(\Pi)$ and $M_{2n}(\Pi, \theta)$, respectively.

Lemma 2 $\sqrt{n}||\hat{\Pi}(\tau_1) - \Pi(\tau_1)|| = O_p(1)$.

Proof: It follows from the fact that $\hat{\Pi}(\tau_1)$ is the standard linear quantile regression estimator. $\square$

Lemma 3

$$\frac{1}{nh_n}\sum_{i=1}^{n}|1\{|D_i - W_i\hat{\Pi}| \leq \frac{h_n}{2}\} - 1\{|D_i - W_i\Pi_0| \leq \frac{h_n}{2}\}|\,||S_i|| = o_p(1).$$

Proof: Note that

$$|1\{|R_i(\tau_1) - W_i'(\hat{\Pi} - \Pi_0)| \leq \frac{h_n}{2}\} - 1\{|R_i(\tau_1)| \leq \frac{h_n}{2}\}|$$
$$\leq 1\{-\frac{h_n}{2} \leq R_i(\tau_1) \leq ||W_i||\,||\hat{\Pi} - \Pi_0|| - \frac{h_n}{2}\} + 1\{\frac{h_n}{2} \leq R_i(\tau_1) \leq ||W_i||\,||\hat{\Pi} - \Pi_0|| + \frac{h_n}{2}\}$$
$$+ 1\{-\frac{h_n}{2} - ||W_i||\,||\hat{\Pi} - \Pi_0|| \leq R_i(\tau_1) \leq -\frac{h_n}{2}\} + 1\{\frac{h_n}{2} - ||W_i||\,||\hat{\Pi} - \Pi_0|| \leq R_i(\tau_1) \leq \frac{h_n}{2}\}$$

for sufficiently large n, because $|W_i'(\hat{\Pi} - \Pi_0)| = o_p(h_n)$. Since the four terms are similar, I only consider the last one. Let δ_n be a (positive) sequence such that $||\hat{\Pi} - \Pi_0|| = o_p(\delta_n)$ and $\delta_n = o(h_n)$. It then follows that $|W_i|\,||\hat{\Pi} - \Pi_0|| \leq ||W_i||\delta_n$ with probability approaching to 1. I will show that

$$E\big(\frac{1}{h_n}1\{\frac{h_n}{2} - ||W_i||\delta_n \leq R_i(\tau_1) \leq \frac{h_n}{2}\}||S_i||\big) = o(1).$$

Note first that

$$E\big(\frac{1}{h_n}1\{\frac{h_n}{2} - ||W_i||\delta_n \leq R_i(\tau_1) \leq \frac{h_n}{2}\}||S_i||\big)$$
$$\leq E\big(\frac{1}{h_n}1\{\frac{h_n}{2} - ||W_i||\delta_n \leq R_i(\tau_1) \leq \frac{h_n}{2}\}|D_i|\big) + E\big(\frac{1}{h_n}1\{\frac{h_n}{2} - ||W_i||\delta_n \leq R_i(\tau_1) \leq \frac{h_n}{2}\}||X_i||\big).$$

Since the second term is similar, I only consider the first term. Note that

$$E\big(\frac{1}{h_n}1\{\frac{h_n}{2} - ||W_i||\delta_n \leq R_i(\tau_1) \leq \frac{h_n}{2}\}|D_i|\,|W_i\big) = \int_{\frac{1}{2}-\frac{\delta_n}{h_n}}^{\frac{1}{2}}|th_n + W_i'\Pi_0|f_{R(\tau_1)|W}(th_n|W_i)dt$$
$$\leq h_n\int_{\frac{1}{2}-\frac{\delta_n}{h_n}}^{\frac{1}{2}}|t|f_{R(\tau_1)|W}(th_n|W_i)dt + |W_i'\Pi_0|\int_{\frac{1}{2}-\frac{\delta_n}{h_n}}^{\frac{1}{2}}f_{R(\tau_1)|W}(th_n|W_i)dt \to 0$$

with probability one. Applying dominated convergence theorem completes the proof. □

Lemma 4 *Let $\tilde{m}_n(\Pi, \theta) = c'S(1\{Y \leq S'\theta\} - \tau_2)1\{|D - W'\Pi| \leq \frac{h_n}{2}\}\frac{1}{\sqrt{h_n}}$, where c is an arbitrary*

conformable vector and $(\Pi, \theta) \in (\Pi_0 + \mathcal{N}) \times \Theta$. *Then, for any sequence* $\delta_{1n} = o(h_n)$ *and* $\delta_{2n} = o(1)$,

$$\sup_{||\Pi - \tilde{\Pi}|| < \delta_{1n}, ||\theta - \tilde{\theta}|| < \delta_{2n}} E\big((\tilde{m}_n(\Pi, \theta) - \tilde{m}_n(\tilde{\Pi}, \tilde{\theta}))^2\big) = o(1).$$

Proof: Note that

$$
\begin{aligned}
(\tilde{m}_n(\Pi, \theta) &- \tilde{m}_n(\tilde{\Pi}, \tilde{\theta}))^2 \\
&\leq 2\Big((c'S)^2 \big|1\{|D - W'\Pi| \leq \tfrac{h_n}{2}\} - 1\{|D - W'\tilde{\Pi}| \leq \tfrac{h_n}{2}\}\big|\tfrac{1}{h_n} \\
&\qquad + (c'S)^2 \big|1\{Y \leq S'\theta\} - 1\{Y \leq S'\tilde{\theta}\}\big|1\{|D - W'\tilde{\Pi}| \leq \tfrac{h_n}{2}\}\tfrac{1}{h_n}\Big). \quad (14)
\end{aligned}
$$

Since $|W'(\Pi - \tilde{\Pi})| = o_p(h_n)$, the first term in the right–hand side of (14) is bounded by

$$
\begin{aligned}
(c'S)^2 \Big(&1\{W'\Pi - \tfrac{h_n}{2} \leq D \leq W'\tilde{\Pi} - \tfrac{h_n}{2}\} + 1\{W'\Pi + \tfrac{h_n}{2} \leq D \leq W'\tilde{\Pi} + \tfrac{h_n}{2}\} \\
&+ 1\{W'\tilde{\Pi} - \tfrac{h_n}{2} \leq D \leq W'\Pi - \tfrac{h_n}{2}\} + 1\{W'\tilde{\Pi} + \tfrac{h_n}{2} \leq D \leq W'\Pi + \tfrac{h_n}{2}\}\Big)\tfrac{1}{h_n}.
\end{aligned}
$$

Since all four terms are similar, I only consider the last one. Taking the expectation of the last term yields

$$
\begin{aligned}
E\Big((c_1 D + c_2'X)^2 &1\{W'\tilde{\Pi} + \tfrac{h_n}{2} \leq D \leq W'\Pi + \tfrac{h_n}{2}\}\tfrac{1}{h_n}\Big) \\
&= E\Big(\int_{W'\tilde{\Pi}/h_n + \frac{1}{2}}^{W'\Pi/h_n + \frac{1}{2}} (c_1 th_n + c_2'X)^2 f_{D|W}(th_n|W)dt\Big) \leq C(\Pi, \tilde{\Pi})\frac{||\Pi - \tilde{\Pi}||}{h_n},
\end{aligned}
$$

where $C(\Pi, \tilde{\Pi}) = E\Big(\sup(c_1 s + c_2'X)^2 f_{D|W}(s|W)||W||\Big)$ with sup taken over s between $W'\tilde{\Pi} + \frac{1}{2}$ and $W'\Pi + \frac{1}{2}$. Since $\sup_{(\Pi, \tilde{\Pi}) \in \mathcal{N} \times \mathcal{N}} C(\Pi, \tilde{\Pi}) < \infty$, it follows that

$$\sup_{||\Pi - \tilde{\Pi}|| < \delta_{1n}} C(\Pi, \tilde{\Pi})\frac{||\Pi - \tilde{\Pi}||}{h_n} = O(\frac{\delta_{1n}}{h_n}) = o(1).$$

The second term of the right–hand side of (14) is similar and omitted. □

Lemma 5 *For any sequence $h_n > 0$ with $h_n \downarrow 0, nh_n \to \infty$,*

(i) $\sup_{\theta \in \Theta} ||M_{2n}(\hat{\Pi}, \theta) - M_2(\Pi_0, \theta)|| = o_p(1)$ as $n \to \infty$.

Moreover, if we let let $r_n = \sqrt{n}h_n^{\delta}$ such that $\delta > 0$ and $\sqrt{n}h_n^{1+\delta} \to \infty$, then for any sequences $\delta_{1n} \downarrow 0$ and $\delta_{2n} \downarrow 0$ such that $r_n \delta_{1n} \downarrow 0$, we have

(ii) $\sup ||\sqrt{nh_n}\Big(M_{2n}(\Pi, \theta) - E\big(M_{2n}(\Pi, \theta)\big)\Big) - \sqrt{nh_n}\Big(M_{2n}(\Pi_0, \theta_0) - E\big(M_{2n}(\Pi_0, \theta_0)\big)\Big)|| = o_p(1)$

as $n \to \infty$, where $\sup$ is taken over $||\Pi - \Pi_0|| \leq \delta_{1n}$ and $||\theta - \theta_0|| \leq \delta_{2n}$.

Proof: Note first that

$$\sup_{\theta} ||M_{2n}(\hat{\Pi}, \theta) - M_{2n}(\Pi_0, \theta)||$$
$$\leq \frac{1}{nh_n} \sum_{i=1}^{n} ||S_i|||1\{|D_i - W_i'\hat{\Pi}| \leq \frac{h_n}{2}\} - 1\{|D_i - W_i'\Pi_0| \leq \frac{h_n}{2}\}| = o_p(1)$$

by lemma 3. For part (i), I note that

$$\sup_{\theta} ||M_{2n}(\Pi_0, \theta) - M_2(\Pi_0, \theta)|| \leq \sup_{(\pi, \theta) \in \mathcal{N} \times \Theta} ||M_{2n}(\Pi_0 + \frac{\pi}{r_n}, \theta) - M_2(\Pi_0 + \frac{\pi}{r_n}, \theta)||$$

and I will show that the right–hand side is $o_p(1)$ in the following discussion.

Consider the following class of functions

$$\mathcal{F}_n = \{m_n(\pi, \theta) = c'S(1\{Y \leq S'\theta\} - \tau_2)1\{|R(\tau_1) - W'\frac{\pi}{r_n}| \leq \frac{h_n}{2}\}\frac{1}{\sqrt{h_n}} : (Y, D, X, Z) \times (\pi, \theta) \mapsto \mathbb{R}\},$$

where c is an arbitrary conformable vector and $(\pi, \theta) \in \mathcal{N} \times \Theta$. Note that working with the local parameter π is sufficient, because $\hat{\Pi}$ is already $\sqrt{n}$–consistent for Π_0.

Let $\mu_n(\pi, \theta) = \frac{1}{nh_n} \sum_i^n c'S_i(1\{Y_i \leq S_i'\theta\} - \tau_2)1\{|R_i(\tau_1) - W_i'\frac{\pi}{r_n}| \leq \frac{h_n}{2}\}$ and consider

$$\mathbb{G}_n m_n = \sqrt{n}(\mathbb{P}_n - P)m_n = \sqrt{nh_n}\Big(\mu_n(\pi, \theta) - E\big(\mu_n(\pi, \theta)\big)\Big),$$

which is an empirical process with a sequential class of functions (see e.g., van der Vaart and Wellner (1996, p220-221), van der Vaart (1998, p282)). In particular, corollary 19.35 and theorem

19.28 of van der Vaart (1998) provide sufficient conditions for uniform convergence and stochastic equicontinuity of $\mathcal{F}_n$. Define $\pi_n^*(D, W) = \arg\min_{\pi \in \mathcal{N}} |R(\tau_1) - W'\frac{\pi}{r_n}|$, and $\mathcal{F}_n$ has an envelope function $F_n = 2c'S1\{|R(\tau_1) - W'\frac{\pi_n^*(D,W)}{r_n}| \leq \frac{h_n}{2}\}\frac{1}{\sqrt{h_n}}$. Since $R(\tau_1) - W'\frac{\pi_n^*(D,W)}{r_n}$ has a density, the standard change-of-variable technique shows

$$E(F_n^2) = O(1)$$
$$E(F_n^2 1\{F_n > \epsilon\sqrt{n}\}) = o(1) \text{ for every } \epsilon > 0,$$

where the second equality is because $nh_n \to \infty$. Therefore, in view of lemma 4[12], the uniform entropy conditions of theorem 19.28 of van der Vaart (1998) will imply stochastic equicontinuity of $\mathbb{G}_n m_n$ and the uniform convergence of part (i) will follow from the maximal inequality.

To be more specific, let $J(\delta, \mathcal{F}_n, L_2)$ be the uniform entropy integral of $\mathcal{F}_n$. Since corollary 19.35 of van der Vaart (1998)[13] shows that

$$E(\sup_{\pi,\theta} |\mathbb{G}_n m_n|) \leq CJ(||F_n||_2, \mathcal{F}_n, L_2)$$

for some constant C, $J(||F_n||_2, \mathcal{F}_n, L_2) = O(1)$ will imply that $\sup_{(\pi,\theta)} |\mathbb{G}_n m_n| = o_p(\sqrt{nh_n})$, which will prove part (i) by lemma 1. Also, by theorem 19.28 of van der Vaart (1998), $J(\delta_n, \mathcal{F}_n, L_2) = o(1)$ for every $\delta_n \downarrow 0$ will show stochastic equicontinuity of $\mathbb{G}_n m_n(\pi, \theta)$ so that

$$\sup_{||(\pi,\theta)-(0,\theta_0)|| \leq \delta_n} |\mathbb{G}_n m_n(\pi, \theta) - \mathbb{G}_n m_n(0, \theta_0)| = o_p(1)$$

for any $\delta_n \downarrow 0$. It will then prove part (ii), because setting $r_n(\Pi - \Pi_0) = \pi$ shows that the supremum over $||\Pi - \Pi_0|| \leq \delta_{1n}$ is bounded by the supremum over $||\pi|| \leq r_n\delta_{1n}$.

Therefore, showing $J(||F_n||_2, \mathcal{F}_n, L_2) = O(1)$ and $J(\delta_n, \mathcal{F}_n, L_2) = o(1)$ for every $\delta_n \downarrow 0$ will complete the proof of the lemma. By definition of the uniform entropy integral, these two conditions are satisfied when the ϵ–uniform covering number $\sup_Q N(\epsilon||F_n||_2, \mathcal{F}_n, L_2(Q))$ is bounded by a

---

[12]Note that $r_n h_n \to \infty$.

[13]See also the comments on p289 of van der Vaart (1998). It is worth noting that the maximal inequality is not an asymptotic one.

polynomial of $\frac{1}{\epsilon}$ that does not depend on n.[14] A sufficient condition for this is that $\mathcal{F}_n$ is a Vapnik–Cervonenkis (VC) class with its VC index not depending on n (see e.g. theorem 2.6.7 of van der Vaart and Wellner (1996, hereafter VW)).

In the following four steps, I will show that the VC index of $\mathcal{F}_n$ is finite and that it does not depend on n. VC indices will be denoted by $\mathcal{V}(\cdot)$. First, I state a few useful properties of VC indices, which are used in the following steps.

VC Properties: Let $\mathcal{F}$ and $\mathcal{G}$ be VC classes of sets. Let $\mathcal{H}$ be a VC class of functions, and let g be an arbitraray fixed function. Then,

(i) $\mathcal{I} = \{1_A(x) : A \in \mathcal{F}\}$ is a VC class of functions with $\mathcal{V}(\mathcal{I}) = \mathcal{V}(\mathcal{F})$.

(ii) $\mathcal{F}^c = \{A^c : A \in \mathcal{F}\}$ is a VC class of sets with $\mathcal{V}(\mathcal{F}^c) = \mathcal{V}(\mathcal{F})$.

(iii) $\mathcal{F} \sqcap \mathcal{G} = \{A \cap B : A \in \mathcal{F}, B \in \mathcal{G}\}$ is a VC class of sets with $\mathcal{V}(\mathcal{F} \sqcap \mathcal{G}) \leq \mathcal{V}(\mathcal{F}) + \mathcal{V}(\mathcal{G}) - 1$.

(iv) The collection of all half–spaces in $\mathbb{R}^d$ is a VC class of sets of index $d + 2$. In particular, $\{1\{x'b \leq c\} : b \in \mathbb{R}^d, \ c \in \mathbb{R}\}$ is a VC class of functions of index $d + 2$.

(v) $\mathcal{H} \cdot g = \{f \cdot g : f \in \mathcal{H}\}$ is a VC class of functions with $\mathcal{V}(\mathcal{H} \cdot g) \leq 2\mathcal{V}(\mathcal{H}) - 1$.

These VC properties are now standard in empirical process theory, and their proofs can be found in many places. See e.g. lemmas 2.6.17 and 2.6.18 of VW, and exercise 14 on page 152 of VW. Using these properties, the following four steps prove that the VC index of $\mathcal{F}_n$ is finite and that it does not depend on n.

Step 1:   Let $\mathcal{C} = \{f_{A,B}(x) = 1_{A \cap B}(x) - \tau_1 1_B(x) : A \in \mathcal{F}, B \in \mathcal{G}\}$, where $\mathcal{F}$ and $\mathcal{G}$ are VC classes of sets. Then, $\mathcal{V}(\mathcal{C}) \leq 2\big(\mathcal{V}(\mathcal{F}) + 2\mathcal{V}(\mathcal{G}) - 2\big)$.

To prove this, suppose that a collection $\{(x_1, t_1), (x_2, t_2), ..., (x_k, t_k)\}$ is shattered by the collection of subgraphs of $\mathcal{C}$. Then, there should not be any t_j less than $-\tau_1$ or larger than $1 - \tau_1$. Moreover, this collection can be partitioned into two groups. Those points whose t_j is between $1 - \tau_1$ and 0 will be shattered by the collection of subgraphs of $\mathcal{C}_1 = \{(1 - \tau_1)1_{A \cap B}(x) : A \in \mathcal{F}, B \in \mathcal{G}\}$. Similarly, those points whose t_j is between 0 and $-\tau_1$ will be shattered by the collection of subgraphs of $\mathcal{C}_2 = \{-\tau_1 1_{B \cap (A \cap B)^c}(x) : A \in \mathcal{F}, B \in \mathcal{G}\}$. It then follows that $\mathcal{V}(\mathcal{C}) \leq \mathcal{V}(\mathcal{C}_1) + \mathcal{V}(\mathcal{C}_2)$. Now, the conclusion of step 1 follows from the VC properties (i), (iii), and (v) above.

---

[14]Since the ϵ–uniform covering number cannot increase as ϵ becomes larger, it is in fact sufficient to consider $0 < \epsilon < 1$.

Step 2: Let $\mathcal{D}_n = \{1\{|R(\tau_1) - W'\frac{\pi}{r_n}| \leq \frac{h_n}{2}\} : \pi \in \mathbb{R}^{k_x + k_z}\}$, $\mathcal{D}^- = \{1\{R(\tau_1) \leq W'\Pi + h\} : \Pi \in \mathbb{R}^{k_x + k_z}, \ h \in \mathbb{R}\}$, and $\mathcal{D}^+ = \{1\{R(\tau_1) \geq W'\Pi - h\} : \Pi \in \mathbb{R}^{k_x + k_z}, \ h \in \mathbb{R}\}$. Then, for every n, $\mathcal{V}(\mathcal{D}_n) \leq \mathcal{V}(\mathcal{D}^-) + \mathcal{V}(\mathcal{D}^+) - 1$.

Note first that $\mathcal{D}_n = \{1\{|R(\tau_1) - W'\frac{\pi}{r_n}| \leq \frac{h_n}{2}\} : \pi \in \mathbb{R}^{k_x + k_z}\}$ is a subclass of $\mathcal{D}^* = \{1\{|R(\tau_1) - W'\Pi| \leq h\} : \Pi \in \mathbb{R}^{k_x + k_z}, \ h \in \mathbb{R}\}$ for every n. Note also that $\mathcal{D}^-$ and $\mathcal{D}^+$ are VC classes by the VC property (iv) above. It then follows that $\mathcal{V}(\mathcal{D}_n) \leq \mathcal{V}(\mathcal{D}^*) \leq \mathcal{V}(\mathcal{D}^-) + \mathcal{V}(\mathcal{D}^+) - 1$, where the second inequality is due to the VC property (iii) above.

Step 3: Let $\mathcal{F}_n^* = \{(1\{Y \leq S'\theta\} - \tau_1)1\{|R(\tau_1) - W'\frac{\pi}{r_n}| \leq \frac{h_n}{2}\} : \pi \in \mathbb{R}^{k_z + k_x}, \theta \in \mathbb{R}^{1 + k_x}\}$, and $\mathcal{E} = \{(1\{Y \leq S'\theta\} - \tau_1) : \theta \in \mathbb{R}^{1 + k_x}\}$. Then, $\mathcal{V}(\mathcal{F}_n^*) \leq 2\Big(\mathcal{V}(\mathcal{E}) + 2\mathcal{V}(\mathcal{D}^-) + 2\mathcal{V}(\mathcal{D}^+) - 4\Big)$ for every n.

It follows from steps 1 and 2, because $(1_A - \tau_1)1_B = 1_{A \cap B} - \tau_1 1_B$.

Step 4: $\mathcal{F}_n$ is a VC class, and its VC index does not depend on n.

By the VC property (v) above, $\mathcal{V}(\mathcal{F}_n) \leq 2\mathcal{V}(\mathcal{F}_n^*) - 1$. Therefore, it follows from step 3 that $\mathcal{V}(\mathcal{F}_n) \leq 4\Big(\mathcal{V}(\mathcal{E}) + 2\mathcal{V}(\mathcal{D}^-) + 2\mathcal{V}(\mathcal{D}^+) - 4\Big) - 1$ for all n. Lastly, note that $\mathcal{E}, \mathcal{D}^-, \mathcal{D}^+$ are all VC classes and they do not depend on n. $\square$

Lemma 6 *As $n \to \infty$, $h_n \downarrow 0$, $nh_n \to \infty$,*

$$\Upsilon_{n,h_n}\Big(M_n(\Pi_0, \theta_0) - E\big(M_n(\Pi_0, \theta_0)\big)\Big) \xrightarrow{d} N\Big(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}\Big),$$

where $V_{11} = E(W_i W_i' | R_i(\tau_1) = 0) f_{R(\tau_1)}(0)$ and $V_{22} = \tau_2(1 - \tau_2) E(S_i S_i' | R_i(\tau_1) = 0) f_{R(\tau_1)}(0)$.

Remark: Since $R(\tau_1)$ has a conditional density given W, $V_{11} = E\big(W_i W_i' f_{R(\tau_1)|W}(0|W_i)\big)$, which is a more common expression in the quantile literature (e.g. Koenker (2005)). If $R(\tau_1)$ has a density conditional on S, which requires that there is at least one continuous instrument, then $V_{22} = \tau_2(1 - \tau_2) E\big(S_i S_i' f_{R(\tau_1)|S}(0|S_i)\big)$.

Proof: For any (conformable) vectors c, d, consider

$$S_n = \frac{[c' : d']\, \Upsilon_{n,h_n}\Big(M_n(\Pi_0, \theta_0) - E\big(M_n(\Pi_0, \theta_0)\big)\Big)}{\sqrt{Var\Big([c' : d']\, \Upsilon_{n,h_n}\Big(M_n(\Pi_0, \theta_0) - E\big(M_n(\Pi_0, \theta_0)\big)\Big)\Big)}}.$$

By the Cramer-Wold device, it suffices to show that

$$S_n \quad \xrightarrow{d} N(0,1) \tag{15}$$

$$Var\Big(\quad \Upsilon_{n,h_n}\Big(M_n(\Pi_0, \theta_0) - E\big(M_n(\Pi_0, \theta_0)\big)\Big)\Big) \to \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}. \tag{16}$$

Part (16) follows from direct computation, and will be omitted here.

For part (15), write

$$N_n = [c' : d']\, \Upsilon_{n,h_n}\Big(M_n(\Pi_0, \theta_0) - E\big(M_n(\Pi_0, \theta_0)\big)\Big) = \sum_{i=1}^{n} \frac{c'W_i m_{1i}}{\sqrt{n}} + \frac{d'S_i m_{2i} k_{in}}{\sqrt{nh_n}},$$

where $m_{1i} = 1\{D_i \le W_i'\Pi_0\} - \tau_1$, $m_{2i} = 1\{Y_i \le S_i'\theta_0\} - \tau_2$, and $k_{in} = 1\{|D_i - W_i\Pi_0| \le \frac{h_n}{2}\}$. By the Liapounov central limit theorem (see e.g., Pagan and Ullah (1999, p 358)), it suffices to shows that

$$\sum_{i}^{n} E\big(|\frac{c'W_i m_{1i}}{s_n\sqrt{n}} + \frac{d'S_i m_{2i} k_{in}}{s_n\sqrt{nh_n}}|^{2+\delta}\big) = o(1),$$

for some $\delta > 0$, where $s_n^2 = Var(N_n)$. By the C_r inequality,[15]

$$\sum_{i}^{n} E\big(|\frac{c'W_i m_{1i}}{s_n\sqrt{n}} + \frac{d'S_i m_{2i} k_{in}}{s_n\sqrt{nh_n}}|^{2+\delta}\big) \le C \sum_{i}^{n} E\big(|\frac{c'W_i m_{1i}}{s_n\sqrt{n}}|^{2+\delta} + |\frac{d'S_i m_{2i} k_{in}}{s_n\sqrt{nh_n}}|^{2+\delta}\big)$$

$$\le \frac{C}{(ns_n^2)^{1+\delta/2}} \sum_{i}^{n} E\big(|c'W_i m_{1i}|^{2+\delta}\big) + \frac{C}{(nh_n s_n^2)^{1+\delta/2}} \sum_{i}^{n} E\big(|d'S_i m_{2i} k_{in}|^{2+\delta}\big),$$

___
[15]See e.g. Davidson (1994, p 140).

where C is a constant. In view of part (16), s_n^2 is convergent to a positive number, and hence it futher follows that

$$\frac{C}{(ns_n^2)^{1+\delta/2}}\sum_i^n E\big(|c'W_im_{1i}|^{2+\delta}\big) + \frac{C}{(nh_ns_n^2)^{1+\delta/2}}\sum_i^n E\big(|d'S_im_{2i}k_{in}|^{2+\delta}\big)$$

$$\leq \frac{1}{n^{1+\delta/2}}O(n) + \frac{1}{(nh_n)^{1+\delta/2}}O(nh_n) = o(1),$$

where δ is chosen to satisfy assumption H. $\square$

Lemma 7 *Let the Jacobian of $M(\Pi,\theta)$ evaluated at Π_0 and θ_0 be given by $\Gamma = \begin{bmatrix} \Gamma_{11} & 0 \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}$. Then, $\Gamma_{11} = E\big(W_iW_i'f_{R(\tau_1)|W}(0|W_i)\big)$, and $\Gamma_{22} = E\big(S_iS_i'f_{\epsilon(\tau_1,\tau_2)|R(\tau_1),S}(0|R_i(\tau_1),S_i)|R_i(\tau_1) = 0\big)f_{R(\tau_1)}(0)$.*

Proof: Γ_{11} and Γ_{22} follows from direct calculation, and it will be omitted here. $\square$

Remark: Let $T(D_i,W_i) = E\big(S_i(1\{\epsilon(\tau_1,\tau_2)\leq 0\} - \tau_2)|D_i,W_i\big)$, and assume that $T(D_i,W_i)$ is twice continuously differentiable with respect to D_i. Then, differentiability of $M(\Pi,\theta)$ at Π_0 and θ_0 follows. To see this, consider Γ_{21}. For each $\Pi \in \Pi_0 + \mathcal{N}$,

$$E\big(T(D_i,W_i)1\{|D_i - W_i'\Pi| \leq \frac{h_n}{2}\}\frac{1}{h_n}\big) \to E\big(T(D_i,W_i)|D_i - W_i'\Pi = 0\big)f_{D_i-W_i'\Pi}(0)$$

$$= E\big(S_i(1\{\epsilon(\tau_1,\tau_2)\leq 0\} - \tau_2)|D_i - W_i'\Pi = 0\big)f_{D_i-W_i'\Pi}(0) = M_2(\Pi,\theta_0). \quad (17)$$

Therefore, it suffices to show that the derivatives of the left–hand side of (17) uniformly converges over $\Pi_0 + \mathcal{N}$. Note that

$$E\big(T(D_i,W_i)1\{|D_i - W_i'\Pi| \leq \frac{h_n}{2}\}\frac{1}{h_n}|W_i\big) = \int T(t,W_i)1\{|t - W_i'\Pi| \leq \frac{h_n}{2}\}\frac{1}{h_n}f_{D|W}(t|W_i)dt$$

$$= \int_{-1/2}^{1/2} T(th_n + W_i'\Pi,W_i)f_{D|W}(th_n + W_i'\Pi|W_i)dt,$$

which has a derivative

$$\int_{-1/2}^{1/2} W_i T_D(th_n + W_i'\Pi, W_i) f_{D|W}(th_n + W_i'\Pi|W_i) + W_i T(th_n + W_i\Pi, W_i) f_{D|W}'(th_n + W_i'\Pi|W_i)dt, \tag{18}$$

where T_D denotes the derivative of $T(D_i, W_i)$ with respect to D_i. Since $\Pi_0 + \mathcal{N}$ is compact, uniform convergence of (18) follows from cotinuous differentiability of T_D and $f_{D|W}'(s|W_i)$. $\square$

Remark: $\Gamma_{21} = E\Big(W_i T_D(W_i'\Pi_0, W_i) f_{D|W}(W_i'\Pi_0|W_i) + W_i T(W_i\Pi_0, W_i) f_{D|W}'(W_i'\Pi_0|W_i)\Big)$. In fact, if $R(\tau_1)$ has a conditional density given Y_i, S_i, which is the case when there is at least one continuous instrument, it is simply $\Gamma_{21} = E\big(S_i W_i'(1\{Y_i \leq S_i'\theta_0\} - \tau_2) f_{R(\tau_1)|Y,S}'(0|Y_i, S_i)\big)$. Similarly, Γ_{22} can be written as $E\big(S_i S_i' f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(0,0|S_i)\big)$ in that case.

Lemma 8 $\sqrt{nh_n}(\hat{\theta} - \theta_0) = O_p(1) + O(\sqrt{nh_n}h_n^2)$.

Proof: Since

$$\sup_{\theta \in \Theta} ||M_{2n}(\hat{\Pi}, \theta) - M_2(\Pi_0, \theta)|| = o_p(1) \tag{19}$$

by lemma 5, consistency of $\hat{\theta}$ follows from the standard arguments. Choose an arbitrary $\delta > 0$. Then, continuity of $M_2(\Pi_0, \theta)$ and the compact parameter space guarantees that there is some $\theta^* \in \{\theta \in \Theta : ||\theta - \theta_0|| \geq \delta\}$ such that $\inf_{||\theta-\theta_0||\geq\delta} ||M_2(\Pi_0, \theta)|| = ||M_2(\Pi_0, \theta^*)||$. Since $M_2(\Pi_0, \theta)$ is equal to 0 uniquly at θ_0, it follows that $\inf_{||\theta-\theta_0||\geq\delta} ||M_2(\Pi_0, \theta)|| \geq \epsilon$ for some $\epsilon > 0$. It then follows that $\Pr(||\hat{\theta} - \theta_0|| \geq \delta) \leq \Pr(||M_2(\Pi_0, \hat{\theta})|| \geq \epsilon)$. Now, note that

$$||M_2(\Pi_0, \hat{\theta})|| \leq \sup ||M_{2n}(\hat{\Pi}, \theta) - M_2(\Pi_0, \theta)|| + ||M_{2n}(\hat{\Pi}, \hat{\theta})|| = o_p(1),$$

where the last equality is due to equations (10) and (19). Therefore, $||\hat{\theta} - \theta_0|| = o_p(1)$.

To establish the convergence rate, let $\delta_{2n} \downarrow 0$ be an arbitrary sequence and let $\delta_{1n} \downarrow 0$ be a sequence satisfying the conditions of lemma 5. It then follows from lemma 5 that

$$\sup ||\sqrt{nh_n}\big(M_{2n}(\Pi, \theta) - M_2(\Pi, \theta)\big) - \sqrt{nh_n}\big(M_{2n}(\Pi_0, \theta_0) - M_2(\Pi_0, \theta_0)\big)||$$

$$\leq \sup ||\sqrt{nh_n}\Big(M_{2n}(\Pi, \theta) - E\big(M_{2n}(\Pi, \theta)\big)\Big) - \sqrt{nh_n}\Big(M_{2n}(\Pi_0, \theta_0) - E\big(M_{2n}(\Pi_0, \theta_0)\big)\Big)||$$

$$+ 2\sqrt{nh_n} \sup ||E\big(M_{2n}(\Pi, \theta)\big) - M_2(\Pi, \theta)|| = o_p(1) + O(\sqrt{nh_n}h_n^2), \quad (20)$$

where sup is taken over $||\Pi - \Pi_0|| < \delta_{1n}$ and $||\theta - \theta_0|| < \delta_{2n}$; for the last equality, I used lemma 1. Since $\hat{\Pi}$ is $\sqrt{n}$–consistent and $\hat{\theta}$ is consistent, there exist some sequences $\delta_{1n} \downarrow 0$ and $\delta_{2n} \downarrow 0$ such that

$$||\sqrt{nh_n}M_2(\hat{\Pi}, \hat{\theta})|| \leq ||\sqrt{nh_n}\big(M_{2n}(\hat{\Pi}, \hat{\theta}) - M_2(\hat{\Pi}, \hat{\theta})\big)|| + ||\sqrt{nh_n}M_{2n}(\hat{\Pi}, \hat{\theta})||$$

$$\leq \sup ||\sqrt{nh_n}\big(M_{2n}(\Pi, \theta) - M_2(\Pi, \theta)\big)|| + o_p(1)$$

$$\leq \sup ||\sqrt{nh_n}\big(M_{2n}(\Pi, \theta) - M_2(\Pi, \theta)\big) - \sqrt{nh_n}M_{2n}(\Pi_0, \theta_0)|| + ||\sqrt{nh_n}M_{2n}(\Pi_0, \theta_0)|| + o_p(1)$$

$$= o_p(1) + O(\sqrt{nh_n}h_n^2) + O_p(1) + o_p(1)$$

with probability approaching to 1, where sup is taken over $||\Pi - \Pi_0|| < \delta_{1n}$ and $||\theta - \theta_0|| < \delta_{2n}$.[16] Therefore,

$$O(h_n^2) + O_p(\frac{1}{\sqrt{nh_n}}) = ||M_2(\hat{\Pi}, \hat{\theta})|| = ||\Gamma_{21}(\hat{\Pi} - \Pi_0) + \Gamma_{22}(\hat{\theta} - \theta_0) + O(||\hat{\Pi} - \Pi_0||||\hat{\theta} - \theta_0||)||$$

$$= ||\Gamma_{22}(\hat{\theta} - \theta_0) + O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{||\hat{\theta} - \theta_0||}{\sqrt{n}})|| \geq ||\Gamma_{22}(\hat{\theta} - \theta_0)|| - O_p(\frac{1}{\sqrt{n}}) \geq \sqrt{\lambda_2}||\hat{\theta} - \theta_0|| - O_p(\frac{1}{\sqrt{n}}),$$

where λ_2 is the smallest eigenvalue of $\Gamma'_{22}\Gamma_{22}$. Since the full rank of Γ_{22} implies that $\lambda_2 > 0$, it follows that $\sqrt{nh_n}||\hat{\theta} - \theta_0|| = O_p(1) + O(\sqrt{nh_n}h_n^2) + O_p(\sqrt{h_n})$. $\square$

Proposition 3 *The proposed estimators have the asymptotic expansion of theorem 1.*

---

[16]Any $\delta_{1n} \downarrow 0$ satisfying $\sqrt{n}\delta_{1n} \to \infty$ and $r_n\delta_{1n} \to 0$ will do, where r_n is defined in lemma 5.

Proof: Since the asymptotic expansion of $\hat{\Pi}(\tau_1)$ follows from standard quantile regression, its proof will be omitted and I only consider $\hat{\theta}(\tau_1, \tau_2)$ here. In the following proof, the asymptotic expansion $\sqrt{n}(\hat{\Pi} - \Pi_0) = \Gamma_{11} M_{1n}(\Pi_0) + o_p(1)$ will be taken for granted.

From differentiability of $M_2(\Pi, \theta)$, consider the expansion

$$M_2(\Pi, \theta) = \Gamma_{21}(\Pi - \Pi_0) + \Gamma_{22}(\theta - \theta_0) + O(||\Pi - \Pi_0||||\theta - \theta_0||).$$

Define

$$L_{2n}(\theta) = \Gamma_{22}(\theta - \theta_0) + M_{2n}(\Pi_0, \theta_0).$$

Then,

$$M_{2n}(\Pi, \theta) = M_2(\Pi, \theta) + \big(M_{2n}(\Pi, \theta) - M_2(\Pi, \theta)\big)$$

$$= L_{2n}(\theta) + \big(M_{2n}(\Pi, \theta) - M_2(\Pi, \theta)\big) - M_{2n}(\Pi_0, \theta_0) + O(||\Pi - \Pi_0||) + O(||\Pi - \Pi_0||||\theta - \theta_0||)$$

$$= L_{2n}(\theta) + \Big(M_{2n}(\Pi, \theta) - E\big(M_{2n}(\Pi, \theta)\big) - M_{2n}(\Pi_0, \theta_0) + E\big(M_{2n}(\Pi_0, \theta_0)\big)\Big)$$

$$+ E\big(M_{2n}(\Pi, \theta)\big) - M_2(\Pi, \theta) - E\big(M_{2n}(\Pi_0, \theta_0) + O(||\Pi - \Pi_0||) + O(||\Pi - \Pi_0||||\theta - \theta_0||). \quad (21)$$

In fact, equation (21) shows that $M_{2n}(\Pi, \theta)$ is decomposed into four different components; the linear term $L_{2n}(\theta)$, the stochastic equicontinuity term $\Big(M_{2n}(\Pi, \theta) - E\big(M_{2n}(\Pi, \theta)\big)\Big) - \Big(M_{2n}(\Pi_0, \theta_0) - E\big(M_{2n}(\Pi_0, \theta_0)\big)\Big)$, the bias term $\Big(E\big(M_{2n}(\Pi_0, \theta)\big) - M_2(\Pi, \theta)\Big) - E\big(M_{2n}(\Pi_0, \theta_0)\big)$, and the remainer term. Therefore, for any estimator θ_n such that $\sqrt{nh_n}(\theta_n - \theta_0) = O_p(1)$,

$$\sqrt{nh_n}||M_{2n}(\hat{\Pi}, \theta_n) - L_{2n}(\theta_n)|| \leq \sqrt{nh_n}||\big(M_{2n}(\hat{\Pi}, \theta_n) - M_2(\hat{\Pi}, \theta_n)\big) - M_{2n}(\Pi_0, \theta_0)|| + O_p(\sqrt{h_n})$$

$$\leq \sup \sqrt{nh_n}||\big(M_{2n}(\Pi, \theta) - M_2(\Pi, \theta)\big) - M_{2n}(\Pi_0, \theta_0)|| + o_p(1) = O(\sqrt{nh_n}h_n^2) + o_p(1),$$

where sup is taken over $||\Pi - \Pi_0|| < \delta_{1n}$ and $||\theta - \theta_0|| < \delta_{2n}$ for some sequences $\delta_{1n} \downarrow 0$ and $\delta_{2n} \downarrow 0$; the second inequality holds with probability approaching to 1 due to consistency, and the

last equality is by equation (20). In particular, undersmoothing ensures that

$$\sqrt{nh_n}||M_{2n}(\hat{\Pi}, \theta_n) - L_{2n}(\theta_n)|| = o_p(1). \tag{22}$$

Now, consider an (infeasible) estimator $\tilde{\theta}$ such that

$$\tilde{\theta} - \theta_0 = \arg\min_{\theta \in \mathbb{R}^{1+k_x}} ||L_{2n}(\theta)|| = -\Gamma_{22}^{-1} M_{2n}(\Pi_0, \theta_0).$$

Note here that although I minimized over $\mathbb{R}^{1+k_x}$, $\tilde{\theta} - \theta_0$ will belong to Θ as n increases, because θ_0 is in the interior of Θ. In the following, I will show that $\hat{\theta}$ is distributionally equivalent to $\tilde{\theta}$. First, note that $\sqrt{nh_n}(\hat{\theta} - \theta_0) = O_p(1)$ and $\sqrt{nh_n}(\tilde{\theta} - \theta_0) = O_p(1)$. Therefore, by equation (22),

$$\sqrt{nh_n}||L_{2n}(\hat{\theta})|| = \sqrt{nh_n}||M_{2n}(\hat{\Pi}, \hat{\theta})|| + o_p(1) = o_p(1).$$

Now, note that $L_{2n}(\hat{\theta}) = L_{2n}(\tilde{\theta}) + \Gamma_{22}(\hat{\theta} - \tilde{\theta})$ by definition, where $L_{2n}(\tilde{\theta})$ is in fact equal to 0. Therefore, it follows that

$$o_p(1) = \sqrt{nh_n}||\Gamma_{22}(\hat{\theta} - \tilde{\theta})|| \geq \sqrt{nh_n}\sqrt{\lambda_2}||(\hat{\theta} - \tilde{\theta})|| \geq \sqrt{\lambda_2}\left|||\sqrt{nh_n}(\hat{\theta} - \theta_0)|| - ||\sqrt{nh_n}(\tilde{\theta} - \theta_0)||\right|,$$

where $\lambda_2 > 0$ is the smallest eigenvalue of $\Gamma_{22}'\Gamma_{22}$. Therefore,

$$\sqrt{nh_n}(\hat{\theta} - \theta_0) = \sqrt{nh_n}(\tilde{\theta} - \theta_0) + o_p(1),$$

which completes the proof. □

E   Proof of Proposition 2

Proposition 2 follows from lemmas 9 and 10. □

Lemma 9

$$\frac{1}{nb_{1n}}\sum_{i=1}^{n}S_iS_i'k\big(\frac{-\hat{R}_i(\tau_1)}{b_{1n}}\big) = \frac{1}{nb_{1n}}\sum_{i=1}^{n}S_iS_i'k\big(\frac{-R_i(\tau_1)}{b_{1n}}\big) + o_p(1),$$

$$\frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k\big(\frac{-\hat{R}_i(\tau_1)}{b_{2n}}\big)k\big(\frac{-\hat{\epsilon}_i(\tau_1,\tau_2)}{b_{2n}}\big) = \frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k\big(\frac{-R_i(\tau_1)}{b_{2n}}\big)k\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big) + o_p(1).$$

Proof: Since two statements are similar, we only prove the second one. Let $\Delta_1 = \Pi - \Pi_0$, $\Delta_2 = \theta - \theta_0$ and $\hat{\Delta}_1 = \hat{\Pi} - \Pi_0$, $\hat{\Delta}_2 = \hat{\theta} - \theta_0$. By the mean value theorem, we can write

$$\frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k\big(\frac{-\hat{R}_i(\tau_1)}{b_{2n}}\big)k\big(\frac{-\hat{\epsilon}_i(\tau_1,\tau_2)}{b_{2n}}\big) = \frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k\big(\frac{-R_i(\tau_1)}{b_{2n}}\big)k\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big)$$

$$+ \frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k'\big(\frac{W_i'\tilde{\Delta}_1 - R_i(\tau_1)}{b_{2n}}\big)k\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big)\frac{W_i'\hat{\Delta}_1}{b_{2n}}$$

$$+ \frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k\big(\frac{-R_i(\tau_1)}{b_{2n}}\big)k'\big(\frac{S_i'\tilde{\Delta}_2 - \epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big)\frac{S_i'\hat{\Delta}_2}{b_{2n}}$$

$$+ \frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k'\big(\frac{W_i'\tilde{\Delta}_1 - R_i(\tau_1)}{b_{2n}}\big)k'\big(\frac{S_i'\tilde{\Delta}_2 - \epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big)\frac{W_i'\hat{\Delta}_1}{b_{2n}}\frac{S_i'\hat{\Delta}_2}{b_{2n}} \quad (23)$$

for some $\tilde{\Delta}_1$ and $\tilde{\Delta}_2$.

We claim that the last three terms in (23) are all $o_p(1)$, which easily follows from the stronger bandwidth requirements when W_i has unbounded support. Therefore, we only consider the case where b_{1n}, b_{2n} satisfy the weaker bandwidth requirements but W_i has bounded support.

Consider the last term in (23) since the other two terms are easier. Since $||\frac{\hat{\Delta}_1}{b_{2n}}|| = O_p(\frac{1}{\sqrt{n}b_{2n}}) = o_p(1)$ and $\frac{\hat{\Delta}_2}{b_{2n}} = O_p(\frac{1}{\sqrt{nh_nb_{2n}}}) = o_p(1)$, for some sequence $\delta_n \downarrow 0$, we have

$$||\frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k'\big(\frac{W_i'\tilde{\Delta}_1 - R_i(\tau_1)}{b_{2n}}\big)k'\big(\frac{S_i'\tilde{\Delta}_2 - \epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big)\frac{W_i'\hat{\Delta}_1}{b_{2n}}\frac{S_i'\hat{\Delta}_2}{b_{2n}}||$$

$$\leq \sup_{||\frac{\hat{\Delta}_1}{b_{2n}}||\leq\delta_n,\ ||\frac{\hat{\Delta}_2}{b_{2n}}||\leq\delta_n}||\frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k'\big(\frac{W_i'\tilde{\Delta}_1 - R_i(\tau_1)}{b_{2n}}\big)k'\big(\frac{S_i'\tilde{\Delta}_2 - \epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big)\frac{W_i'\hat{\Delta}_1}{b_{2n}}\frac{S_i'\hat{\Delta}_2}{b_{2n}}||$$

$$\leq \frac{1}{nb_{2n}^2}\sum_{i=1}^{n}||S_i||^3||W_i||\sup_{|t|\leq||W_i||\delta_n}|k'\big(\frac{-R_i(\tau_1)}{b_{2n}} + t\big)|\sup_{|t|\leq||S_i||\delta_n}|k'\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}} + t\big)|\delta_n^2$$

with probability approaching to 1. Note here that

$$
\frac{1}{nb_{2n}^2}\sum_{i=1}^n ||S_i||^3||W_i|| \sup_{|t|\le ||W_i||\delta_n} |k'\big(\frac{-R_i(\tau_1)}{b_{2n}}+t\big)| \sup_{|t|\le ||S_i||\delta_n} |k'\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}+t\big)|
$$

$$
\le C\frac{1}{nb_{2n}^2}\sum_{i=1}^n ||S_i||^3 \sup_{|t|\le C\delta_n} |k'\big(\frac{-R_i(\tau_1)}{b_{2n}}+t\big)| \sup_{|t|\le ||S_i||\delta_n} |k'\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}+t\big)| \quad (24)
$$

for some $C>0$, because W_i has bounded support. Therefore, it suffices to shows that the RHS in (24) is $O_p(1)$. Note here that

$$
\sup_{|t|\le C\delta_n} |k'\big(\frac{-R_i(\tau_1)}{b_{2n}}+t\big)|
$$

$$
= \max_{c_j\in\mathcal{S}}\big(|k'(c_j)|1\{|c_j+\frac{R_i(\tau_1)}{b_{2n}}|<C\delta_n\}, \ |k'\big(\frac{-R_i(\tau_1)}{b_{2n}}+C\delta_n\big)|, \ |k'\big(\frac{-R_i(\tau_1)}{b_{2n}}-C\delta_n\big)|\big)
$$

and that

$$
\sup_{|t|\le ||S_i||\delta_n} |k'\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}+t\big)|
$$

$$
= \max_{c_j\in\mathcal{S}}\big(|k'(c_j)|1\{|c_j+\frac{\epsilon_i(\tau_1,\tau_2)}{b_{2n}}|<||S_i||\delta_n\}, \ |k'\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}+||S_i||\delta_n\big)|, \ |k'\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}-||S_i||\delta_n\big)|\big).
$$

Therefore, plugging in these expressions to the RHS of (24), taking expectations, and using the usual change of variables technique finishes the proof. □

Lemma 10 *If $\epsilon_i(\tau_1,\tau_2)$ and $R_i(\tau_1)$ have a joint density conditional on S_i (e.g. when there is at least one continuous instrument),*

$$
\frac{1}{nb_{1n}}\sum_{i=1}^n S_iS_i'k\big(\frac{-R_i(\tau_1)}{b_{1n}}\big) = E\Big(S_iS_i'f_{R(\tau_1)|S}(0|S_i)\Big) + o_p(1),
$$

$$
\frac{1}{nb_{2n}^2}\sum_{i=1}^n S_iS_i'k\big(\frac{-R_i(\tau_1)}{b_{2n}}\big)k\big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\big) = E\Big(S_iS_i'f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(0,0|S_i)\Big) + o_p(1).
$$

Proof: Since the two statements are similar, we only consider the second one. Note first that

$$\frac{1}{nb_{2n}^2}\sum_{i=1}^{n}S_iS_i'k\Big(\frac{-R_i(\tau_1)}{b_{2n}}\Big)k\Big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\Big) = E\Big(\frac{1}{b_{2n}^2}S_iS_i'k\Big(\frac{-R_i(\tau_1)}{b_{2n}}\Big)k\Big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\Big)\Big)+O_p\Big(\frac{1}{\sqrt{n}b_{2n}}\Big),$$

$$(25)$$

because letting ξ_{tsi} be the t–s element of S_iS_i', squaring and taking expectations yields

$$\frac{1}{nb_{2n}^2}E\Big(\xi_{tsi}^2k\Big(\frac{-R_i(\tau_1)}{b_{2n}}\Big)^2k\Big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\Big)^2\frac{1}{b_{2n}^2}\Big)$$

$$= \frac{1}{nb_{2n}^2}E\Big(\xi_{tsi}^2\int k\Big(\frac{-r}{b_{2n}}\Big)^2k\Big(\frac{-e}{b_{2n}}\Big)^2\frac{1}{b_{2n}^2}f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(r,e|S_i)drde\Big)$$

$$= \frac{1}{nb_{2n}^2}E\Big(\xi_{tsi}^2\int k(t)^2k(s)^2f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(-b_{2n}t,-b_{2n}s|S_i)dtds\Big) = O\Big(\frac{1}{nb_{2n}^2}\Big).$$

Therefore, I only consider the expectation in the RHS of equation (25).

$$E\Big(\frac{1}{b_{2n}^2}S_iS_i'k\Big(\frac{-R_i(\tau_1)}{b_{2n}}\Big)k\Big(\frac{-\epsilon_i(\tau_1,\tau_2)}{b_{2n}}\Big)\Big) = E\Big(S_iS_i'\int k(t)k(s)f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(-b_{2n}t,-b_{2n}s|S_i)dtds\Big)$$

$$= E\Big(S_iS_i'f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(0,0|S_i)\Big)\int k(t)dt\int k(s)ds$$

$$+ b_{2n}E\Big(S_iS_i'\int k(t)k(s)\big(t\mathcal{D}_1f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(\bar{t},\bar{s}|S_i)+s\mathcal{D}_2f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(\bar{t},\bar{s}|S_i)\big)dtds\Big),$$

where $\mathcal{D}_jf_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}$ is the partial derivative with respect to the jth argument and $\bar{t}$, $\bar{s}$ denote the mean values. Since $\sup_{t,s}|\mathcal{D}_jf_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(t,s|S_i)| \leq \phi(S_i)$, we have

$$b_{2n}||E\Big(S_iS_i'\int k(t)k(s)\big(t\mathcal{D}_1f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(\bar{t},\bar{s}|S_i)+s\mathcal{D}_2f_{R(\tau_1),\epsilon(\tau_1,\tau_2)|S}(\bar{t},\bar{s}|S_i)\big)dtds\Big)||$$

$$\leq b_{2n}E\Big(||S_i||^2\phi(S_i)\int |k(t)||k(s)|(|t|+|s|)dtds\Big) = o(1). \quad \square$$

Table I: Experiments using the Angrist and Krueger data

p-values of the significance of instruments in the first stage quantile regression

$\tau_1 =$	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$\sigma = 0.01$	0.0000	0.0000	0.0000	0.0000	0.0000	0.2641	0.3110	0.3792	0.1254
$\sigma = 0.02$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0031	0.1060	0.1550	0.0111
$\sigma = 0.03$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sigma = 0.04$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
$\sigma = 0.05$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

$\tau_1 =$	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
$\sigma = 0.01$	0.0000	0.0000	0.0000	0.0000	0.0000	0.8133	0.9997	0.0000	0.0000
$\sigma = 0.02$	0.0000	0.0000	0.0000	0.0000	0.0000	0.5650	0.9840	0.0014	0.0000
$\sigma = 0.03$	0.0000	0.0000	0.0000	0.0000	0.0000	0.4630	0.9400	0.0020	0.0000
$\sigma = 0.04$	0.0000	0.0000	0.0000	0.0001	0.0000	0.3775	0.6384	0.0028	0.0000
$\sigma = 0.05$	0.0000	0.0000	0.0000	0.0021	0.0000	0.3010	0.4350	0.0173	0.0000

Note

H_0: $\Pi_2(\tau_1) = 0$, where $Q_{D_i^*|X_i,Z_i}(\tau_1) = X_i'\Pi_1(\tau_1) + Z_i'\Pi_2(\tau_1)$ with $D_i^* = D_i + \epsilon_i$.

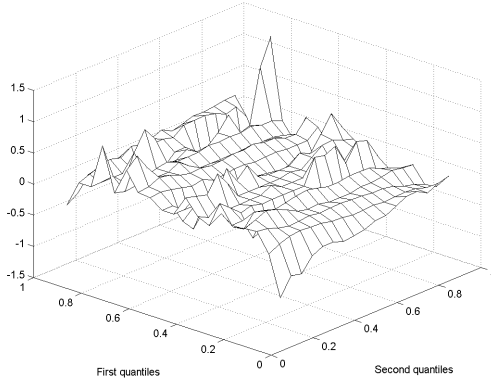D_i is the education variable of the Angrist–Krueger data, and ϵ_i is a random noise from $N(0, \sigma^2)$.

X_i contains 10 dummies indicating birth-years.

Z_i contains 30 instruments of birth-quarters interacted with birth-years.

Figure I: Experiments using the Angrist and Krueger data
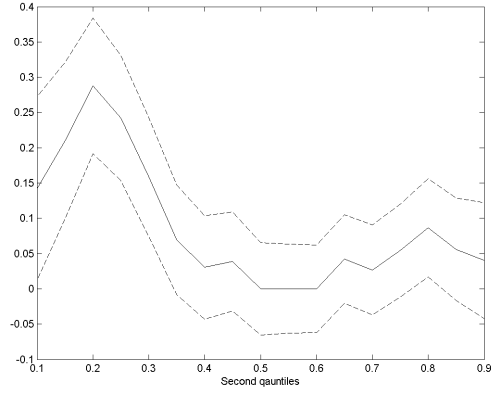
Point Estimates of $\alpha(\tau_1, \tau_2)$ with $\sigma = 0.01$ Confidence Intervals for $\alpha(0.20, \tau_2)$ with $\sigma = 0.01$
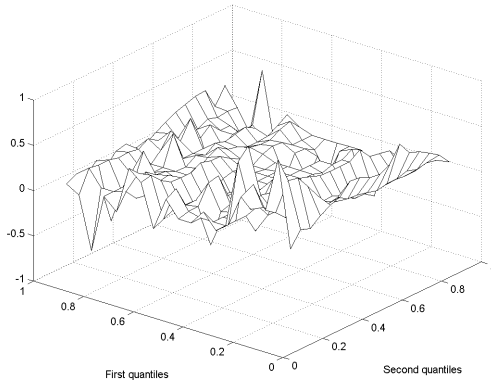
Point Estimates of $\alpha(\tau_1, \tau_2)$ with $\sigma = 0.02$ Confidence Intervals for $\alpha(0.20, \tau_2)$ with $\sigma = 0.02$
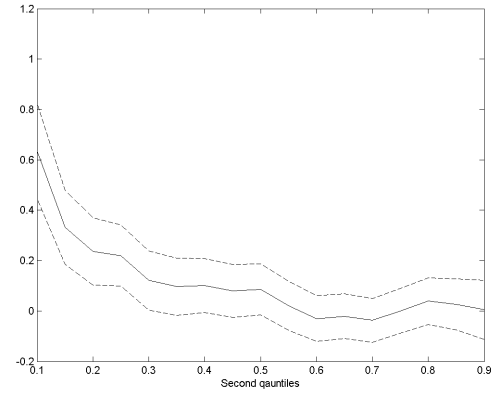
Note:

$Q_{Y_i|Z_i, X_i, V_i = \tau_1}(\tau_2) = Q_{D_i^*|Z_i, X_i}(\tau_1)\alpha(\tau_1, \tau_2) + X_i'\beta(\tau_1, \tau_2)$, where $D_i^* = D_i + \epsilon_i$.

$\epsilon_i \sim N(0, \sigma^2)$ is the same as those used in Table I.

Y_i is the log wage variable of the Angrist–Krueger data.

Table II-1: Monte Carlo Results of DGP1

τ_1	$\alpha(\tau_1, 0.5)$	Bias			Variance			RMSE		
		$n = 3000$	$n = 5000$	$n = 8000$	$n = 3000$	$n = 5000$	$n = 8000$	$n = 3000$	$n = 5000$	$n = 8000$
0.10	1.8152	-0.0018	-0.0039	-0.0070	0.0175	0.0115	0.0092	0.1322	0.1075	0.0960
0.15	1.7343	-0.0045	-0.0039	-0.0052	0.0144	0.0093	0.0071	0.1201	0.0964	0.0846
0.20	1.6614	-0.0011	0.0015	0.0021	0.0131	0.0084	0.0059	0.1144	0.0919	0.0768
0.25	1.5965	-0.0064	-0.0028	-0.0014	0.0096	0.0074	0.0051	0.0979	0.0859	0.0717
0.30	1.5399	0.0031	-0.0028	-0.0038	0.0100	0.0069	0.0048	0.0998	0.0831	0.0691
0.35	1.4916	-0.0001	-0.0034	0.0009	0.0096	0.0060	0.0043	0.0982	0.0773	0.0659
0.40	1.4518	-0.0022	-0.0010	-0.0007	0.0078	0.0058	0.0044	0.0884	0.0760	0.0665
0.45	1.4208	0.0000	0.0000	-0.0011	0.0075	0.0054	0.0040	0.0869	0.0736	0.0629
0.50	1.3987	-0.0030	0.0022	0.0015	0.0075	0.0052	0.0038	0.0864	0.0722	0.0620
0.55	1.3858	0.0022	0.0030	0.0009	0.0067	0.0048	0.0037	0.0816	0.0697	0.0606
0.60	1.3821	-0.0018	-0.0033	0.0004	0.0068	0.0052	0.0038	0.0823	0.0722	0.0617
0.65	1.3880	0.0011	0.0008	0.0033	0.0078	0.0054	0.0038	0.0883	0.0737	0.0619
0.70	1.4038	-0.0009	-0.0043	-0.0007	0.0081	0.0060	0.0042	0.0902	0.0776	0.0648
0.75	1.4295	-0.0030	0.0043	0.0004	0.0100	0.0061	0.0050	0.1003	0.0781	0.0711
0.80	1.4655	0.0066	-0.0024	-0.0036	0.0109	0.0080	0.0057	0.1045	0.0896	0.0756
0.85	1.5121	0.0039	0.0025	0.0032	0.0153	0.0094	0.0067	0.1239	0.0972	0.0817
0.90	1.5696	-0.0029	0.0029	-0.0036	0.0184	0.0131	0.0082	0.1358	0.1143	0.0909

Table II-2: Monte Carlo Results of DGP2

τ_1	$\alpha(\tau_1, 0.5)$	Bias			Variance			RMSE		
		$n=3000$	$n=5000$	$n=8000$	$n=3000$	$n=5000$	$n=8000$	$n=3000$	$n=5000$	$n=8000$
0.10	1.8152	-0.0205	-0.0049	-0.0153	0.0473	0.0309	0.0234	0.2185	0.1757	0.1537
0.15	1.7343	-0.0341	-0.0298	0.0024	0.0892	0.0584	0.0436	0.3005	0.2436	0.2089
0.20	1.6614	-0.0753	-0.0380	-0.0290	0.1917	0.1342	0.0901	0.4442	0.3683	0.3016
0.25	1.5965	-0.2930	-0.1692	-0.1128	0.5849	0.4531	0.2984	0.8190	0.6941	0.5577
0.30	1.5399	-1.1475	-1.0843	-1.0642	2.0345	1.9117	2.0003	1.8306	1.7571	1.7700
0.35	1.4916	-0.4447	-0.3165	-0.2343	0.9930	0.8596	0.5473	1.0912	0.9797	0.7760
0.40	1.4518	-0.0843	-0.0575	-0.0193	0.2372	0.1695	0.1400	0.4943	0.4157	0.3747
0.45	1.4208	-0.0462	-0.0162	-0.0143	0.1066	0.0764	0.0521	0.3298	0.2770	0.2286
0.50	1.3987	-0.0260	-0.0184	-0.0077	0.0539	0.0397	0.0287	0.2337	0.2002	0.1695
0.55	1.3858	-0.0078	0.0000	-0.0088	0.0388	0.0250	0.0167	0.1972	0.1580	0.1294
0.60	1.3821	-0.0071	0.0012	-0.0091	0.0223	0.0163	0.0127	0.1494	0.1278	0.1129
0.65	1.3880	-0.0046	-0.0014	-0.0059	0.0192	0.0139	0.0087	0.1386	0.1180	0.0935
0.70	1.4038	-0.0055	-0.0044	-0.0029	0.0141	0.0103	0.0079	0.1188	0.1015	0.0890
0.75	1.4295	0.0012	0.0028	-0.0005	0.0132	0.0093	0.0067	0.1147	0.0963	0.0820
0.80	1.4655	-0.0020	0.0021	0.0003	0.0117	0.0080	0.0060	0.1082	0.0897	0.0777
0.85	1.5121	0.0056	0.0008	0.0016	0.0113	0.0071	0.0053	0.1065	0.0840	0.0729
0.90	1.5696	0.0045	0.0061	0.0027	0.0099	0.0067	0.0047	0.0994	0.0819	0.0683

Table II-3: Monte Carlo Results of DGP3

τ_1	$\alpha(\tau_1, 0.5)$	Bias			Variance			RMSE		
		$n = 3000$	$n = 5000$	$n = 8000$	$n = 3000$	$n = 5000$	$n = 8000$	$n = 3000$	$n = 5000$	$n = 8000$
0.10	1.8152	-0.0266	-0.0040	-0.0053	0.0508	0.0340	0.0264	0.2269	0.1843	0.1627
0.15	1.7343	-0.0234	-0.0252	-0.0033	0.0915	0.0668	0.0491	0.3035	0.2597	0.2216
0.20	1.6614	-0.0280	-0.0169	-0.0278	0.2305	0.1481	0.1079	0.4809	0.3852	0.3297
0.25	1.5965	-0.1940	-0.0756	-0.0995	0.9424	0.5139	0.4170	0.9900	0.7208	0.6534
0.30	1.5399	-1.5233	-1.5528	-1.4428	4.2290	4.4756	4.6431	2.5592	2.6243	2.5932
0.35	1.4916	-0.3367	-0.2186	-0.1702	1.9904	1.3365	0.8952	1.4504	1.1766	0.9614
0.40	1.4518	-0.0495	-0.0445	-0.0373	0.4108	0.2348	0.1841	0.6429	0.4866	0.4307
0.45	1.4208	-0.0133	0.0014	-0.0079	0.1440	0.0968	0.0695	0.3797	0.3111	0.2638
0.50	1.3987	0.0036	-0.0143	-0.0209	0.0719	0.0549	0.0383	0.2682	0.2347	0.1968
0.55	1.3858	-0.0173	-0.0044	0.0020	0.0483	0.0338	0.0234	0.2205	0.1839	0.1530
0.60	1.3821	-0.0053	-0.0075	-0.0007	0.0348	0.0220	0.0166	0.1867	0.1484	0.1288
0.65	1.3880	-0.0057	0.0025	-0.0018	0.0224	0.0163	0.0115	0.1496	0.1278	0.1072
0.70	1.4038	-0.0053	0.0020	-0.0015	0.0184	0.0139	0.0094	0.1358	0.1177	0.0969
0.75	1.4295	-0.0013	0.0079	0.0011	0.0166	0.0115	0.0081	0.1288	0.1075	0.0900
0.80	1.4655	0.0034	0.0026	-0.0005	0.0138	0.0093	0.0075	0.1174	0.0964	0.0867
0.85	1.5121	0.0040	0.0041	-0.0012	0.0120	0.0082	0.0065	0.1098	0.0905	0.0808
0.90	1.5696	0.0129	0.0052	0.0027	0.0109	0.0076	0.0053	0.1052	0.0876	0.0728

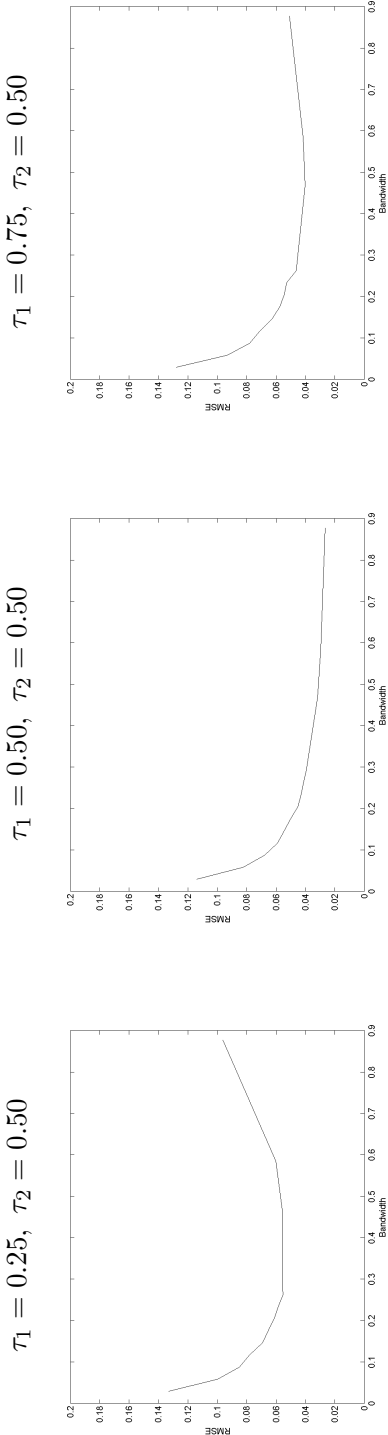Figure II-1: Sensitivity of RMSE to bandwidth choices (DGP1 with $n = 5000$)

$\tau_1 = 0.25, \ \tau_2 = 0.50$

$\tau_1 = 0.50, \ \tau_2 = 0.50$

$\tau_1 = 0.75, \ \tau_2 = 0.50$

Figure II-2: Sensitivity of RMSE to bandwidth choices (DGP2 with $n = 5000$)

$\tau_1 = 0.30, \ \tau_2 = 0.50$

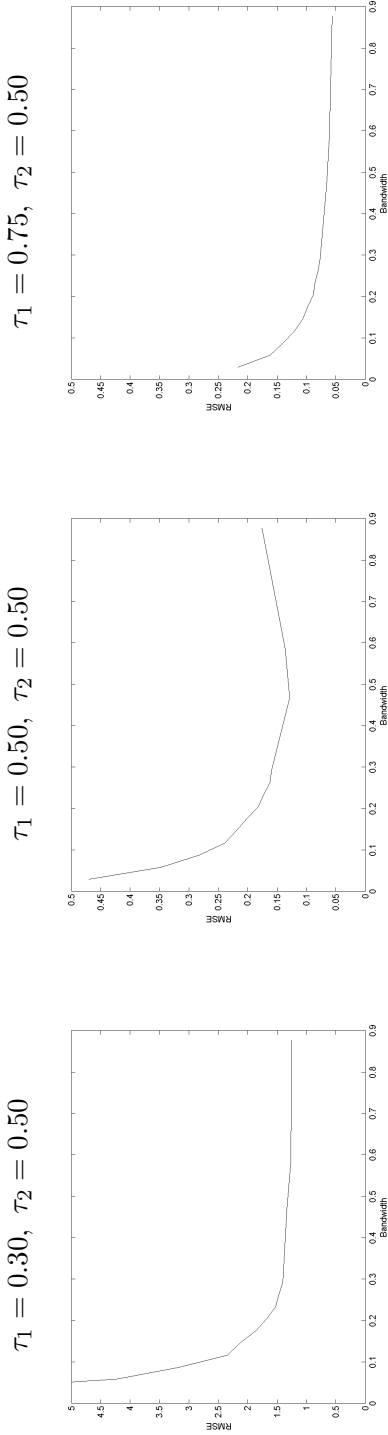$\tau_1 = 0.50, \ \tau_2 = 0.50$

$\tau_1 = 0.75, \ \tau_2 = 0.50$

Figure III: Rejection Rates of t-statistic testing the truth with nominal size 5% for the various values of τ_1

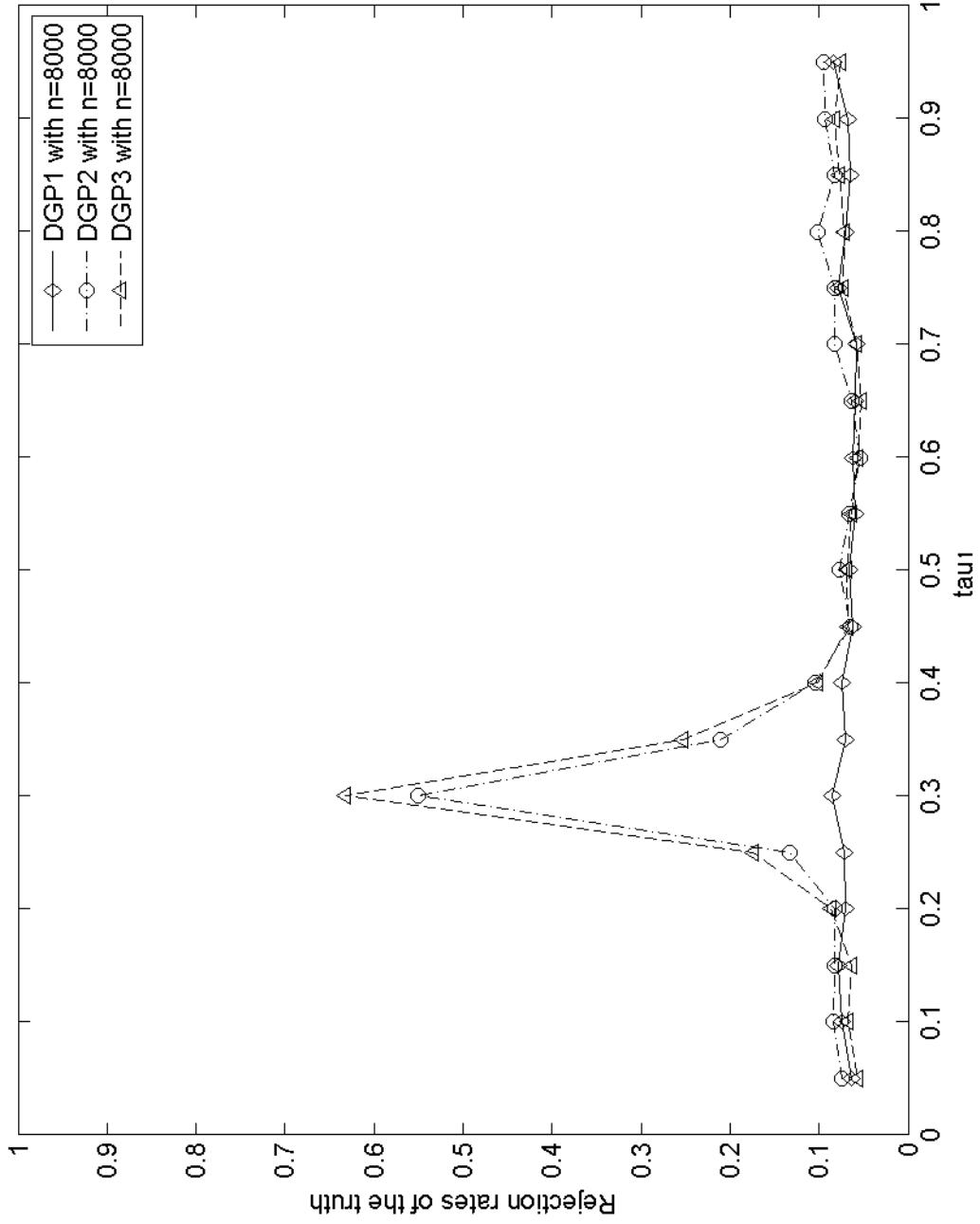Figure IV: Estimated Densities of t-statistics from Monte Carlo

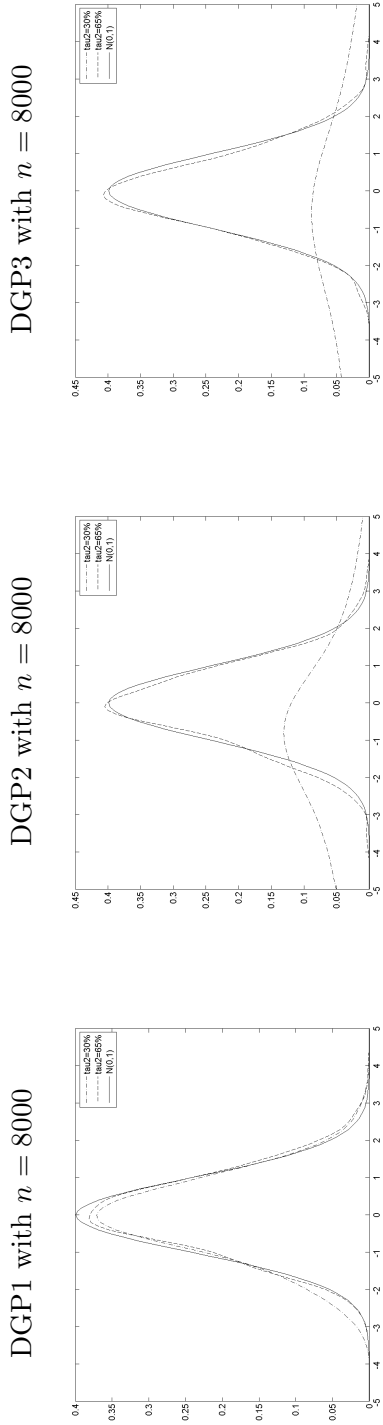DGP1 with $n = 8000$ DGP2 with $n = 8000$ DGP3 with $n = 8000$

Figure V: Examples of 95% Confidence Intervals

DGP1 with $n = 8000$ DGP2 with $n = 8000$ DGP3 with $n = 8000$