

INSTRUMENTAL VARIABLE QUANTILE REGRESSION WITH MISCLASSIFICATION*

TAKUYA URA^{*}
University of California, Davis

This article investigates the instrumental variable quantile regression model (Chernozhukov and Hansen, 2005, *Econometrica* 73, 245–261; 2013, *Annual Review of Economics*, 5, 57–81) with a binary endogenous treatment. It offers two identification results when the treatment status is not directly observed. The first result is that, remarkably, the reduced-form quantile regression of the outcome variable on the instrumental variable provides a lower bound on the structural quantile treatment effect under the stochastic monotonicity condition. This result is relevant, not only when the treatment variable is subject to misclassification, but also when any measurement of the treatment variable is not available. The second result is for the structural quantile function when the treatment status is measured with error; the sharp identified set is characterized by a set of moment conditions under widely used assumptions on the measurement error. Furthermore, an inference method is provided in the presence of other covariates.

1. INTRODUCTION

The instrumental variable quantile regression model (Chernozhukov and Hansen, 2005, 2013) aims to investigate heterogeneous treatment effects in the presence of an endogenous binary treatment variable. In many empirical applications, the treatment variable is potentially mismeasured, so it is empirically relevant how researchers can use the instrumental variable quantile regression model with a mismeasured treatment variable. For example, Chernozhukov and Hansen (2004) use the instrumental variable quantile regression model to investigate the quantile treatment effect of 401(k) participation on saving behaviors, but the pension plan

*First version: October, 2016. I am grateful to Peter C.B. Phillips, Arthur Lewbel, and four anonymous referees for comments and suggestions that have significantly improved this article. I would like to thank Alexandre Belloni, Stéphane Bonhomme, Federico A. Bugni, A. Colin Cameron, V. Joseph Hotz, Shakeeb Khan, Jia Li, Matthew A. Masten, Arnaud Maurel, Adam M. Rosen, and seminar participants at Duke, UC Berkeley, Shanghai University of Finance and Economics, Hakodate Conference in Econometrics, IAAE Annual Conference, Econometric Society North American Summer Meeting, and Econometric Society Asian Meeting for very helpful comments. The usual disclaimer applies. Address correspondence to Takuya Ura, Department of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616-5270, USA; e-mail: takura@ucdavis.edu.

© Cambridge University Press 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

type is subject to measurement error in survey datasets. Using the Health and Retirement Study, Gustman, Steinmeier, and Tabatabai (2008) estimate that around one-fourth of the survey respondents misclassified their pension plan type. To the best of my knowledge, however, no paper has investigated the instrumental variable quantile regression model when a binary regressor is potentially misclassified and endogenous.

This article has two identification results on the structural quantile function under the rank similarity condition.¹ The first identification result considers a reduced-form parameter, that is, the coefficient of the instrumental variable when running the quantile regression of the outcome variable on the instrumental variable. Under the rank similarity condition and the stochastic monotonicity condition (Small and Tan, 2007; DiNardo and Lee, 2011), this reduced-form parameter can be used as a lower bound for the structural quantile treatment effect. Although it has been used by empirical studies (e.g., Bitler, Hoynes, and Domina, 2016), the reduced-form quantile regression on the instrumental variable has not been formally related to the structural quantile treatment effect. Moreover, this result does not depend on the treatment variable or its measurement, and therefore it is relevant even when a measurement does not exist.

The second identification result is to derive moment conditions for the structural quantile function when the treatment variable is measured with exogenous errors. The exogeneity of the measurement error is widely assumed in the measurement error literature (e.g., Bound, Brown, and Mathiowetz, 2001) and yields exclusion restrictions similar to Henry, Kitamura, and Salanié (2014). Given the structure of the moment conditions, the structural quantile function can be under-identified even if the order condition for point identification holds. In other words, additional assumptions or variables are necessary to achieve point identification for the structural quantile function. As an example of additional restrictions, this article considers two observed measurements for one latent treatment variable, and point-identify the structural quantile function. Point identification results from combining two existing methods: misclassification correction techniques (Mahajan, 2006; Lewbel, 2007; Hu, 2008), and the identification results in Chernozhukov and Hansen (2005, 2013).

Based on the partial identification result, an inference method for the structural quantile function is provided by incorporating the misclassification probabilities in the inference method of Chernozhukov and Hansen (2008). The proposed inference method can include covariates other than the treatment variable,

¹Wüthrich (2019) investigates the instrumental variable quantile regression model without the rank similarity condition and characterizes the estimand of Chernozhukov and Hansen (2005) when the rank similarity condition fails. Dong and Shen (2018), Frandsen and Lefgren (2018), Kim and Park (2018), and Yu (2017) propose a test for the rank invariance or similarity condition.

and it is computationally feasible by imposing a linear-in-parameters structure on the structural quantile function. Simulation studies and an empirical illustration demonstrate the finite sample performance for the proposed inference method.

Related to this article, several papers have considered the problem of mismeasured regressors in the quantile regression framework, for example, Chesher (1991, 2017), Schennach (2008), Galvao and Montes-Rojas (2009), Wei and Carroll (2009), Firpo, Galvao, and Song (2017), and Song (2018). They focus on the case in which the mismeasured regressor is continuously distributed, whereas this article focuses on a binary treatment variable in which the measurement error has to be nonclassical.

Mahajan (2006), Lewbel (2007), and Hu (2008) consider identification of the average treatment effect (or, more generally, the conditional density function of the outcome variable given the true treatment variable) when a discrete treatment variable is mismeasured. Their identification strategy is based on the assumption that the true treatment variable (or the individual treatment effect in Lewbel, 2007) is exogenous, and there is no straightforward way to modify their results to the endogenous treatment.

Calvi, Lewbel, and Tommasi (2017), Yanagi (2019), and Ura (2018) investigate the local average treatment effect model with a mismeasured binary treatment. The local average treatment effect model is also a model for heterogeneous treatment effects in the presence of endogeneity but has a different structure than the instrumental variable quantile regression model.

Frazis and Loewenstein (2003), DiTraglia and García-Jimeno (2019), and Nguimkeu, Denteh, and Tchernis (2019) study a linear regression model in which a binary regressor is potentially misclassified and endogenous. Their approach is based on a homogenous treatment effect, which does not hold in the quantile treatment effect framework.

The remainder of this article is organized as follows. Section 2 introduces the instrumental variable quantile regression model (Chernozhukov and Hansen, 2005, 2013) with a misclassified treatment variable. Section 3 studies the reduced-form quantile regression of the outcome variable on the instrumental variable. Section 4 presents the identified set for the structural quantile function. Section 5 proposes an inference method based on the identification analysis in Section 4. Section 6 provides an empirical illustration and simulation studies. Section 7 concludes. The Appendix includes the proofs and additional results.

The rest of this article uses the following notations. Pr denotes the true probability measure for the observed and unobserved random variables. $Q_{RV_1|RV_2}(\tau)$ is the τ th conditional quantile of a continuous random variable RV_1 given a random variable RV_2 . $F_{RV_1|RV_2}$ is the conditional cumulative distribution function of a random variable RV_1 given a random variable RV_2 . $f_{RV_1|RV_2}$ is the conditional probability density function of a random variable RV_1 given a random variable RV_2 .

2. INSTRUMENTAL VARIABLE QUANTILE REGRESSION MODEL WITH MISCLASSIFICATION

This section presents notations and assumptions that are essentially those of the instrumental variable quantile regression model (Chernozhukov and Hansen, 2005, 2013), although the treatment variable D^* is not observed in this article. For the sake of simplicity, covariates are omitted other than the treatment variable when analyzing the identification problem. In the instrumental variable quantile regression model, Y is an outcome variable, D^* is a binary (true but latent) treatment variable taking values in $\{0, 1\}$, and Z is an instrumental variable. $D^* = 1$ means that the individual is treated; otherwise, $D^* = 0$.

In Chernozhukov and Hansen (2004) and Section 6.1, the outcome variable Y is the net amount of financial assets in dollars, the treatment variable D^* is the participation status in a 401(k) program, and the instrumental variable Z is the 401(k) eligibility indicator of whether an employer offers a 401(k) program to employees.

The goal of this article is to investigate treatment effects of D^* on Y . The error term (U_0, U_1) and the (unknown) structural quantile function $q(d^*, u)$ are used to model the relationship between the outcome variable Y and the binary treatment variable D^* :

$$Y = q(D^*, U) \text{ where } U = (1 - D^*)U_0 + D^*U_1.$$

The random variable $q(d^*, U_{d^*})$ is the potential outcome variable when $D^* = d^*$. The parameter of interest is the τ th quantile of the counterfactual outcome variable, $q(d^*, \tau)$, for a given value of $\tau \in (0, 1)$.

To identify $q(d^*, u)$ even partially, it is necessary to impose some structure on the unknown function, $q(d^*, u)$, and on the unobserved variable, U_{d^*} . The following assumptions are based on Chernozhukov and Hansen (2005, 2013). Unlike their papers, the condition is local at the given value of τ , which is sufficient for deriving the testable implication in Chernozhukov and Hansen (2005, Thm. 1) for the structural quantile function at τ .²

Assumption 1. *The mapping $u \mapsto q(d^*, u)$ is strictly increasing and left-continuous for every $u \in [0, 1]$, and has the inverse $y \mapsto q^{-1}(y, d^*)$.*

Assumption 2. (i) $Pr(U_0 \leq \tau | Z) = Pr(U_1 \leq \tau | Z) = \tau$. (ii) $Pr(U_0 \leq \tau | D^*, Z) = Pr(U_1 \leq \tau | D^*, Z)$.

Assumption 1 requires that the outcome variable Y is continuously distributed. Assumption 2(i) is the exogeneity of the instrumental variable Z . This assumption allows for the endogeneity of the treatment variable D^* . Assumption 2(ii) is the rank similarity condition on $U_{d^*} \leq \tau$. It is a relaxation of the rank invariance

²The local restriction is a weaker condition than full independence between Z and U_{d^*} (Chesher, 2003). Technically speaking, U_{d^*} 's are not necessarily uniform random variables in this article since the assumptions are only about the given value of τ , but the subsequent discussions in terms of the quantile treatment effect hold only when U_{d^*} 's are uniform random variables.

condition, $U_0 = U_1$, that an individual's rank, U_{d^*} , is the same regardless of whether the individual is treated or controlled. The rank similarity condition allows the two unobserved heterogeneity terms, U_1 and U_0 , to be different, although it still requires them to have the same distribution given the endogenous treatment assignment and the instrumental variable. It enables the τ th quantile of the counterfactual outcome variable, $q(d^*, \tau)$, to be compared between the control group ($d^* = 0$) and the treatment group ($d^* = 1$). The rank similarity condition is a restriction on the unobserved heterogeneity in the outcome variable equation and has been widely used for investigating heterogeneous treatment effects (e.g., Doksum, 1974; Heckman, Smith, and Clements, 1997; Chernozhukov and Hansen, 2004).

Under the rank similarity condition, Chernozhukov and Hansen (2005) obtain the following relationship between the distribution of (Y, D^*, Z) and the structural quantile function $q(d^*, \tau)$.

LEMMA 1 (Chernozhukov and Hansen, 2005, Thm. 1). *Under Assumptions 1 and 2,*

$$\Pr(U \leq \tau | Z) = \tau \text{ and } \Pr(Y \leq q(D^*, \tau) | Z) = \tau. \quad (1)$$

The rest of this article adds the complication that the binary treatment variable D^* may not be observed. Then the equality (1) cannot be directly used for identifying the structural quantile function.

3. QUANTILE REGRESSION OF Y ON Z

This section does not use any measurement of D^* ; instead the relationship between the outcome variable Y and the instrumental variable Z is used to provide a lower bound on the structural quantile treatment effect $q(1, \tau) - q(0, \tau)$. Namely, $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$ can be a lower bound on $q(1, \tau) - q(0, \tau)$, when Z is a binary variable taking z_0 and z_1 . This analysis provides a new structural interpretation to $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$, which is computed as the regression coefficient from the quantile regression of Y on Z . When Z takes more than two values, the discussion in this section can be applied by selecting any two values in the support of Z or partitioning the support into two parts. It is worth clarifying that the results in this section and the following sections are valid regardless of whether Z is binary, discrete, continuous, or mixed.

The result in this section uses the stochastic monotonicity condition (Small and Tan, 2007 and DiNardo and Lee, 2011). It assumes a positive relationship between the treatment variable D^* and the instrumental variable Z in which, for every possible realization u of U , the probability of being treated $f_{U_1, D^*|Z=z}(u, 1)$ is weakly increasing in z , and the probability of being untreated $f_{U_0, D^*|Z=z}(u, 0)$ is weakly decreasing in z . The condition is weaker than the deterministic monotonicity condition (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996)

because it allows for defiers, that is, some individuals who change D^* from 1 to 0 when Z increases.

Definition 1. *The stochastic monotonicity condition is that*

$$f_{U_0, D^*|Z=z_1}(u, 0) \leq f_{U_0, D^*|Z=z_0}(u, 0) \text{ and } f_{U_1, D^*|Z=z_1}(u, 1) \geq f_{U_1, D^*|Z=z_0}(u, 1) \tag{2}$$

for every $u \in [0, 1]$.

Theorem 1 shows that, under the stochastic monotonicity condition in (2), $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$ is biased toward zero compared with the structural quantile treatment effect $q(1, \tau) - q(0, \tau)$.

THEOREM 1. *Suppose that Assumptions 1 and 2 hold and that the stochastic monotonicity condition holds.*

(a) *There is some unknown constant $\kappa \in [0, 1]$ such that*

$$Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau) = \kappa \times (q(1, \tau) - q(0, \tau)). \tag{3}$$

(b) *If $f_{U_0, D^*|Z=z_1}(u, 0) < f_{U_0, D^*|Z=z_0}(u, 0)$ and $f_{U_1, D^*|Z=z_1}(u, 1) > f_{U_1, D^*|Z=z_0}(u, 1)$ in a neighborhood of τ , then $\kappa \neq 0$.*

This theorem provides a one-sided bound on $q(1, \tau) - q(0, \tau)$: $q(1, \tau) - q(0, \tau) \geq Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$ if $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau) \geq 0$; and $q(1, \tau) - q(0, \tau) \leq Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$ if $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau) \leq 0$. This bound gives researchers a justification for using $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$, which is a lower bound for the structural quantile treatment effect. Note that $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$ is a simple object to compute; it is obtained as the quantile regression coefficient on Z and various statistical software packages include linear and nonlinear quantile regressions (e.g., the `qreg` command in Stata).

The stochastic monotonicity condition cannot be removed from Theorem 1, but Lemma B.1 in the appendix shows that equation (3) holds with $\kappa \in [-1, 1]$ even if the stochastic monotonicity condition does not hold. In other words, the inequality $|Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)| \leq |q(1, \tau) - q(0, \tau)|$ still holds without the stochastic monotonicity condition. It is possible to use this inequality to test the significance of D^* by testing $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau) = 0$.

4. IDENTIFIED SET FOR THE STRUCTURAL QUANTILE FUNCTION

This section considers use of a potentially misclassified treatment variable D and provides the sharp identified set for the structural quantile function $q(\cdot, \tau)$. To extract some information about the true treatment D^* from its measurement D , the following restrictions on the misclassification probabilities are imposed.

Assumption 3. (i) $(Pr(D = 1 | D^* = 0, Y, Z), Pr(D = 0 | D^* = 1, Y, Z)) = (\pi_0, \pi_1)$ for some constants (π_0, π_1) . (ii) $Pr(D \neq D^* | D^* = 0, Y, Z) + Pr(D \neq D^* | D^* = 1, Y, Z) < 1$.

Assumption 3(i) is that the measurement error does not depend on (Y, Z) . Assumption 3(ii) is that the measurement D is positively correlated with the true treatment variable D^* . These assumptions are widely used in the literature on misclassification (e.g., Hausman, Abrevaya, and Scott-Morton, 1998, Mahajan, 2006; Lewbel, 2007; and Hu, 2008).

The sharp identified set for $q(\cdot, \tau)$ is the set of values for $q(\cdot, \tau)$ that exhausts all the information from the model and the data distribution. Let \mathcal{Q} be a subset of the set of functions of $\{0, 1\} \times [0, 1]$ to \mathbb{R} , and \mathcal{P}^* be a subset of the set of probability distributions for (D, Z, U_0, U_1, D^*) .³ Given a distribution P for (Y, D, Z) , the sharp identified set for (q, P^*) is the set of elements (\tilde{q}, \tilde{P}^*) of $\mathcal{Q} \times \mathcal{P}^*$ such that P is the distribution for $(\tilde{q}(D^*, U), D, Z)$ under \tilde{P}^* .⁴ The sharp identified set for $q(\cdot, \tau)$ is defined as the projection of the sharp identified set for (q, P^*) on the component $q(\cdot, \tau)$.

The following theorem characterizes the sharp identified set for $q(\cdot, \tau)$ under Assumptions 1, 2, and 3, by using moment equalities and inequalities.

THEOREM 2. *Assume that all the elements in $\mathcal{Q} \times \mathcal{P}^*$ satisfy Assumptions 1, 2, and 3. (a) Given a distribution P for the observed variables, if (y_0, y_1) belongs to the sharp identified set for $q(\cdot, \tau)$, then*

$$P(Y \leq y_D | Z) - \tau = p_1(P(Y \leq y_0 | Z) - \tau) + p_0(P(Y \leq y_1 | Z) - \tau) \tag{4}$$

for some (p_0, p_1) with $p_0 + p_1 < 1$ such that

$$0 \leq p_0 \leq P(D = 1 | Y, Z) \text{ a.s. and } 0 \leq p_1 \leq P(D = 0 | Y, Z) \text{ a.s.}$$

(b) *The converse is also true if $\mathcal{Q} \times \mathcal{P}^*$ includes all (q, P^*) 's satisfying Assumptions 1, 2, and 3.*

The moment equality condition in equation (4) is equivalent to the main testable implication in Chernozhukov and Hansen (2005) when $(p_0, p_1) = (\pi_0, \pi_1)$ where (π_0, π_1) are the true unknown misclassification probabilities defined in Assumption 3. The moment inequality conditions about (π_0, π_1) are derived from the following calculations:

³ \mathcal{Q} and \mathcal{P}^* are subsets because there can be restrictions on q and the distribution for (D, Z, U_0, U_1, D^*) . Also note that the distribution for (D, Z, U_0, U_1, D^*) can be characterized by

$$Pr(D = d, Z \leq z, U_0 \leq u_0, U_1 \leq u_1, D^* = d^*)$$

and the distribution for (Y, D, Z) can be characterized by

$$Pr(Y \leq y, D = d, Z \leq z).$$

The distribution P for (Y, D, Z) is induced by (q, P^*) via

$$P(Y \leq y, D = 0, Z \leq z) = P^*(D = 0, q(0, U_0) \leq y, D^* = 0, Z \leq z) + P^*(D = 0, q(1, U_1) \leq y, D^* = 1, Z \leq z)$$

$$P(Y \leq y, D = 1, Z \leq z) = P^*(D = 1, q(0, U_0) \leq y, D^* = 0, Z \leq z) + P^*(D = 1, q(1, U_1) \leq y, D^* = 1, Z \leq z).$$

⁴ In this article, \tilde{q} is a generic element of \mathcal{Q} and \tilde{P}^* is a generic element of \mathcal{P}^* , whereas q is the true structural quantile function and P^* is the true distribution for (D, Z, U_0, U_1, D^*) .

$$\begin{aligned} Pr(D = 1 \mid Y, Z) &= \pi_0 + (1 - \pi_0 - \pi_1)Pr(D^* = 1 \mid Y, Z) \geq \pi_0 \\ Pr(D = 0 \mid Y, Z) &= \pi_1 + (1 - \pi_0 - \pi_1)Pr(D^* = 0 \mid Y, Z) \geq \pi_1. \end{aligned}$$

As a corollary to Theorem 2, it is possible to compare the identified set for $q(\cdot, \tau)$ and the estimand in Chernozhukov and Hansen (2005), which does not consider measurement error in D .

Corollary 1. *Suppose all the assumptions in Theorem 2(b) hold. Every solution (y_0, y_1) to $P(Y \leq y_D \mid Z) - \tau = 0$, belongs to the sharp identified set for $q(\cdot, \tau)$.*

As another corollary to Theorem 2, it is possible to relate $Q_{Y|Z}(\tau)$ to the identified set for $q(\cdot, \tau)$. Although $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$ can be used as a lower bound for $q(1, \tau) - q(0, \tau)$, it does not always belong to the identified set.

Corollary 2. *Consider two points, z_0 and z_1 , in the support of Z , and suppose all the assumptions in Theorem 2 (b) hold. Then $(y_0, y_1) = (Q_{Y|Z=z_0}(\tau), Q_{Y|Z=z_1}(\tau))$ belongs to the sharp identified set for $q(\cdot, \tau)$ if and only if $P(D = 1 \mid \{y_0 < Y \leq y_1 \text{ or } y_1 < Y \leq y_0\}, Z = z_0) \leq P(D = 1 \mid Y, Z)$ a.s. and $P(D = 0 \mid \{y_0 < Y \leq y_1 \text{ or } y_1 < Y \leq y_0\}, Z = z_1) \leq P(D = 0 \mid Y, Z)$ a.s.*

Note that, by the exogeneity of Z , $(q(0, \tau), q(1, \tau)) = (Q_{Y|Z=z_0}(\tau), Q_{Y|Z=z_1}(\tau))$ if $D^* = 1\{Z = z_1\}$. Corollary 2 is roughly related to the observation that, under Assumption 3, $D^* = 1\{Z = z_1\}$ implies

$$\begin{aligned} P(D = 1 \mid \{y_0 < Y \leq y_1 \text{ or } y_1 < Y \leq y_0\}, Z = z_0) &= P(D = 1 \mid \{y_0 < Y \leq y_1 \text{ or } y_1 < Y \leq y_0\}, D^* = 0) \\ &= \pi_0 \\ &\leq \pi_0 + (1 - \pi_0 - \pi_1)P(D^* = 1 \mid Y, Z) \\ &= P(D = 1 \mid Y, Z) \\ P(D = 0 \mid \{y_0 < Y \leq y_1 \text{ or } y_1 < Y \leq y_0\}, Z = z_1) &= P(D = 0 \mid \{y_0 < Y \leq y_1 \text{ or } y_1 < Y \leq y_0\}, D^* = 1) \\ &= \pi_1 \\ &\leq \pi_1 + (1 - \pi_0 - \pi_1)P(D^* = 0 \mid Y, Z) \\ &= P(D = 0 \mid Y, Z). \end{aligned}$$

The precise derivations are found in the proof in the Appendix.

4.1. Under-Identification Even with Large Variation in Z

This section shows that the structural quantile function is not point identified in general unless there is additional information on the model primitives (q, P^*) . The failure of point identification is due to the rank condition and happens regardless of the order condition based on equation (4), where the number of the parameters is 4,

and the number of the equations is the number of the support points of Z . Theorem 3 presents this failure for a class of data generating processes. In particular, the theorem states that the quantile treatment effect can be under-identified if one cannot exclude that the treatment is exogenous. Based on the result, it is necessary to impose additional assumptions (other than the ones maintained in this article) to achieve point identification when the treatment variable is potentially misclassified.

THEOREM 3. Consider $\bar{d}^* = 0$ or $\bar{d}^* = 1$. Assume (i) the mapping $u \mapsto q^{-1}(q(1, u), 0)$ is Lipschitz continuous, (ii) $q(d^*, \tau) \neq q(1 - d^*, \tau)$, (iii) (U_0, U_1) is independent of (D^*, Z) , (iv) $\pi_{\bar{d}^*} > 0$, and (v) for sufficiently small $\varepsilon > 0$, the following three statements hold:

1. $\tilde{q} \in \mathcal{Q}$ for every strictly increasing bijection t of $[0, 1]$ to $[0, 1]$ such that $|t(u) - u| \leq \varepsilon$ for every $u \in (0, 1)$, where $\tilde{q}(d^*, \cdot) = q(d^*, t(\cdot))$ and $\tilde{q}(1 - d^*, \cdot) = q(1 - d^*, \cdot)$.
2. $\tilde{P}^* \in \mathcal{P}^*$ for every distribution \tilde{P}^* for (D, Z, U_0, U_1, D^*) such that \tilde{P}^* satisfies Assumption 2 and that

$$\tilde{P}^*(D = 1 - \bar{d}^* \mid U_0, U_1, D^* = \bar{d}^*, Z) = \pi_{\bar{d}^*} - \varepsilon$$

$$\tilde{P}^*(D = \bar{d}^* \mid U_0, U_1, D^* = 1 - \bar{d}^*, Z) = \pi_{1 - \bar{d}^*}$$

$$|\tilde{P}^*(U_0 \leq u_0, U_1 \leq u_1, D^* = d^*, Z \leq z)$$

$$- P^*(U_0 \leq u_0, U_1 \leq u_1, D^* = d^*, Z \leq z)| \leq \varepsilon.$$

Then the sharp identified set for $q(\cdot, \tau)$ has more than one element.

Condition (i) is a regularity condition. Condition (ii) is that the treatment variable can have a nonzero effect on the outcome variable at quantile index τ . Condition (iii) is that the treatment variable can be exogenous. Condition (iv) is that there is a nonzero measurement error. Condition (v) is a condition about the size of the parameter space $\mathcal{Q} \times \mathcal{P}^*$. A sufficient condition for (v) is that $\mathcal{Q} \times \mathcal{P}^*$ includes all (q, P^*) 's satisfying Assumptions 1, 2, and 3.

Condition (iii) needs a careful discussion. Theorem 3 states that the quantile treatment effect is not always point identified unless the treatment is assumed to be endogenous. This theorem is more relevant when one cannot exclude that D^* is exogenous than when D^* is known to be exogenous. When one cannot exclude that D^* is exogenous, there is a possibility for the lack of point identification. It can be possible to point-identify the quantile treatment function if one can assume that D^* is not exogenous.

4.2. Point Identification with Second Measurement

Given the under-identification result in Theorem 3, this section considers the case of two measurements for D^* to achieve point identification of the structural quantile

function. The identification strategy is based on existing results in the econometric literature. First, the results in Mahajan (2006), Lewbel (2007), and Hu (2008) are applied to identify $f_{Y, D^*|Z}$. Given identification of $f_{Y, D^*|Z}$, the identification result in Chernozhukov and Hansen (2005, 2013) recovers the structural quantile function.

The following assumption and Lemma 2 are based on Theorem 1 in Hu (2008).

Assumption 4. (i) *The two measurements, D and V , are conditionally independent given D^* .* (ii) $0 < f_{D^*|Z=z_1}(0) < 1$. (iii) *There are two points, v_0 and v_1 , in the support of V such that*

$$\begin{pmatrix} f_{V|D^*=0}(v_0) & f_{V|D^*=0}(v_1) \\ f_{V|D^*=1}(v_0) & f_{V|D^*=1}(v_1) \end{pmatrix}$$

is invertible. (iv) $P(D = D^* | D^*) > 1/2$. (v) *There are two points, z_0 and z_1 , in the support of Z .*

LEMMA 2. *Under Assumptions 3 and 4, $f_{(Y, D^*)|Z}$ is point identified.*

Chernozhukov and Hansen (2013) provide a simple sufficient condition for the global identification of the structural quantile function given $f_{(Y, D^*)|Z}$. The following assumption and identification result are borrowed from Chernozhukov and Hansen (2013, Sect. 3.1).

Assumption 5. *There is a cube \mathcal{L} with $(q(0, \tau), q(1, \tau)) \in \mathcal{L}$ such that*

$$\frac{f_{(Y, D^*)|Z=z_1}(y_1, 1)}{f_{(Y, D^*)|Z=z_1}(y_0, 0)} > \frac{f_{(Y, D^*)|Z=z_0}(y_1, 1)}{f_{(Y, D^*)|Z=z_0}(y_0, 0)}, f_{(Y, D^*)|Z=z_1}(y_1, 1) > 0 \text{ and } f_{(Y, D^*)|Z=z_0}(y_0, 0) > 0$$

for all $(y_0, y_1) \in \mathcal{L}$.

LEMMA 3. *Under Assumptions 1, 2, and 5, $(q(0, \tau), q(1, \tau))$ is uniquely determined from $f_{(Y, D^*)|Z}$.*

By Lemmas 2 and 3, the structural quantile function can be identified with two measurements for D^* .

THEOREM 4. *Under Assumptions 1, 2, 3, 4, and 5, $(q(0, \tau), q(1, \tau))$ is identified.*

5. INFERENCE PROCEDURE WITH COVARIATES

This section proposes an inference method for the structural quantile function. The method extends the inference method in Chernozhukov and Hansen (2008) to incorporate misclassification probabilities. To include control variables X , a linear-in-parameters structure is imposed on the structural quantile function:

$$Y = q(D^*, X, U) \text{ with } q(d^*, x, \tau) = \alpha_0 d^* + x' \beta_0. \tag{5}$$

This section focuses on constructing a confidence interval for α_0 .

With control variables X , Assumptions 1–3 are modified into the following assumptions:

Assumption 6. *With probability one, the mapping $u \mapsto q(d^*, X, u)$ is strictly increasing and left-continuous for every $u \in [0, 1]$.*

Assumption 7. (i) $Pr(U_0 \leq \tau \mid Z, X) = Pr(U_1 \leq \tau \mid Z, X) = \tau$. (ii) $Pr(U_0 \leq \tau \mid D^*, Z, X) = Pr(U_1 \leq \tau \mid D^*, Z, X)$.

Assumption 8. (i) For each $d^* = 0, 1$, $Pr(D \neq D^* \mid D^* = d^*, Y, Z, X)$ is a constant, denoted by π_{d^*} . (ii) $Pr(D \neq D^* \mid D^* = 0, Y, Z, X) + Pr(D \neq D^* \mid D^* = 1, Y, Z, X) < 1$.

Given n i.i.d. copies $\{(Y_i, D_i, X_i, Z_i) : i = 1, \dots, n\}$ of (Y, D, X, Z) , a confidence interval for α_0 is constructed via the following two steps. The first step constructs a confidence interval for (π_0, π_1) . Given each point in the confidence interval for (π_0, π_1) , the second step constructs a confidence interval for α_0 . The size control comes from the Bonferroni correction for the first and second steps.

The following condition is imposed on a $(1 - \text{size}_1)$ confidence region, CI_1 , for (π_0, π_1) .

Assumption 9. $\liminf_{n \rightarrow \infty} Pr((\pi_0, \pi_1) \in CI_1) \geq 1 - \text{size}_1$.

In the empirical illustration, CI_1 is constructed by inverting the one-tailed t -tests based on $\pi_0 \leq E[D \mid Z = 0]$ and $\pi_1 \leq E[1 - D \mid Z = 1]$, where $\text{size}_1/2$ for each t -test. In the empirical illustration, the confidence interval for α_0 is bounded by using $E_n[D \mid Z = 0] = 0$ and $E_n[1 - D \mid Z = 1] \approx 0.3$.

At the true value (α_0, π_0, π_1) of (α, p_0, p_1) , the following testable implications (cf. Chernozhukov and Hansen, 2008) hold. As in Chernozhukov and Hansen (2008), it is possible to replace Z with a function $g(X, Z)$ of (X, Z) .

LEMMA 4. *Under Assumptions 6–8,*

$$0 \in \arg \min_{\gamma} \min_{\beta} Q_0(\theta; \alpha_0, \pi_0, \pi_1).$$

where $v_0(y, x, z) = E[D \mid Y = y, X = x, Z = z]$, $t_+ = 1\{t \geq 0\} \cdot t$, $\rho_\tau(t) = (\tau - 1\{t \leq 0\})t$, $\theta = (\beta', \gamma')'$, $W = [X', Z']'$ and

$$\begin{aligned} Q_0(\theta; \alpha, p_0, p_1) &= E[\rho_\tau(Y - W'\theta)(1 - p_1 - D)] + E[\rho_\tau(Y - \alpha - W'\theta)(D - p_0)]. \\ &= E[\rho_\tau(Y - W'\theta)(1 - p_1 - v_0(Y, X, Z))_+] \\ &\quad + E[\rho_\tau(Y - \alpha - W'\theta)(v_0(Y, X, Z) - p_0)_+]. \end{aligned}$$

This article assumes that $Q_0(\theta; \alpha_0, \pi_0, \pi_1)$ has a unique minimizer over θ for the true parameter value (α_0, π_0, π_1) , which is implied by Assumption 12(1). Since (α_0, π_0, π_1) is unknown, an estimator for θ is computed as a function of (α, p_0, p_1) :

$$\hat{\theta}(\alpha; p_0, p_1) = \arg \min_{\theta \in \Theta} Q_n(\theta; \alpha, p_0, p_1),$$

where $\hat{v}(y, x, z)$ is an estimator for $v_0(y, x, z)$, and

$$Q_n(\theta; \alpha, p_0, p_1) = E_n[\rho_\tau(Y - W'\theta)(1 - p_1 - \hat{v}(Y, X, Z))_+] + E_n[\rho_\tau(Y - \alpha - W'\theta)(\hat{v}(Y, X, Z) - p_0)_+].$$

The objective function $Q_n(\theta; \alpha, p_0, p_1)$ is convex in θ , which is the result of using $\hat{v}(Y, X, Z)$ instead of D . This transformation comes from Buchinsky and Hahn (1998) and Abadie, Angrist, and Imbens (2002).⁵ To simplify the arguments, a parametric model $v_0(y, x, z) = v_{\delta_0}(y, x, z)$ is imposed with a parametric estimator $\hat{\delta}$ for δ_0 and the following assumptions.

Assumption 10. (i) $E[D^*|Y, X, Z]$ is bounded away from zero and one. (ii) $\sup_{(y,x,z)} |\hat{v}(y, x, z) - v_0(y, x, z)| = o_p(1)$ where $\hat{v}(y, x, z) = v_{\hat{\delta}}(y, x, z)$. (iii) $E_n \left[W(\xi_0(\hat{v}) - \xi_0(v_0) - \Xi_0(\delta_0)(\hat{\delta} - \delta_0)) \right] = o_p(n^{-1/2})$, where

$$\xi_0(v) = (\tau - 1\{Y - W'\theta_0 \leq 0\})(1 - \pi_1 - v(Y, X, Z)) + (\tau - 1\{Y - \alpha_0 - W'\theta_0 \leq 0\})(v(Y, X, Z) - \pi_0)$$

$$\Xi_0(\delta) = (1\{Y - W'\theta_0 \leq 0\} - 1\{Y - \alpha_0 - W'\theta_0 \leq 0\}) \frac{\partial}{\partial \delta} v_{\delta}(Y, X, Z).$$

(iv) There are n random variables, $\psi_{\delta, 1}, \dots, \psi_{\delta, n}$, such that $\hat{\delta} - \delta_0 = E_n[\psi_{\delta}] + o_p(n^{-1/2})$ with $E[\|\psi_{\delta}\|^2] < \infty$. (v) There is an estimator, $(\widehat{\psi}_{\delta, 1}, \dots, \widehat{\psi}_{\delta, n})$, for $(\psi_{\delta, 1}, \dots, \psi_{\delta, n})$, that satisfies $E[\|\widehat{\psi}_{\delta} - \psi_{\delta}\|^2] = o(1)$. (vi) There is an estimator $\frac{\partial}{\partial \delta} v_{\delta}$ for $\frac{\partial}{\partial \delta} v_{\delta}$ that satisfies $E \left[\left\| W \left(\widehat{\frac{\partial}{\partial \delta} v_{\delta}}(Y, X, Z) - \frac{\partial}{\partial \delta} v_{\delta}(Y, X, Z) \right) \right\|^2 \right] = o(1)$. (vii) $E \left[\left\| W \frac{\partial}{\partial \delta} v_{\delta}(Y, X, Z) \right\|^2 \right] < \infty$.

The optimization of $Q_n(\theta; \alpha, p_0, p_1)$ is implemented in the same way as the linear quantile regression. Namely, the objective function can be written as $Q_n(\theta; \alpha, p_0, p_1) = n^{-1} \sum_{i=1}^{2n} \rho_\tau(\check{Y}_i - \check{W}_i'\theta)$, where

$$\check{Y}_i = \begin{cases} Y_i(1 - p_1 - \hat{v}(Y_i, X_i, Z_i))_+ & \text{if } i \leq n \\ Y_{i-n}(\hat{v}(Y_{i-n}, X_{i-n}, Z_{i-n}) - p_0)_+ & \text{if } i \geq n + 1 \end{cases}$$

$$\check{W}_i = \begin{cases} W_i(1 - p_1 - \hat{v}(Y_i, X_i, Z_i))_+ & \text{if } i \leq n \\ W_{i-n}(\hat{v}(Y_{i-n}, X_{i-n}, Z_{i-n}) - p_0)_+ & \text{if } i \geq n + 1. \end{cases}$$

for $i = 1, \dots, n, n + 1, \dots, 2n$.⁶

⁵I am thankful to a referee for proposing this transformation.

⁶Note that $\rho_\tau(t)c = (\tau - 1\{t \leq 0\})tc = (\tau - 1\{ct \leq 0\})ct = \rho_\tau(ct)$ for every $t \in \mathbb{R}$ and $c \geq 0$. Since the weights, $(1 - p_1 - \hat{v}(Y, X, Z))_+$ and $(\hat{v}(Y, X, Z) - p_0)_+$, are non-negative,

$$\rho_\tau(Y - W'\theta)(1 - p_1 - \hat{v}(Y, X, Z))_+ = \rho_\tau(Y(1 - p_1 - \hat{v}(Y, X, Z))_+ - (W(1 - p_1 - \hat{v}(Y, X, Z))_+)' \theta)$$

The asymptotic variance for $\hat{\theta}(\alpha; p_0, p_1)$ is estimated with a kernel function $K(\cdot)$ and a bandwidth h . For every value of (α, p_0, p_1) , the asymptotic variance for $\hat{\theta}(\alpha; p_0, p_1)$ estimated by

$$\hat{\Omega}(\alpha; p_0, p_1) = E_n[\hat{\lambda}(\alpha; p_0, p_1)WW']^{-1}E_n[\hat{s}(\alpha; p_0, p_1)\hat{s}(\alpha; p_0, p_1)'] \\ E_n[\hat{\lambda}(\alpha; p_0, p_1)WW']^{-1},$$

where

$$\hat{\xi}(\alpha; p_0, p_1) = (\tau - 1\{Y - W'\hat{\theta}(\alpha; p_0, p_1) \leq 0\})(1 - p_1 - \hat{v}(Y, X, Z)) \\ + (\tau - 1\{Y - \alpha - W'\hat{\theta}(\alpha; p_0, p_1) \leq 0\})(\hat{v}(Y, X, Z) - p_0)$$

$$\hat{\Xi}(\alpha; p_0, p_1) = 1\{Y - W'\hat{\theta}(\alpha; p_0, p_1) \leq 0\} \frac{\partial}{\partial \delta} \hat{v}_\delta(Y, X, Z) \\ - 1\{Y - \alpha - W'\hat{\theta}(\alpha; p_0, p_1) \leq 0\} \frac{\partial}{\partial \delta} \hat{v}_\delta(Y, X, Z)$$

$$\hat{s}(\alpha; p_0, p_1) = \hat{\xi}(\alpha; p_0, p_1)W + E_n[W\hat{\Xi}(\alpha; p_0, p_1)]\hat{\psi}_\delta$$

$$\hat{\lambda}(\alpha; p_0, p_1) = K_h(Y - W'\hat{\theta}(\alpha; p_0, p_1))(1 - p_1 - D) \\ + K_h(Y - \alpha - W'\hat{\theta}(\alpha; p_0, p_1))(D - p_0).$$

Denote by $\hat{\Omega}_\gamma(\alpha; p_0, p_1)$, the asymptotic variance for $\hat{\gamma}(\alpha; p_0, p_1)$.

The proposed confidence interval for α is

$$CI_\alpha(\text{size}_1 + \text{size}_2) = \bigcup_{(p_0, p_1) \in CI_1} \{\alpha \in \mathcal{A} : T(\alpha; p_0, p_1) \leq cv\},$$

where \mathcal{A} is the parameter space for α , the test statistic is

$$T(\alpha; p_0, p_1) = n\hat{\gamma}(\alpha; p_0, p_1)' \hat{\Omega}_\gamma(\alpha; p_0, p_1)^{-1} \hat{\gamma}(\alpha; p_0, p_1),$$

and the critical value cv is the $(1 - \text{size}_2)$ quantile of the χ^2 distribution with $\dim(\gamma)$ degrees of freedom. The proposed confidence interval satisfies the asymptotic size control under the following assumptions.

Assumption 11. $\theta_0 \equiv (\beta_0', 0)'$ is in the interior of a compact parameter space Θ .

$$\rho_\tau(Y - \alpha - W'\theta)(\hat{v}(Y, X, Z) - p_0)_+ = \rho_\tau((Y - \alpha)(\hat{v}(Y, X, Z) - p_0)_+ - (W(\hat{v}(Y, X, Z) - p_0)_+)' \theta).$$

Therefore,

$$Q_n(\theta; \alpha, p_0, p_1) = E_n[\rho_\tau(Y - W'\theta)(1 - p_1 - \hat{v}(Y, X, Z))_+] + E_n[\rho_\tau(Y - \alpha - W'\theta)(\hat{v}(Y, X, Z) - p_0)_+] \\ = \frac{1}{n} \sum_{i=1}^{2n} \rho_\tau(\check{Y}_i - \check{W}'_i \theta).$$

Assumption 12. (i) $E[f_{Y-\alpha_0 D^*|Z,X}(W'\theta_0)WW']$ is invertible. (ii) $E[\|W\|^4]$ is finite. (iii) $\lim_{\epsilon \rightarrow 0} Pr(|Y - W'\theta_0| \leq \epsilon \cdot \|W\|) = 0$ and $\lim_{\epsilon \rightarrow 0} Pr(|Y - \alpha_0 - W'\theta_0| \leq \epsilon \cdot \|W\|) = 0$. (iv) There is a constant C such that $\max\{f_{Y|X,Z,D}, f_{Y|X,Z,D}^{(1)}\} < C$ a.s.

Assumption 13. (i) $h \rightarrow 0$ and $\sqrt{nh} \rightarrow \infty$ as $n \rightarrow \infty$. (ii) K is differentiable with $\sup_v |K^{(1)}(v)| < \infty$, $\int K(v)dv = 1$, $\int |K(v)v|dv < \infty$, and $\int K(v)^2 dv < \infty$.

THEOREM 5. Under Assumptions 6–13, $\liminf_{n \rightarrow \infty} Pr(\alpha_0 \in CI_\alpha(\text{size}_1 + \text{size}_2)) \geq 1 - (\text{size}_1 + \text{size}_2)$.

Assumption 11 is a regularity condition on the parameter. Assumption 12(i) is that the Hessian matrix is nonsingular, and it implies point identification of θ_0 given the true value (α_0, π_0, π_1) . Assumption 12(ii)–(iv) is a regularity condition on the distribution of the observables. Assumption 13 is a restriction on the bandwidth and the kernel function, which is used to estimate the asymptotic variance $\hat{\Omega}_\gamma(\alpha; p_0, p_1)$.

6. EMPIRICAL ILLUSTRATION AND MONTE CARLO SIMULATIONS

This section investigates the finite sample performance of the proposed method using an existing empirical application and simulated datasets. As emphasized in Section 4.1, the inference results presented in this section are valid regardless of whether the structural quantile function is point or partially identified.

6.1. Empirical Illustration

This empirical illustration studies the quantile treatment effects of the 401(k) participation on financial savings (Chernozhukov and Hansen, 2004) and consider the problem of mismeasured 401(k) participation.⁷ It uses the same dataset as Chernozhukov and Hansen (2004), which is an extract from the Survey of Income and Program Participation of 1991. The sample contains 9,915 households with at least one person in employment and without any income from self-employment.

The model is based on Chernozhukov and Hansen (2004). The outcome variable Y is the net amount of financial assets in dollars, the measured treatment variable D is self-reported participation in a 401(k) program, and the instrumental variable Z is 401(k) eligibility, an indicator variable of whether a 401(k) program is available in any of the companies for which the household member work. The summary statistics for (Y, D, Z) are in Table 1.

The details for the confidence intervals are as follows. The prespecified sizes are $(\text{size}_1, \text{size}_2) = (1\%, 4\%)$, where size_1 is used to construct a confidence interval for unknown misclassification probabilities (π_0, π_1) and size_2 is used for the

⁷Ura (2018) uses the same empirical setting to investigate the local average treatment effect under treatment misclassification.

TABLE 1. Summary statistics for Y , D , and Z

	Sample size	Mean	SD
Y : family net financial assets (in \$1,000)	9,915	18.05	63.52
D : 401(k) participation	9,915	0.26	0.44
Z : 401(k) eligibility	9,915	0.37	0.48

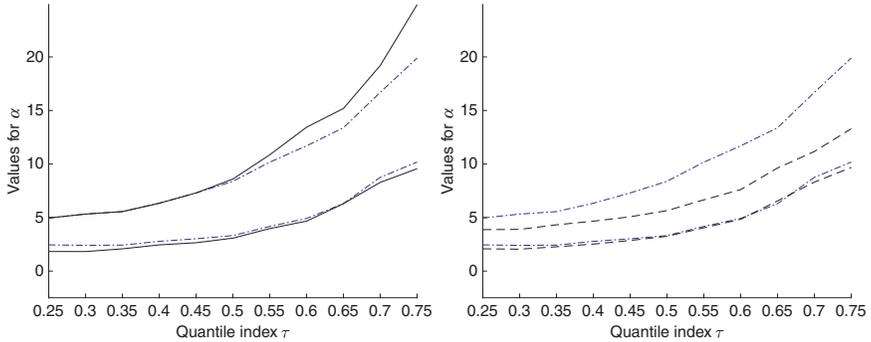


FIGURE 1. (Color online) Three point-wise 95% confidence intervals for α_0 . The solid curves represent the proposed inference method, the dash-dotted line (— · —) curves represent the inference method that assumes $\pi_0 = \pi_1 = 0$, and the dashed (---) curves represent the inference method that assumes $(\pi_0, \pi_1) = (0, 0.31)$.

critical value cv . The confidence interval for (π_0, π_1) is $CI_1 = \{0\} \times [0, 0.31]$ for (π_0, π_1) , where $\{0\}$ follows from $E_n[D | Z = 0] = 0$ and $[0, 0.31]$ comes from the one-tailed t -test for $\pi_1 \leq E[1 - D | Z = 1]$ with size 0.5%. The conditional probability of D given (Y, X, Z) is estimated by probit regression of D on all the interactions of $(1, X, Z)$ and the cubic polynomials of Y .⁸ Using $p_0 = 0$ and grid points $p_1 = 0, 0.01, 0.02, \dots, 0.31$ for CI_1 , $CI_\alpha(\text{size}_1 + \text{size}_2)$ is constructed by $\bigcup_{p_1=0, 0.01, 0.02, \dots, 0.31} \{\alpha \in \mathcal{A} : T(\alpha; p_0, p_1) \leq cv\}$.

Figure 1 (left) shows the 95% confidence intervals based on Section 4, with the specification in equation (5) and the same list of covariates as Benjamin (2003) and Chernozhukov and Hansen (2004).⁹ The figure also shows the 95% confidence intervals that assume $\pi_0 = \pi_1 = 0$, that is, no misclassification. The confidence intervals that assume $\pi_0 = \pi_1 = 0$ are exactly the same as the confidence intervals proposed in Chernozhukov and Hansen (2008, Sect. 3). All the confidence intervals are point-wise, that is, computed separately for each quantile index τ .

⁸The interaction terms can be written as $(1, X', Z', Y, Y \cdot X', Y \cdot Z', Y^2, Y^2 \cdot X', Y^2 \cdot Z', Y^3, Y^3 \cdot X', Y^3 \cdot Z')$.

⁹ X includes a constant, age, age², income categories, family size, education dummies, marital status dummy, two-earner status dummy, defined benefit pension status dummy, individual retirement account participation status dummy, and homeownership status dummy. The number of the covariates in X is 18.

In this empirical exercise, the confidence intervals proposed in this article are comparable in lengths to those that assume no misclassification. There are several features of this empirical exercise that make the confidence intervals tight (compared with those that assume no misclassification). First, in this empirical exercise, $E_n[D | Z = z_0] = 0$ implies $\pi_0 = 0$, so under-identification for $q(\cdot, \tau)$ mainly comes from under-identification for the scalar parameter π_1 . It makes the degree of under-identification smaller than when π_0 and π_1 are both under-identified. Second, the confidence intervals that assume $\pi_0 = \pi_1 = 0$, are large enough to include the confidence interval with another value of (p_0, p_1) . Figure 1 (right) shows the confidence intervals that assume $(p_0, p_1) = (0, 0)$ and that with $(p_0, p_1) = (0, 0.31)$, where the points $(0, 0)$ and $(0, 0.31)$ are the two endpoints for CI_1 . The confidence intervals that assume $(p_0, p_1) = (0, 0)$ include those that assume $(p_0, p_1) = (0, 0.31)$ for almost all the values of τ .

6.2. Monte Carlo Simulations

Monte Carlo simulations are based on the following data generating process. The instrumental variable Z takes z_0 with probability 0.5 and z_1 with probability 0.5. The error term (\tilde{U}, V) is a two-dimensional mean-zero normal random vector with variance 1 and correlation coefficient 0.5, independent of Z . The covariate X is $X = (1, \tilde{X})$, where \tilde{X} is a four-dimensional standard normal random vector with identity covariance matrix, independent of (Z, \tilde{U}, V) . The latent treatment variable D^* is determined by

$$D^* = 1\{1\{Z = z_1\} + X'\beta + V > 0\},$$

and the outcome variable Y is determined by

$$Y = \exp(\Phi(\tilde{U}) - 0.5)D^* + X'\beta + \tilde{U},$$

where Φ is the standard normal cumulative distribution function and $\beta = (-(2 \times 1)^{-1}, -(2 \times 2)^{-1}, \dots, -(2 \times 5)^{-1})'$. The binary measurement D is determined by $Pr(D \neq D^* | D^*) = \pi_{D^*}$, where $(\pi_0, \pi_1) = (0, 0), (0.1, 0), (0, 0.1), (0.2, 0), (0.1, 0.1), (0, 0.2), (0.2, 0.1), (0.1, 0.2)$. The true values for α_0 is $\exp(\tau - 0.5)$.

In this simulation exercise, coverage frequencies are computed for two inference methods over $[\alpha_0 - 1, \alpha_0 + 1]$. One is the proposed inference method and the other is the inference method that assumes no misclassification, that is, $\pi_0 = \pi_1 = 0$. All the results are based on $n = 1,000$, $(size_1, size_2) = (1\%, 4\%)$, and 5,000 simulations.¹⁰

¹⁰The details about the confidence interval is as follows. $CI_1 = [0, \bar{c}_0] \times [0, \bar{c}_1]$ where the value of \bar{c}_0 comes from the one-tailed t -test for $\pi_0 \leq E[D | Z = 0]$ with size 0.5% and the value of \bar{c}_1 comes from the one-tailed t -test for $\pi_1 \leq E[1 - D | Z = 1]$ with size 0.5%. As in the empirical exercise, $v_0(y, x, z)$ is estimated by the probit regression of D on all the interactions of $(1, X, Z)$ and the cubic polynomials of Y . Using 1% grid points $(p_0, p_1) \in \{0, 0.01, 0.02, \dots, \bar{c}_0\} \times \{0, 0.01, 0.02, \dots, \bar{c}_1\}$, $CI_\alpha(size_1 + size_2)$ is $\bigcup_{p_1=0, 0.01, 0.02, \dots, \bar{c}_0} \bigcup_{p_0=0, 0.01, 0.02, \dots, \bar{c}_1} \{\alpha \in \mathcal{A} : T(\alpha; p_0, p_1) \leq cv\}$.

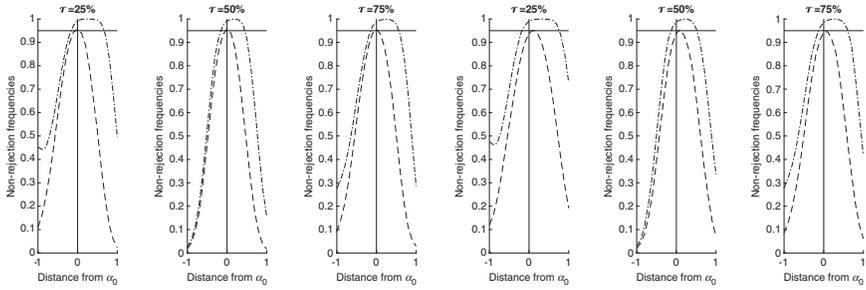


Figure 2.a: $\pi_0 = \pi_1 = 0$.

Figure 2.b: $\pi_0 = 0.1$ and $\pi_1 = 0$.

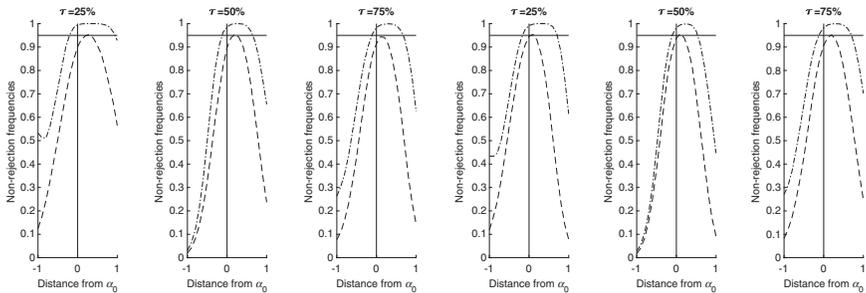


Figure 2.c: $\pi_0 = 0$ and $\pi_1 = 0.1$.

Figure 2.d: $\pi_0 = 0.2$ and $\pi_1 = 0$.

FIGURE 2. Coverage frequencies. The dash-dot (–.) curve represents the proposed inference method, and the dashed (—) curve represents the inference method that assumes $\pi_0 = \pi_1 = 0$.

Figure 2.a summarizes the simulation results when there is no misclassification, that is, $(\pi_0, \pi_1) = (0, 0)$. In this case, both the proposed method and the method that assumes $\pi_0 = \pi_1 = 0$ have correct size, that is, the coverage frequency at the true value of α is at least 95%. The proposed inference method is less powerful than that with $\pi_0 = \pi_1 = 0$, but this is the cost for achieving robustness to misclassification.

Figure 2.b–h summarizes the simulation results when there is some misclassification. The method that assumes $\pi_0 = \pi_1 = 0$ does not have correct size as (π_0, π_1) becomes far from $(0, 0)$, but the proposed method always has correct size. This is consistent with Theorem 5, which shows that the proposed method has correct size even in the presence of misclassification.

To summarize these simulation results, the proposed inference method covers the true parameter value at least with the prespecified significance level in finite samples. A practitioner could obtain a narrower confidence interval by assuming no misclassification, but the confidence interval may not cover the true parameter with correct size when there is non-negligible misclassification.

Additional simulation results, which provide a comparison between the proposed inference method and the infeasible method with knowing $(p_0, p_1) = (\pi_0, \pi_1)$, are provided in the online [Supplementary Material](#) associated with this article.

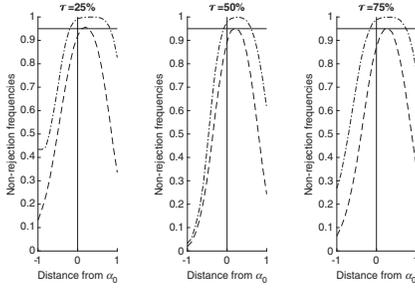


Figure 2.e: $\pi_0 = \pi_1 = 0.1$.

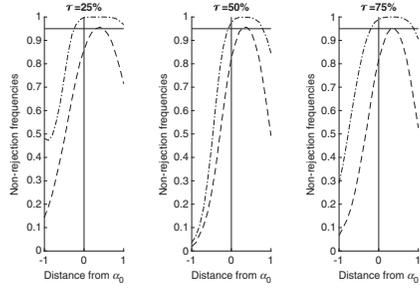


Figure 2.f: $\pi_0 = 0$ and $\pi_1 = 0.2$.

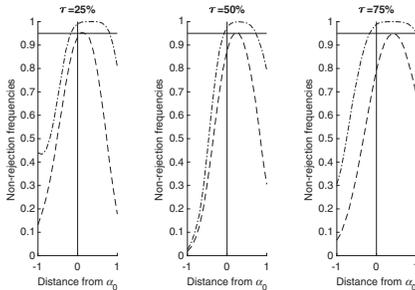


Figure 2.g: $\pi_0 = 0.2$ and $\pi_1 = 0.1$.

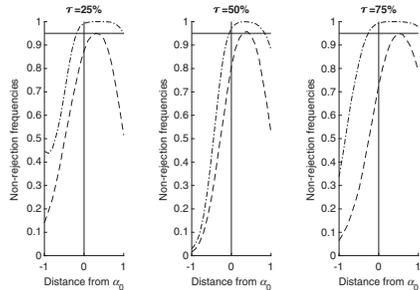


Figure 2.h: $\pi_0 = 0.1$ and $\pi_1 = 0.2$.

FIGURE 2 (continued): Coverage frequencies. The dash-dot (— · —) curve represents the proposed inference method, and the dashed (---) curve represents the inference method that assumes $\pi_0 = \pi_1 = 0$.

7. CONCLUSION

This article extends the instrumental variable quantile regression model (Chernozhukov and Hansen, 2005, 2013) for a binary regressor, to situations when this binary regressor is potentially misclassified. The first identification result is that under the rank similarity condition and the stochastic monotonicity condition, the reduced-form question effect, $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau)$, is biased toward zero compared with the structural quantile treatment effect $q(1, \tau) - q(0, \tau)$. The second identification result characterizes the sharp identified set for $q(d^*, \tau)$ under widely used assumptions. An inference method for the structural quantile function is provided, and its finite sample performance is demonstrated in simulation studies and an empirical illustration.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S026646662000002X>

REFERENCES

- Abadie, A., J. Angrist, & G. Imbens (2002) Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70, 91–117.
- Angrist, J.D., G.W. Imbens, & D.B. Rubin (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.
- Benjamin, D.J. (2003) Does 401(k) eligibility increase saving? Evidence from propensity score subclassification. *Journal of Public Economics* 87, 1259–1290.
- Bitler, M.P., H.W. Hoynes, & T. Domina (2016) Experimental Evidence on Distributional Effects of Head Start, Working paper.
- Bound, J., C. Brown, & N. Mathiowetz (2001) Measurement error in survey data. In J. Heckman and E. Leamerin (eds.) *Handbook of Econometrics*, vol. 5, chap. 59, pp. 3705–3843. Elsevier.
- Buchinsky, M. & J. Hahn (1998) An alternative estimator for the censored quantile regression model. *Econometrica* 66, 653.
- Calvi, R., A. Lewbel, & D. Tommasi (2017) LATE with Mismeasured or Misspecified Treatment: An Application To Women's Empowerment in India, Working paper.
- Chernozhukov, V. & C. Hansen (2004) The effects of 401(K) participation on the wealth distribution: An instrumental quantile regression analysis. *Review of Economics and Statistics* 86, 735–751.
- Chernozhukov, V. & C. Hansen (2005) An IV model of quantile treatment effects. *Econometrica* 73, 245–261.
- Chernozhukov, V. & C. Hansen (2008) Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics* 142, 379–398.
- Chernozhukov, V. & C. Hansen (2013) Quantile models with endogeneity. *Annual Review of Economics* 5, 57–81.
- Chesher, A. (1991) The effect of measurement error. *Biometrika* 78, 451–462.
- Chesher, A. (2003) Identification in nonseparable models. *Econometrica* 71, 1405–1441.
- Chesher, A. (2017) Understanding the effect of measurement error on quantile regressions. *Journal of Econometrics* 200, 223–237.
- DiNardo, J. & D.S. Lee (2011) Program evaluation and research designs. In O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, vol. 4, part A, chap. 5, pp. 463–536, Elsevier.
- DiTraglia, F.J. & C. García-Jimeno (2019) On mis-measured binary regressors: New results and some comments on the literature. *Journal of Econometrics* 209, 376–390.
- Doksum, K. (1974) Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Annals of Statistics* 2, 267–277.
- Dong, Y. & S. Shen (2018) Testing for rank invariance or similarity in program Evaluation. *Review of Economics and Statistics* 100, 78–85.
- Firpo, S., A.F. Galvao, & S. Song (2017) Measurement Errors in Quantile Regression Models. *Journal of Econometrics* 198, 146–164.
- Frandsen, B.R. & L.J. Lefgren (2018) Testing rank similarity. *Review of Economics and Statistics* 100, 86–91.
- Frazis, H. & M.A. Loewenstein (2003) Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables. *Journal of Econometrics* 117, 151–178.
- Galvao, A.F. & G. Montes-Rojas (2009) Instrumental Variables Quantile Regression for Panel Data with Measurement Errors, Working paper.
- Gustman, A.L., T. Steinmeier, & N. Tabatabai (2008) Do workers know about their pension plan type? Comparing workers' and employers' pension information. In A. Lusardi (ed.) *Overcoming the Saving Slump; How to Increase the Effectiveness of Financial Education and Saving Programs*. University of Chicago Press, Chicago, IL, pp. 47–81.
- Hausman, J.A., J. Abrevaya, & F.M. Scott-Morton (1998) Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* 87, 239–269.
- Heckman, J.J., J. Smith, & N. Clements (1997) Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies* 64, 487–535.

- Henry, M., Y. Kitamura, & B. Salanié (2014) Partial identification of finite mixtures in econometric models. *Quantitative Economics* 5, 123–144.
- Hu, Y. (2008) Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics* 144, 27–61.
- Imbens, G.W. & J.D. Angrist (1994) Identification and estimation of local average treatment effects. *Econometrica* 62, 467–75.
- Kim, J.H. & B. Park (2018) Testing Rank Similarity in the Local Average Treatment Effect Model, Working paper.
- Lewbel, A. (2007) Estimation of average treatment effects with misclassification. *Econometrica* 75, 537–551.
- Mahajan, A. (2006) Identification and estimation of regression models with misclassification. *Econometrica* 74, 631–665.
- Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (eds.) *Handbook of Econometrics*, vol. 4, chap. 36, pp. 2111–2245. Elsevier.
- Ngumkeu, P., A. Denteh, & R. Tchernis (2019) On the estimation of treatment effects with endogenous misreporting. *Journal of Econometrics* 208, 487–506.
- Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
- Schennach, S.M. (2008) Quantile regression with mismeasured covariates. *Econometric Theory* 24, 1010–1043.
- Small, D.S. & Z. Tan (2007) A Stochastic Monotonicity Assumption for the Instrumental Variables Method, Working paper.
- Song, S. (2018) Nonseparable Triangular Models with Errors in Endogenous Variables, Working paper.
- Ura, T. (2018) Heterogeneous treatment effects with mismeasured endogenous treatment. *Quantitative Economics* 9, 1335–1370.
- Wei, Y. & R.J. Carroll (2009) Quantile regression with measurement error. *Journal of the American Statistical Association* 104, 1129–1143.
- Wüthrich, K. (2019) A comparison of two quantile models with endogeneity. *Journal of Business and Economic Statistics*, forthcoming, <https://doi.org/10.1080/07350015.2018.1514307>.
- Yanagi, T. (2019) Inference on local average treatment effects for misclassified treatment. *Econometric Reviews* 38, 938–959.
- Yu, P. (2017) Testing Conditional Rank Similarity With and Without Covariates, Working paper.

APPENDIX

Appendix A discusses the identified set when the outcome variable is discrete, and Appendix B provides the proofs for the results in the main text.

APPENDIX A: IDENTIFIED SET WITH A DISCRETE OUTCOME VARIABLE

This appendix demonstrates how to modify Theorem 2 when the outcome variable is discrete. Assumptions 1 and 2 are modified into the following two conditions.

Assumption A.1. *The mapping $u \rightarrow q(d^*, u)$ is weakly increasing and left-continuous for every $u \in [0, 1]$.*

Assumption A.2. *(i) $Pr(U_0 \leq \tau | Z) \geq \tau$ and $Pr(U_1 \leq \tau | Z) \geq \tau$. (ii) $Pr(U_0 \leq \tau | D^*, Z) = Pr(U_1 \leq \tau | D^*, Z)$.*

Under the above two assumptions, Theorem 2 can be modified as follows:

THEOREM 6. Assume that all the elements in $\mathcal{Q} \times \mathcal{P}^*$ satisfy Assumptions 3, A.1, and A.2. (a) Given a distribution P for the observed variables, if (y_0, y_1) belongs to the sharp identified set for $q(\cdot, \tau)$, then

$$P(Y \leq y_D | Z) - \tau \geq p_1(P(Y \leq y_0 | Z) - \tau) + p_0(P(Y \leq y_1 | Z) - \tau) \tag{A.1}$$

for some (p_0, p_1) with $p_0 + p_1 < 1$ such that $0 \leq p_0 \leq P(D = 1 | Y, Z)$ a.s. and $0 \leq p_1 \leq P(D = 0 | Y, Z)$ a.s. (b) The converse is also true if $\mathcal{Q} \times \mathcal{P}^*$ includes all (q, P^*) 's satisfying Assumptions 3, A.1, and A.2.

Proof. The proof for (a) is as follows. Define $(p_0, p_1) = (\pi_0, \pi_1)$. By Assumption 3,

$$\begin{pmatrix} P^*(D^* = 0 | Y, Z) \\ P^*(D^* = 1 | Y, Z) \end{pmatrix} = (1 - p_0 - p_1)^{-1} \begin{pmatrix} P(D = 0 | Y, Z) - p_1 \\ P(D = 1 | Y, Z) - p_0 \end{pmatrix}. \tag{A.2}$$

Using equation (A.2),

$$P(Y \leq y_D | Z) - p_1 P(Y \leq y_0 | Z) - p_0 P(Y \leq y_1 | Z) = (1 - p_0 - p_1)P(Y \leq y_{D^*} | Z).$$

Since Assumption (A.1) implies $P(Y \leq y_{D^*} | Z) = P(q(D^*, U) \leq q(D^*, \tau) | Z) \geq P(U \leq \tau | Z) = P(U_0 \leq \tau | Z) \geq \tau$, it follows that $P(Y \leq y_D | Z) - p_1 P(Y \leq y_0 | Z) - p_0 P(Y \leq y_1 | Z) \geq (1 - p_0 - p_1)\tau$, and then equation (A.1) holds. Moreover, by equation (A.2), $P^*(D^* = d^* | Y, Z) \geq 0$ implies $P(D = 0 | Y, Z) \geq p_1$ and $P(D = 1 | Y, Z) \geq p_0$.

In the proof for (b), it is necessary to find $(\tilde{q}, \tilde{P}^*) \in \mathcal{Q} \times \mathcal{P}^*$ such that $\tilde{q}(d^*, \tau) = y_{d^*}$ and that P is the distribution for (Y, D, Z) under \tilde{P}^* . For each $d^* = 0, 1$, there is a strictly increasing bijection $t_{d^*} : [0, 1] \rightarrow [0, 1]$ such that $t_{d^*}(\tau) = F_{Y|D=d^*}(y_{d^*})$. For each $d^* = 0, 1$ and every $u \in [0, 1]$, define $\tilde{q}(d^*, u) = Q_{Y|D=d^*}(t_{d^*}(u))$. Define the distribution \tilde{P}^* for (D, Z, U_0, U_1, D^*) by

$$\tilde{P}^*(D = 1 - d^* | Z, U_0, U_1, D^*) = p_{D^*}$$

$$\tilde{P}^*(Z \leq z, U_0 \leq u_0, U_1 \leq u_1, D^* = d^*) = \tilde{P}^*(Y \leq q(d^*, \min\{u_0, u_1\}), Z \leq z, D^* = d^*),$$

where

$$\begin{pmatrix} \tilde{P}^*(Y \leq y, Z \leq z, D^* = 0) \\ \tilde{P}^*(Y \leq y, Z \leq z, D^* = 1) \end{pmatrix} = \begin{pmatrix} 1 - p_0 & p_1 \\ p_0 & 1 - p_1 \end{pmatrix}^{-1} \begin{pmatrix} P(Y \leq y, Z \leq z, D = 0) \\ P(Y \leq y, Z \leq z, D = 1) \end{pmatrix}. \tag{A.3}$$

By construction, the distribution for (Y, D, Z) under (\tilde{q}, \tilde{P}^*) is P , and $\tilde{q}(d^*, \tau) = y_{d^*}$ for each $d^* = 0, 1$. To show $(\tilde{q}, \tilde{P}^*) \in \mathcal{Q} \times \mathcal{P}^*$, it suffices to show that Assumptions A.1 and 3 hold for (\tilde{q}, \tilde{P}^*) . The rest of the proof is going to show Assumption 2(i).¹¹ By rearranging equation (A.1),

$$\begin{aligned} \tau &\leq \frac{1 - p_1}{1 - p_0 - p_1} P(Y \leq y_0, D = 0 | Z) - \frac{p_1}{1 - p_0 - p_1} P(Y \leq y_0, D = 1 | Z) \\ &\quad - \frac{p_0}{1 - p_0 - p_1} P(Y \leq y_1, D = 0 | Z) + \frac{1 - p_0}{1 - p_0 - p_1} P(Y \leq y_1, D = 1 | Z). \end{aligned}$$

¹¹ Assumption A.1 follows from the definition of \tilde{q} . Assumption 2(ii) follows from the definition of \tilde{P}^* . Assumption 3 follows from $\tilde{P}^*(D \neq D^* | Z, U_0, U_1, D^*) = p_{D^*}$.

Using equation (A.3), $\tau \leq \tilde{P}^*(Y \leq y_0, D^* = 0 | Z) + \tilde{P}^*(Y \leq y_1, D^* = 1 | Z)$. The definition of \tilde{q} and \tilde{P}^* implies

$$\begin{aligned} \tilde{P}^*(U_0 \leq \tau | Z) &= \tilde{P}^*(Y \leq y_0, D^* = 0 | Z) + \tilde{P}^*(Y \leq y_1, D^* = 1 | Z) \geq \tau \\ \tilde{P}^*(U_1 \leq \tau | Z) &= \tilde{P}^*(Y \leq y_0, D^* = 0 | Z) + \tilde{P}^*(Y \leq y_1, D^* = 1 | Z) \geq \tau, \end{aligned}$$

which is Assumption 2(i). □

APPENDIX B: PROOFS OF THE RESULTS IN THE MAIN TEXT

Proof of Theorem 1

LEMMA B.1. *Under Assumptions 1 and 2, $Q_{Y|Z}(\tau)$ is a convex combination of $q(1, \tau)$ and $q(0, \tau)$.*

Proof. By the monotonicity of $q(\cdot, d^*)$ and Lemma 1,

$$\begin{aligned} \tau &= P(Y \leq q(D^*, \tau) | Z) \geq P(Y \leq \min\{q(1, \tau), q(0, \tau)\} | Z) = F_{Y|Z}(\min\{q(1, \tau), q(0, \tau)\}) \\ \tau &= P(Y \leq q(D^*, \tau) | Z) \leq P(Y \leq \max\{q(1, \tau), q(0, \tau)\} | Z) = F_{Y|Z}(\max\{q(1, \tau), q(0, \tau)\}). \end{aligned}$$

Since $Q_{Y|Z}(\cdot)$ is monotonic, it follows that $\min\{q(1, \tau), q(0, \tau)\} \leq Q_{Y|Z}(\tau) \leq \max\{q(1, \tau), q(0, \tau)\}$. □

By Lemma B.1, the statement of theorem holds if $q(1, \tau) = q(0, \tau)$. The rest of the proof is going to focus on $q(1, \tau) > q(0, \tau)$. By Lemma B.1, there is some $\kappa \in [-1, 1]$ such that $Q_{Y|Z=z_1}(\tau) - Q_{Y|Z=z_0}(\tau) = \kappa(q(1, \tau) - q(0, \tau))$. Here, it is sufficient to show $Q_{Y|Z=z_1}(\tau) \geq Q_{Y|Z=z_0}(\tau)$ for Theorem 1(a) and $Q_{Y|Z=z_1}(\tau) > Q_{Y|Z=z_0}(\tau)$ for Theorem 1(b). Since $Q_{Y|Z}(\tau)$ is a convex combination of $q(1, \tau)$ and $q(0, \tau)$, it follows $q(1, \tau) \geq Q_{Y|Z}(\tau) \geq q(0, \tau)$. Since

$$\begin{aligned} P(Y \leq Q_{Y|Z}(\tau), D^* = 0 | Z) + P(Y \leq Q_{Y|Z}(\tau), D^* = 1 | Z) & \\ = F_{Y|Z}(Q_{Y|Z}(\tau)) & \\ = \tau & \\ = P(q(D^*, U) \leq q(D^*, \tau) | Z) & \\ = P(q(0, U) \leq q(0, \tau), D^* = 0 | Z) + P(q(1, U) \leq q(1, \tau), D^* = 1 | Z), & \end{aligned}$$

it follows that $P(q(0, \tau) < q(0, U) \leq Q_{Y|Z}(\tau), D^* = 0 | Z) = P(Q_{Y|Z}(\tau) < q(1, U) \leq q(1, \tau), D^* = 1 | Z)$. Using the monotonicity of $u \mapsto q(d^*, u)$, $P(\tau < U \leq q^{-1}(Q_{Y|Z}(\tau), 0), D^* = 0 | Z) = P(q^{-1}(Q_{Y|Z}(\tau), 1) < U \leq \tau, D^* = 1 | Z)$. Using the density function $f_{U_{d^*}, D^*|Z}(y, d^*)$, the above equation can be rewritten as

$$\int_{\tau}^{q^{-1}(Q_{Y|Z}(\tau), 0)} f_{U_0, D^*|Z}(u, 0) du = \int_{q^{-1}(Q_{Y|Z}(\tau), 1)}^{\tau} f_{U_1, D^*|Z}(u, 1) du. \tag{B.1}$$

The first half of this proof is going to show Theorem 1(a). Evaluate equation (B.1) at $Z = z_0$ and then

$$\int_{\tau}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} f_{U_0, D^*|Z=z_0}(u, 0)du = \int_{q^{-1}(Q_{Y|Z=z_0}(\tau), 1)}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} f_{U_1, D^*|Z=z_0}(u, 1)du.$$

Since $f_{U_0, D^*|Z=z_1}(u, 0) \leq f_{U_0, D^*|Z=z_0}(u, 0)$ and $f_{U_1, D^*|Z=z_1}(u, 1) \geq f_{U_1, D^*|Z=z_0}(u, 1)$,

$$\int_{\tau}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} f_{U_0, D^*|Z=z_1}(u, 0)du \leq \int_{q^{-1}(Q_{Y|Z=z_0}(\tau), 1)}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} f_{U_1, D^*|Z=z_1}(u, 1)du.$$

Subtracting the above equation from equation (B.1) at $Z = z_1$, and then

$$\int_{q^{-1}(Q_{Y|Z=z_1}(\tau), 0)}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} f_{U_0, D^*|Z=z_1}(u, 0)du \leq \int_{q^{-1}(Q_{Y|Z=z_0}(\tau), 1)}^{q^{-1}(Q_{Y|Z=z_1}(\tau), 1)} f_{U_1, D^*|Z=z_1}(u, 1)du.$$

Using $Y = q(D^*, U)$,

$$\int_{Q_{Y|Z=z_1}(\tau)}^{Q_{Y|Z=z_0}(\tau)} f_{Y, D^*|Z=z_1}(y, 0)dy \leq \int_{Q_{Y|Z=z_0}(\tau)}^{Q_{Y|Z=z_1}(\tau)} f_{Y, D^*|Z=z_1}(y, 1)dy,$$

and then $\int_{Q_{Y|Z=z_0}(\tau)}^{Q_{Y|Z=z_1}(\tau)} f_{Y, D^*|Z=z_1}(u)du \geq 0$, which implies $\tau - F_{Y, D^*|Z=z_1}(Q_{Y|Z=z_0}(\tau)) \geq 0$.

The second half of the proof is going to show Theorem 1(b). To the contrary, suppose $Q_{Y|Z=z_0}(\tau) \leq Q_{Y|Z=z_1}(\tau)$. By Theorem 1(a), $Q_{Y|Z=z_0}(\tau) = Q_{Y|Z=z_1}(\tau)$. Using $\Delta f_{U_{d^*}, D^*|Z}(u, d^*) = f_{U_{d^*}, D^*|Z=z_1}(u, d^*) - f_{U_{d^*}, D^*|Z=z_0}(u, d^*)$, equation (B.1) implies

$$\int_{\tau}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} \Delta f_{U_0, D^*|Z}(u, 0)du = \int_{q^{-1}(Q_{Y|Z=z_0}(\tau), 1)}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} \Delta f_{U_1, D^*|Z}(u, 1)du.$$

Since $f_{U_0, D^*|Z=z_1}(u, 0) \leq f_{U_0, D^*|Z=z_0}(u, 0)$ and $f_{U_1, D^*|Z=z_1}(u, 1) \geq f_{U_1, D^*|Z=z_0}(u, 1)$,

$$\int_{\tau}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} |\Delta f_{U_0, D^*|Z}(u, 0)|du = - \int_{q^{-1}(Q_{Y|Z=z_0}(\tau), 1)}^{q^{-1}(Q_{Y|Z=z_0}(\tau), 0)} |\Delta f_{U_1, D^*|Z}(u, 1)|du.$$

Since $f_{U_0, D^*|Z=z_1}(u, 0) < f_{U_0, D^*|Z=z_0}(u, 0)$ and $f_{U_1, D^*|Z=z_1}(u, 1) > f_{U_1, D^*|Z=z_0}(u, 1)$ in a neighborhood of τ , the above equation implies $q^{-1}(Q_{Y|Z=z_0}(\tau), 0) = \tau = q^{-1}(Q_{Y|Z=z_0}(\tau), 1)$, which contradicts $q(1, \tau) > q(0, \tau)$.

Proof of Theorem 2

The proof for (a) is as follows. By Assumption 3,

$$\left(\begin{matrix} P^*(D^* = 0 | Y, Z) \\ P^*(D^* = 1 | Y, Z) \end{matrix} \right) = (1 - \pi_0 - \pi_1)^{-1} \left(\begin{matrix} P(D = 0 | Y, Z) - \pi_1 \\ P(D = 1 | Y, Z) - \pi_0 \end{matrix} \right).$$

Equation (1) in Lemma 1 becomes equation (4). Moreover, $P^*(D^* = d^* | Y, Z) \geq 0$ implies $P(D = 0 | Y, Z) \geq \pi_1$ and $P(D = 1 | Y, Z) \geq \pi_0$.

In the proof for (b), it is necessary to find $(\tilde{q}, \tilde{P}^*) \in \mathcal{Q} \times \mathcal{P}^*$ such that $\tilde{q}(d^*, \tau) = y_{d^*}$ and that P is the distribution for (Y, D, Z) under \tilde{P}^* . For each $d^* = 0, 1$, there is a strictly increasing bijection $t_{d^*}: [0, 1] \rightarrow [0, 1]$ such that $t_{d^*}(\tau) = F_{Y|D=d^*}(y_{d^*})$. For each $d^* = 0, 1$ and every $u \in [0, 1]$, define $\tilde{q}(d^*, u) = Q_{Y|D=d^*}(t_{d^*}(u))$. Define the distribution

\tilde{P}^* for (D, Z, U_0, U_1, D^*) by

$$\tilde{P}^*(D \neq D^* \mid Z, U_0, U_1, D^*) = p_{D^*}$$

$$\tilde{P}^*(Z \leq z, U_0 \leq u_0, U_1 \leq u_1, D^* = d^*) = \tilde{P}^*(Y \leq q(d^*, \min\{u_0, u_1\}), Z \leq z, D^* = d^*),$$

where

$$\begin{pmatrix} \tilde{P}^*(Y \leq y, Z \leq z, D^* = 0) \\ \tilde{P}^*(Y \leq y, Z \leq z, D^* = 1) \end{pmatrix} = \begin{pmatrix} 1-p_0 & p_1 \\ p_0 & 1-p_1 \end{pmatrix}^{-1} \begin{pmatrix} P(Y \leq y, Z \leq z, D = 0) \\ P(Y \leq y, Z \leq z, D = 1) \end{pmatrix}. \tag{B.2}$$

By construction, the distribution for (Y, D, Z) under (\tilde{q}, \tilde{P}^*) is P , and $\tilde{q}(d^*, \tau) = y_{d^*}$ for each $d^* = 0, 1$. To show $(\tilde{q}, \tilde{P}^*) \in \mathcal{Q} \times \mathcal{P}^*$, it suffices to show that Assumptions 1, 2, and 3 hold for (\tilde{q}, \tilde{P}^*) . The rest of the proof is going to show Assumption 2(i). By rearranging equation (4),

$$\begin{aligned} \tau &= \frac{1-p_1}{1-p_0-p_1} P(Y \leq y_0, D = 0 \mid Z) - \frac{p_1}{1-p_0-p_1} P(Y \leq y_0, D = 1 \mid Z) \\ &\quad - \frac{p_0}{1-p_0-p_1} P(Y \leq y_1, D = 0 \mid Z) + \frac{1-p_0}{1-p_0-p_1} P(Y \leq y_1, D = 1 \mid Z). \end{aligned}$$

Using equation (B.2), $\tau = \tilde{P}^*(Y \leq y_0, D^* = 0 \mid Z) + \tilde{P}^*(Y \leq y_1, D^* = 1 \mid Z)$. The definition of \tilde{q} and \tilde{P}^* implies

$$\tilde{P}^*(U_0 \leq \tau \mid Z) = \tilde{P}^*(Y \leq y_0, D^* = 0 \mid Z) + \tilde{P}^*(Y \leq y_1, D^* = 1 \mid Z) = \tau$$

$$\tilde{P}^*(U_1 \leq \tau \mid Z) = \tilde{P}^*(Y \leq y_0, D^* = 0 \mid Z) + \tilde{P}^*(Y \leq y_1, D^* = 1 \mid Z) = \tau.$$

Proof of Corollary 1

Use $(p_0, p_1) = (0, 0)$. Then (y_0, y_1) satisfies the conditions in Theorem 2.

Proof of Corollary 2

In this proof, assume $Q_{Y|Z=z_0}(\tau) \leq Q_{Y|Z=z_1}(\tau)$ without loss of generality. The ‘‘only if’’ part of this corollary is shown as follows. By Theorem 2, equation (4) holds for some (p_0, p_1) with $0 \leq p_0 \leq P(D = 1 \mid Y, Z)$ a.s. and $0 \leq p_1 \leq P(D = 0 \mid Y, Z)$ a.s. Using $(y_0, y_1) = (Q_{Y|Z=z_0}(\tau), Q_{Y|Z=z_1}(\tau))$, equation (4) becomes

$$\begin{aligned} P(y_0 < Y \leq y_1, D = 1 \mid Z = z_0) &= p_0 P(y_0 < Y \leq y_1 \mid Z = z_0) \\ -P(y_0 < Y \leq y_1, D = 0 \mid Z = z_1) &= -p_1 P(y_0 < Y \leq y_1 \mid Z = z_1). \end{aligned}$$

Since $p_0 \leq P(D = 1 \mid Y, Z)$ a.s. and $p_1 \leq P(D = 0 \mid Y, Z)$ a.s., it follows that $P(D = 1 \mid y_0 < Y \leq y_1, Z = z_0) \leq P(D = 1 \mid Y, Z)$ a.s. and $P(D = 0 \mid y_0 < Y \leq y_1, Z = z_1) \leq P(D = 0 \mid Y, Z)$ a.s.

The ‘‘if’’ part of this corollary is shown by choosing $p_0 = P(D = 1 \mid y_0 < Y \leq y_1, Z = z_0)$ and $p_1 = P(D = 0 \mid y_0 < Y \leq y_1, Z = z_1)$.

Proof of Theorem 3

Assume $\bar{d}^* = 0$ for simplicity. Note that $q(d^*, u) = Q_{Y|D^*=d^*, Z}(u)$ from Condition (iii). Take sufficiently small $\varepsilon > 0$ and define

$$(\tilde{q}(0, u), \tilde{q}(1, u)) = \left(q \left(0, u + \frac{\varepsilon}{1 - \pi_0 - \pi_1} (u - P^*(Y \leq q(1, u) \mid D^* = 0)) \right), q(1, u) \right)$$

$$(p_0, p_1) = (\pi_0 - \varepsilon, \pi_1).$$

Condition (iv) guarantees $p_0 \geq 0$. Consider $U_0 = U_1$ and define \tilde{P}^* by

$$\tilde{P}^*(\tilde{q}(0, U_0) \leq y, D^* = 0, Z \leq z) = \frac{1 - \pi_0 - \pi_1}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(Y \leq y, D^* = 0, Z \leq z)$$

$$\tilde{P}^*(\tilde{q}(1, U_1) \leq y, D^* = 1, Z \leq z) = P^*(Y \leq y, D^* = 1, Z \leq z)$$

$$+ \frac{\varepsilon}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(Y \leq y, D^* = 0, Z \leq z)$$

$$\tilde{P}^*(D = 0 \mid Z, U_1, D^* = 1) = \pi_1$$

$$\tilde{P}^*(D = 1 \mid Z, U_0, D^* = 0) = \pi_0 - \varepsilon.$$

By Condition (ii), $u \neq P^*(Y \leq q(1, u) \mid D^* = 0)$ and $\tilde{q}(0, \tau) \neq q(0, u)$ as long as ε is positive. To establish the statement of this theorem, the rest of the proof is going to show that $(\tilde{q}, \tilde{P}^*) \in \mathcal{Q} \times \mathcal{P}^*$ and that (\tilde{q}, \tilde{P}^*) is observationally equivalent to (q, P^*) .

First, $(\tilde{q}, \tilde{P}^*) \in \mathcal{Q} \times \mathcal{P}^*$. The Lipschitz continuity in Condition (i) guarantees that $u \mapsto t(u)$ is a strictly increasing bijection of $[0, 1]$ into $[0, 1]$ for sufficiently small ε , where $t(u) = u + \frac{\varepsilon}{1 - \pi_0 - \pi_1} (u - P^*(Y \leq q(1, u) \mid D^* = 0))$. Therefore, Condition (v) implies $\tilde{q} \in \mathcal{Q}$ for sufficiently small ε . To show $\tilde{P}^* \in \mathcal{P}^*$, it suffices to show Assumption 2(i) because Assumption 2(ii) holds for $U = U_0 = U_1$. By the definition of \tilde{q} and independence between Z and Y given D^* ,

$$P^*(Y \leq \tilde{q}(0, \tau), D^* = 0 \mid Z) = P^*(D^* = 0 \mid Z) P^*(Y \leq \tilde{q}(0, \tau) \mid D^* = 0)$$

$$= P^*(D^* = 0 \mid Z) \left(\tau + \frac{\varepsilon}{1 - \pi_0 - \pi_1} (\tau - P^*(Y \leq q(1, \tau) \mid D^* = 0)) \right)$$

$$= P^*(D^* = 0 \mid Z) \left(\tau + \frac{\varepsilon}{1 - \pi_0 - \pi_1} (\tau - P^*(Y \leq q(1, \tau) \mid D^* = 0, Z)) \right),$$

where the second equality uses $q(0, u) = Q_{Y \mid D^*=0}(u)$. By the definition of \tilde{P}^* and \tilde{q} ,

$$\tilde{P}^*(U \leq \tau \mid Z) = \frac{1 - \pi_0 - \pi_1}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(D^* = 0 \mid Z)$$

$$\times \left(\tau + \frac{\varepsilon}{1 - \pi_0 - \pi_1} (\tau - P^*(Y \leq q(1, \tau) \mid D^* = 0, Z)) \right)$$

$$+ P^*(Y \leq q(1, \tau), D^* = 1 \mid Z)$$

$$+ \frac{\varepsilon}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(Y \leq q(1, \tau), D^* = 0 \mid Z)$$

$$= \tau P^*(D^* = 0 \mid Z) + P^*(Y \leq q(1, \tau), D^* = 1 \mid Z)$$

$$= \tau,$$

where the last equality uses $q(1, u) = Q_{Y \mid D^*=1, Z}(u)$.

Second, (\tilde{q}, \tilde{P}^*) is observationally equivalent to (q, P^*) . By the definition of \tilde{P}^* ,

$$\begin{aligned} \tilde{P}^*(Y \leq y, D^* = 1, Z \leq z) &= \tilde{P}^*(\tilde{q}(1, U) \leq y, D^* = 1, Z \leq z) = P^*(Y \leq y, D^* = 1, Z \leq z) \\ &\quad + \frac{\varepsilon}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(Y \leq y, D^* = 0, Z \leq z) \\ \tilde{P}^*(Y \leq y, D^* = 0, Z \leq z) &= \tilde{P}^*(\tilde{q}(0, U) \leq y, D^* = 0, Z \leq z) \\ &= \frac{1 - \pi_0 - \pi_1}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(Y \leq y, D^* = 0, Z \leq z). \end{aligned}$$

Therefore,

$$\begin{aligned} \tilde{P}(Y \leq y, D = 0, Z \leq z) &= \tilde{P}^*(D = 0 \mid Z, U_0, U_1, D^* = 0) \tilde{P}(Y \leq y, D^* = 0, Z \leq z) \\ &\quad + \tilde{P}^*(D = 0 \mid Z, U_0, U_1, D^* = 1) \tilde{P}(Y \leq y, D^* = 1, Z \leq z) \\ &= (1 - \pi_0 + \varepsilon) \frac{1 - \pi_0 - \pi_1}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(Y \leq y, D^* = 0, Z \leq z) \\ &\quad + \pi_1 P^*(Y \leq y, D^* = 1, Z \leq z) \\ &\quad + \pi_1 \frac{\varepsilon}{1 - \pi_0 - \pi_1 + \varepsilon} P^*(Y \leq y, D^* = 0, Z \leq z) \\ &= (1 - \pi_0) P^*(Y \leq y, D^* = 0, Z \leq z) \\ &\quad + \pi_1 P^*(Y \leq y, D^* = 1, Z \leq z) \\ &= P^*(Y \leq y, D = 0, Z \leq z) \end{aligned}$$

and it can be similarly shown that $\tilde{P}(Y \leq y, D = 1, Z \leq z) = P^*(Y \leq y, D = 1, Z \leq z)$.

Proof of Lemma 2

Under Assumption 4(i),

$$f_{(D, V) \mid Z}(d, v) = (f_{D \mid D^*=0, Z}(d) f_{D \mid D^*=1, Z}(d)) \begin{pmatrix} f_{D^* \mid Z}(0) & 0 \\ 0 & f_{D^* \mid Z}(1) \end{pmatrix} \begin{pmatrix} f_{V \mid D^*=0}(v) \\ f_{V \mid D^*=1}(v) \end{pmatrix}.$$

Under Assumption 3,

$$\begin{aligned} &\begin{pmatrix} f_{(D, V) \mid Z}(0, v_0) & f_{(D, V) \mid Z}(0, v_1) \\ f_{(D, V) \mid Z}(1, v_0) & f_{(D, V) \mid Z}(1, v_1) \end{pmatrix} \\ &= \begin{pmatrix} 1 - \pi_0 & \pi_1 \\ \pi_0 & 1 - \pi_1 \end{pmatrix} \begin{pmatrix} f_{D^* \mid Z}(0) & 0 \\ 0 & f_{D^* \mid Z}(1) \end{pmatrix} \begin{pmatrix} f_{V \mid D^*=0}(v_0) & f_{V \mid D^*=0}(v_1) \\ f_{V \mid D^*=1}(v_0) & f_{V \mid D^*=1}(v_1) \end{pmatrix}. \end{aligned}$$

Under Assumption 4(ii) and (iii), the above matrix for $Z = z_1$ is invertible, so

$$\begin{aligned} &\begin{pmatrix} f_{(D, V) \mid Z=z_0}(0, v_0) & f_{(D, V) \mid Z=z_0}(0, v_1) \\ f_{(D, V) \mid Z=z_0}(1, v_0) & f_{(D, V) \mid Z=z_0}(1, v_1) \end{pmatrix} \begin{pmatrix} f_{(D, V) \mid Z=z_1}(0, v_0) & f_{(D, V) \mid Z=z_1}(0, v_1) \\ f_{(D, V) \mid Z=z_1}(1, v_0) & f_{(D, V) \mid Z=z_1}(1, v_1) \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 1 - \pi_0 & \pi_1 \\ \pi_0 & 1 - \pi_1 \end{pmatrix} \begin{pmatrix} f_{D^* \mid Z=z_0}(0) / f_{D^* \mid Z=z_1}(0) & 0 \\ 0 & f_{D^* \mid Z=z_0}(1) / f_{D^* \mid Z=z_1}(1) \end{pmatrix} \\ &\times \begin{pmatrix} 1 - \pi_0 & \pi_1 \\ \pi_0 & 1 - \pi_1 \end{pmatrix}^{-1}. \end{aligned}$$

Under Assumption 4(iv), the eigenvalue decomposition of the above matrix is uniquely determined, so (π_0, π_1) is identified. Since $f_{(Y, D) \mid Z}(y, d) = \sum_{d^*=0, 1} f_{D \mid Y=y, D^*=d^*, Z}(d)$

$f_{(Y, D^*)|Z}(y, d^*)$, Assumption 3 implies

$$\begin{aligned} f_{(Y, D)|Z}(y, 0) &= \pi_0 f_{(Y, D^*)|Z}(y, 1) + (1 - \pi_0) f_{(Y, D^*)|Z}(y, 0) \\ f_{(Y, D)|Z}(y, 1) &= (1 - \pi_1) f_{(Y, D^*)|Z}(y, 1) + \pi_1 f_{(Y, D^*)|Z}(y, 0), \end{aligned}$$

so that $f_{(Y, D^*)|Z}$ is point identified.

Proof of Lemma 3

It follows from Theorem 2 of Chernozhukov and Hansen (2013).

Proof of Theorem 4

It follows from Lemmas 2 and 3.

Proof of Lemma 4

Note that

$$\begin{aligned} &(1 - \pi_0 - \pi_1)E[\rho_\tau(Y - \alpha_0 D^* - W'\theta)] \\ &= (1 - \pi_0 - \pi_1)E[\rho_\tau(Y - W'\theta)(1 - D^*)] + (1 - \pi_0 - \pi_1)E[\rho_\tau(Y - \alpha_0 - W'\theta)D^*] \\ &= (1 - \pi_1)E[\rho_\tau(Y - W'\theta)(1 - D)] - \pi_1 E[\rho_\tau(Y - W'\theta)D] \\ &\quad - \pi_0 E[\rho_\tau(Y - \alpha_0 - W'\theta)(1 - D)] + (1 - \pi_0)E[\rho_\tau(Y - \alpha_0 - W'\theta)D] \\ &= E[\rho_\tau(Y - W'\theta)(1 - \pi_1 - D)] + E[\rho_\tau(Y - \alpha_0 - W'\theta)(D - \pi_0)] \\ &= E[\rho_\tau(Y - W'\theta)(1 - \pi_1 - E[D | Y, X, Z])] + E[\rho_\tau(Y - \alpha_0 - W'\theta)(E[D | Y, X, Z] - \pi_0)], \end{aligned}$$

because

$$\begin{pmatrix} 1 - \pi_0 & \pi_1 \\ \pi_0 & 1 - \pi_1 \end{pmatrix}^{-1} = \frac{1}{1 - \pi_0 - \pi_1} \begin{pmatrix} 1 - \pi_1 & -\pi_1 \\ -\pi_0 & 1 - \pi_0 \end{pmatrix}.$$

Since $Pr(Y - \alpha_0 D^* \leq X'\beta_0 + 0 \cdot Z | X, Z) = Pr(q(D^*, X, U) \leq q(D^*, X, \tau) | X, Z) = \tau$, Chernozhukov and Hansen (2008, p. 383) derives $0 \in \arg \min_\gamma (\min_\beta E[\rho_\tau(Y - \alpha_0 D^* - W'\theta)])$. Therefore,

$$\begin{aligned} 0 \in \arg \min_\gamma \left(\min_\beta E[\rho_\tau(Y - W'\theta)(1 - \pi_1 - v_0(Y, X, Z))] \right. \\ \left. + E[\rho_\tau(Y - \alpha_0 - W'\theta)(v_0(Y, X, Z) - \pi_0)] \right). \end{aligned}$$

Note that $v_0(Y, X, Z) - \pi_0 = (v_0(Y, X, Z) - \pi_0)_+$ and $1 - \pi_1 - v_0(Y, X, Z) = (1 - \pi_1 - v_0(Y, X, Z))_+$, because

$$\begin{aligned} E[D | Y, X, Z] &= \pi_0 + (1 - \pi_0 - \pi_1)E[D^* | Y, X, Z] \geq \pi_0 \\ 1 - E[D | Y, X, Z] &= \pi_1 + (1 - \pi_0 - \pi_1)E[1 - D^* | Y, X, Z] \geq \pi_1. \end{aligned}$$

Proof of Theorem 5

By Lemmas B.6 and B.10 below, $P(T(\alpha_0; \pi_0, \pi_1) \leq cv) \rightarrow 1 - \text{size}_2$ as $n \rightarrow \infty$. Then the theorem follows from

$$\begin{aligned} P(\alpha_0 \in CI_\alpha(\text{size}_1 + \text{size}_2)) &\geq P(\{\alpha_0 \in CI_\alpha(\text{size}_1 + \text{size}_2)\} \cap \{(\pi_0, \pi_1) \in CI_1\}) \\ &\geq P(\{T(\alpha_0; \pi_0, \pi_1) \leq cv\} \cap \{(\pi_0, \pi_1) \in CI_1\}) \\ &\geq P(T(\alpha_0; \pi_0, \pi_1) \leq cv) - P((\pi_0, \pi_1) \notin CI_1) \end{aligned}$$

and then $\liminf_{n \rightarrow \infty} P(\alpha_0 \in CI_\alpha(\text{size}_1 + \text{size}_2)) \geq 1 - (\text{size}_1 + \text{size}_2)$.

The following proof assumes $1 - \pi_1 - \hat{v}(Y, X, Z) \geq 0$ and $\hat{v}(Y, X, Z) - \pi_0 \geq 0$ without loss of generality, because $E[D^* | Y, X, Z]$ is bounded away from zero and one, and then

$$\begin{aligned} 1 - \pi_1 - \hat{v}(Y, X, Z) &= (1 - \pi_0 - \pi_1)E[1 - D^* | Y, X, Z] - (\hat{v}(Y, X, Z) - \nu(Y, X, Z)) > 0 \\ \hat{v}(Y, X, Z) - \pi_0 &= (1 - \pi_0 - \pi_1)E[D^* | Y, X, Z] + (\hat{v}(Y, X, Z) - \nu(Y, X, Z)) > 0 \end{aligned}$$

with probability approaching one.

LEMMA B.2. $\sup_{\theta \in \Theta} |\bar{Q}_n(\theta) - \bar{Q}_0(\theta)| = o_p(1)$, where $\bar{Q}_n(\theta) = Q_n(\theta; \alpha_0, \pi_0, \pi_1)$ and $\bar{Q}_0(\theta) = Q_0(\theta; \alpha_0, \pi_0, \pi_1)$.

Proof. The proof of this lemma is to check the conditions in Newey and McFadden (1994, Lem. 2.9). Since Θ is compact and \bar{Q}_0 is continuous, it suffices to show $\bar{Q}_n(\theta) = \bar{Q}_0(\theta) + o_p(1)$ for every $\theta \in \Theta$ and $|\bar{Q}_n(\tilde{\theta}) - \bar{Q}_n(\theta)| \leq 2E_n[\|W\|] \cdot \|\tilde{\theta} - \theta\|$ for every $\tilde{\theta}, \theta \in \Theta$. This proof uses

$$\begin{aligned} \bar{Q}_n^*(\theta) &= E_n[\rho_\tau(Y - W'\theta)(1 - \pi_1 - \nu_0(Y, X, Z))_+] \\ &\quad + E_n[\rho_\tau(Y - \alpha_0 - W'\theta)(\nu_0(Y, X, Z) - \pi_0)_+]. \end{aligned}$$

The pointwise convergence of $\bar{Q}_n(\theta)$ to $\bar{Q}_0(\theta)$ is shown by demonstrating $\bar{Q}_n(\theta) - \bar{Q}_n^*(\theta) = o_p(1)$ and $\bar{Q}_n^*(\theta) - \bar{Q}_0(\theta) = o_p(1)$. Since

$$\begin{aligned} |\bar{Q}_n(\theta) - \bar{Q}_n^*(\theta)| &\leq E_n[\rho_\tau(Y - W'\theta)|\hat{v}(Y, X, Z) - \nu_0(Y, X, Z)|] \\ &\quad + E_n[\rho_\tau(Y - \alpha_0 - W'\theta)|\hat{v}(Y, X, Z) - \nu_0(Y, X, Z)|] \\ &\leq \sup_{(y, x, z)} |\hat{v}(y, x, z) - \nu_0(y, x, z)| \\ &\quad (E_n[|\rho_\tau(Y - W'\theta)| + |\rho_\tau(Y - \alpha_0 - W'\theta)|]), \end{aligned}$$

Assumption 10(ii) implies $\bar{Q}_n(\theta) - \bar{Q}_n^*(\theta) = o_p(1)$. Moreover, $\bar{Q}_n^*(\theta) - \bar{Q}_0(\theta) = \bar{Q}_n^*(\theta) - E[\bar{Q}_n^*(\theta)] = o_p(1)$ is shown by checking the second moment of $\bar{Q}_n^*(\theta)$: $E[\bar{Q}_n^*(\theta)^2] = O(1/n)$ follows from

$$\begin{aligned} &E\left[(\rho_\tau(Y - W'\theta)(1 - \pi_1 - D) + \rho_\tau(Y - \alpha_0 - W'\theta)(D - \pi_0))^2\right]^{1/2} \\ &\leq E\left[(Y - W'\theta)^2\right]^{1/2} + E\left[(Y - \alpha_0 - W'\theta)^2\right]^{1/2} \\ &\leq 2E\left[Y^2\right]^{1/2} + \alpha_0 + 2E\left[\|W\|^2\right]^{1/2} \|\theta\| \\ &< \infty. \end{aligned}$$

The in-probability Lipschitz condition is shown as follows. Since $|\rho_\tau(Y - W'\theta) - \rho_\tau(Y - W'\tilde{\theta})| \leq |W'(\tilde{\theta} - \theta)| \leq \|W\| \cdot \|\tilde{\theta} - \theta\|$ and $|\rho_\tau(Y - \alpha_0 - W'\theta) - \rho_\tau(Y - \alpha_0 - W'\tilde{\theta})| \leq \|W\| \cdot \|\tilde{\theta} - \theta\|$, it follows that $|\bar{Q}_n(\tilde{\theta}) - \bar{Q}_n(\theta)| \leq E_n[\|W\| \cdot \|\tilde{\theta} - \theta\|] + E_n[\|W\| \cdot \|\tilde{\theta} - \theta\|] \leq 2E_n[\|W\| \cdot \|\tilde{\theta} - \theta\|]$. \square

LEMMA B.3. $\hat{\theta}_0 - \theta_0 = o_p(1)$, where $\hat{\theta}_0 = \hat{\theta}(\alpha_0; \pi_0, \pi_1)$.

Proof. The proof of this lemma is to check the conditions in Newey and McFadden (1994, Thm. 2.1) to establish the consistency. Since Θ is compact, \bar{Q}_0 is continuous, and Lemma B.2 establishes the uniform convergence of $\bar{Q}_n(\theta)$, it suffices to show that \bar{Q}_0 is uniquely minimized at θ_0 . As in the proof of Lemma 4,

$$\begin{aligned} \bar{Q}_0(\theta) &= E[\rho_\tau(Y - W'\theta)(1 - \pi_1 - D) + \rho_\tau(Y - \alpha_0 - W'\theta)(D - \pi_0)] \\ &= (1 - \pi_0 - \pi_1)E[\rho_\tau(Y - \alpha_0 D^* - W'\theta)]. \end{aligned}$$

Since

$$\begin{aligned} \frac{\partial}{\partial \theta} \bar{Q}_0(\theta) &= (1 - \pi_0 - \pi_1) \frac{\partial}{\partial \theta} E[\rho_\tau(Y - \alpha_0 D^* - W'\theta)] \\ &= (1 - \pi_0 - \pi_1) E[(f_{Y-\alpha_0 D^*|X,Z}(W'\theta) - \tau) W] \\ \frac{\partial^2}{\partial \theta \partial \theta'} \bar{Q}_0(\theta) &= (1 - \pi_0 - \pi_1) E[f_{Y-\alpha_0 D^*|X,Z}(W'\theta) W W'], \end{aligned}$$

it follows that $\frac{\partial}{\partial \theta} \bar{Q}_0(\theta_0) = 0$ and $\frac{\partial^2}{\partial \theta \partial \theta'} \bar{Q}_0(\theta)$ is positive semidefinite everywhere and positive definite at θ_0 . Therefore, \bar{Q}_0 is uniquely minimized at θ_0 . \square

LEMMA B.4. $E_n[\xi_0(\hat{v})W] = E_n[s_0] + o_p(n^{-1/2})$, where $s_0 = \xi_0(v_0)W + E[W\Xi_0(\delta_0)]\psi_\delta$.

Proof. By Assumption 10(iii) and (iv), $E_n[\xi_0(\hat{v})W] = E_n[\xi_0(v_0)W] + E[W\Xi_0(\delta_0)]E_n[\psi_\delta] + o_p(n^{-1/2})$. \square

LEMMA B.5. Define

$$\begin{aligned} \lambda_0(\tau) &= f_{Y|D,Z,X}(W'\theta_0 - n^{-1/2}W'\tau)(1 - \pi_1 - D) \\ &\quad + f_{Y|D,Z,X}(\alpha_0 + W'\theta_0 - n^{-1/2}W'\tau)(D - \pi_0) \\ g(\tau, v) &= (\rho_\tau(Y - W'\theta_0 - n^{-1/2}W'\tau) - \rho_\tau(Y - W'\theta_0))(1 - \pi_1 - v(Y, X, Z)) + \\ &\quad + (\rho_\tau(Y - \alpha_0 - W'\theta_0 - n^{-1/2}W'\tau) - \rho_\tau(Y - \alpha_0 - W'\theta_0))(v(Y, X, Z) - \pi_0) + \\ r_n(\tau) &= nE_n[g(\tau, \hat{v})] - \frac{1}{2}\tau'E[\lambda_0(0)WW']\tau + \sqrt{n}\tau'E_n[\xi_0(\hat{v})W]. \end{aligned}$$

Then $\sup_\tau |r_n(\tau)| = o_p(1)$.

Proof. If $r_n(\tau) = o_p(1)$ pointwise in τ , it is possible to the convexity lemma in Pollard (1991) to $\tau \mapsto nE_n[g(\tau, \hat{v})] + \sqrt{n}\tau'E_n[\xi_0(\hat{v})W]$, and then $\sup_\tau |r_n(\tau)| = o_p(1)$. Therefore, it suffices to show that $r_n(\tau) = o_p(1)$ pointwise in τ . Note that

$$E[\lambda_0(0)WW'] = E[f_{Y|D,Z,X}(W'\theta_0)(1 - \pi_1 - \nu_0(Y, X, Z))_+ WW'] + E[f_{Y|D,Z,X}(\alpha_0 + W'\theta_0)(\nu_0(Y, X, Z) - \pi_0)_+ WW'],$$

where the equality uses the law of iterated expectation, $(1 - \pi_1 - \nu_0(Y, X, Z))_+ = 1 - \pi_1 - \nu_0(Y, X, Z)$ and $(\nu_0(Y, X, Z) - \pi_0)_+ = \nu_0(Y, X, Z) - \pi_0$. Therefore, $\frac{\partial}{\partial \tau} E[ng(\tau, \nu_0)] \Big|_{\tau=0} = -n^{1/2} E[\xi_0(\nu_0)W] = 0$ and $\frac{\partial^2}{\partial \tau \partial \tau'} E[ng(\tau, \nu_0)] = E[\lambda_0(\tau)WW'] = E[\lambda_0(0)WW'] + o(1)$, so that

$$E[ng(\tau, \nu_0)] = \frac{1}{2} \tau' E[\lambda_0(0)WW'] \tau + o(1).$$

Therefore,

$$\begin{aligned} r_n(\tau) &= nE_n[g(\tau, \hat{\nu})] - nE[g(\tau, \nu_0)] + \sqrt{n}\tau' E_n[\xi_0(\hat{\nu})W] + o(1). \\ &= E_n[n g(\tau, \hat{\nu}) + \sqrt{n}\tau' \xi_0(\hat{\nu})W] - E_n[n g(\tau, \nu_0) + \sqrt{n}\tau' \xi_0(\nu_0)W] \\ &\quad + E_n[n g(\tau, \nu_0) + \sqrt{n}\tau' \xi_0(\nu_0)W] - E[ng(\tau, \nu_0)] + o(1) \\ &= E_n[n(g(\tau, \hat{\nu}) - g(\tau, \nu_0)) + \sqrt{n}\tau'(\xi_0(\hat{\nu}) - \xi_0(\nu_0))W] \\ &\quad + (E_n - E)[ng(\tau, \nu_0) + \sqrt{n}\tau' \xi_0(\nu_0)W] + o(1), \end{aligned}$$

where the last equality follows from $E[\tau' \xi_0(\nu_0)W] = 0$.

First, $E_n[n(g(\tau, \hat{\nu}) - g(\tau, \nu_0)) + \sqrt{n}\tau'(\xi_0(\hat{\nu}) - \xi_0(\nu_0))W]$ converges to zero in L^1 . By the definitions of $g(\tau, \nu)$ and $\xi_0(\nu)$,

$$\begin{aligned} n(g(\tau, \hat{\nu}) - g(\tau, \nu_0)) + \sqrt{n}\tau'(\xi_0(\hat{\nu}) - \xi_0(\nu_0))W \\ = n \times \text{term}_1 \times ((1 - \pi_1 - \hat{\nu}(Y, X, Z))_+ - (1 - \pi_1 - \nu_0(Y, X, Z))) \\ + n \times \text{term}_2 \times ((\hat{\nu}(Y, X, Z) - \pi_0)_+ - (\nu_0(Y, X, Z) - \pi_0)), \end{aligned}$$

where $\text{term}_1 = \rho_\tau(Y - W'(\theta_0 - n^{-1/2}\tau)) - \rho_\tau(Y - W'\theta_0) + (\tau - 1\{Y - W'\theta_0 \leq 0\})n^{-1/2}W'\tau$ and $\text{term}_2 = \rho_\tau(Y - \alpha_0 - W'(\theta_0 - n^{-1/2}\tau)) - \rho_\tau(Y - \alpha_0 - W'\theta_0) + (\tau - 1\{Y - \alpha_0 - W'\theta_0 \leq 0\})n^{-1/2}W'\tau$. By the definition of ρ_τ , the two terms, term_1 and term_2 , can be bounded as follows:

$$\begin{aligned} |\text{term}_1| &\leq 1\{|Y - W'\theta_0| \leq |n^{-1/2}W'\tau|\} (|Y - W'\theta_0| + |n^{-1/2}W'\tau|) \\ &\leq 1\{|Y - W'\theta_0| \leq |n^{-1/2}W'\tau|\} \times 2|n^{-1/2}W'\tau| \end{aligned} \tag{B.3}$$

$$\begin{aligned} |\text{term}_2| &\leq 1\{|Y - \alpha_0 - W'\theta_0| \leq |n^{-1/2}W'\tau|\} (|Y - \alpha_0 - W'\theta_0| + |n^{-1/2}W'\tau|) \\ &\leq 1\{|Y - \alpha_0 - W'\theta_0| \leq |n^{-1/2}W'\tau|\} \times 2|n^{-1/2}W'\tau|. \end{aligned} \tag{B.4}$$

As long as $1 - \pi_1 - \hat{\nu}(Y, X, Z) \geq 0$ and $\hat{\nu}(Y, X, Z) - \pi_0 \geq 0$,

$$\begin{aligned} |(1 - \pi_1 - \hat{\nu}(Y, X, Z))_+ - (1 - \pi_1 - \nu_0(Y, X, Z))| &= |(1 - \pi_1 - \hat{\nu}(Y, X, Z)) \\ &\quad - (1 - \pi_1 - \nu_0(Y, X, Z))| \\ &\leq \sup_{(y,x,z)} |\hat{\nu}(y,x,z) - \nu_0(y,x,z)| \end{aligned} \tag{B.5}$$

$$\begin{aligned} |(\hat{\nu}(Y, X, Z) - \pi_0)_+ - (\nu_0(Y, X, Z) - \pi_0)| &= |(\hat{\nu}(Y, X, Z) - \pi_0) - (\nu_0(Y, X, Z) - \pi_0)| \\ &\leq \sup_{(y,x,z)} |\hat{\nu}(y,x,z) - \nu_0(y,x,z)|. \end{aligned} \tag{B.6}$$

Now it is possible to bound $E[\|E_n[n(g(\tau, \hat{v}) - g(\tau, v_0)) + \sqrt{n}\tau'(\xi_0(\hat{v}) - \xi_0(v_0))W]\|]$ as follows:

$$\begin{aligned} & E[\|E_n[n(g(\tau, \hat{v}) - g(\tau, v_0)) + \sqrt{n}\tau'(\xi_0(\hat{v}) - \xi_0(v_0))W]\|] \\ & \leq E[\|n(g(\tau, \hat{v}) - g(\tau, v_0)) + \sqrt{n}\tau'(\xi_0(\hat{v}) - \xi_0(v_0))\| \|W\|] \\ & \leq 2nE[1\{|Y - W'\theta_0| \leq |n^{-1/2}W'\tau|\}|n^{-1/2}W'\tau| \|W\|] \sup_{(y,x,z)} |\hat{v}(y,x,z) - v_0(y,x,z)| \\ & \quad + 2nE[1\{|Y - \alpha_0 - W'\theta_0| \leq |n^{-1/2}W'\tau|\}|n^{-1/2}W'\tau| \|W\|] \\ & \quad \sup_{(y,x,z)} |\hat{v}(y,x,z) - v_0(y,x,z)|. \end{aligned}$$

Since $|f_{Y,X,Z}| \leq C$, it follows that

$$\begin{aligned} E[1\{|Y - W'\theta_0| \leq |n^{-1/2}W'\tau|\}|n^{-1/2}W'\tau| \|W\|] & \leq 2Cn^{-1}E[|W'\tau|^2 \|W\|] \\ E[1\{|Y - \alpha_0 - W'\theta_0| \leq |n^{-1/2}W'\tau|\}|n^{-1/2}W'\tau| \|W\|] & \leq 2Cn^{-1}E[|W'\tau|^2 \|W\|], \end{aligned}$$

so that

$$\begin{aligned} & E[\|n(g(\tau, \hat{v}) - g(\tau, v_0)) + \sqrt{n}\tau'(\xi_0(\hat{v}) - \xi_0(v_0))\| \|W\|] \\ & \leq 8CE[|W'\tau|^2 \|W\|] \sup_{(y,x,z)} |\hat{v}(y,x,z) - v_0(y,x,z)| \\ & = o(1). \end{aligned}$$

Next, $(E_n - E)[ng(\tau, v_0) + \sqrt{n}\tau'\xi_0(v_0)W]$ converges to zero in L^2 . Using the bounds in equations (B.3)-(B.6),

$$\begin{aligned} & E[(ng(\tau, v_0) + \sqrt{n}\tau'\xi_0(v_0)W)^2]^{1/2} \\ & \leq nE[|\text{term}_1|^2 \times ((1 - \pi_1 - \hat{v}(Y, X, Z))_+ - (1 - \pi_1 - v_0(Y, X, Z)))^2]^{1/2} \\ & \quad + nE[|\text{term}_2|^2 \times ((\hat{v}(Y, X, Z) - \pi_0)_+ - (v_0(Y, X, Z) - \pi_0))^2]^{1/2} \\ & \leq 4nE[1\{|Y - W'\theta_0| \leq |n^{-1/2}W'\tau|\} \times |n^{-1/2}W'\tau|^2]^{1/2} \sup_{(y,x,z)} |\hat{v}(y,x,z) - v_0(y,x,z)| \\ & \quad + 4nE[1\{|Y - \alpha_0 - W'\theta_0| \leq |n^{-1/2}W'\tau|\} \times |n^{-1/2}W'\tau|^2]^{1/2} \\ & \quad \times \sup_{(y,x,z)} |\hat{v}(y,x,z) - v_0(y,x,z)| \\ & \leq 8n^{1/2}E[|W'\tau|^2]^{1/2} \sup_{(y,x,z)} |\hat{v}(y,x,z) - v_0(y,x,z)| \\ & = o(n^{1/2}), \end{aligned}$$

and then $E[((E_n - E)[ng(\tau, v_0) + \sqrt{n}\tau'\xi_0(v_0)W])^2] = n^{-1}E[(ng(\tau, v_0) + \sqrt{n}\tau'\xi_0(v_0)W)^2] = o(1)$. □

LEMMA B.6. $\sqrt{n}(\hat{\theta}_0 - \theta_0) = \sqrt{n}E[\lambda_0(0)WW']^{-1}E_n[s_0] + o_p(1)$ and therefore $\sqrt{n}(\hat{\theta}_0 - \theta_0) \rightarrow_d N(0, \Omega_0)$ as $n \rightarrow \infty$, where $\Omega_0 = E[\lambda_0(0)WW']^{-1}E[s_0s_0']E[\lambda_0(0)WW']^{-1}$.

Proof. Define $\eta_n = \sqrt{n}E[\lambda_0(0)WW']^{-1}E_n[s_0]$. It suffices to show $\sqrt{n}(\hat{\theta}_0 - \theta_0) = \eta_n + o_p(1)$. By Lemma B.4, $\eta_n = \sqrt{n}E[\lambda_0(0)WW']^{-1}E_n[\xi_0(\hat{v})W] + o_p(1)$. Using the definition of $r_n(\cdot)$,

$$\begin{aligned} -\frac{1}{2}\eta_n'E[\lambda_0(0)WW']\eta_n + r_n(\eta_n) &= nE_n[g(\eta_n, \hat{v})] \\ &\geq nE_n[g(\sqrt{n}(\hat{\theta}_0 - \theta_0), \hat{v})] \\ &= \frac{1}{2}(\sqrt{n}(\hat{\theta}_0 - \theta_0) - \eta_n)'E[\lambda_0(0)WW'](\sqrt{n}(\hat{\theta}_0 - \theta_0) - \eta_n) \\ &\quad - \frac{1}{2}\eta_n'E[\lambda_0(0)WW']\eta_n + r_n(\sqrt{n}(\hat{\theta}_0 - \theta_0)) \\ &\geq \frac{1}{2}\|\sqrt{n}(\hat{\theta}_0 - \theta_0) - \eta_n\|^2 \text{eig}_{\min}(E[\lambda_0(0)WW']) \\ &\quad - \frac{1}{2}\eta_n'E[\lambda_0(0)WW']\eta_n + r_n(\sqrt{n}(\hat{\theta}_0 - \theta_0)), \end{aligned}$$

where the first inequality uses $\sqrt{n}(\hat{\theta}_0 - \theta_0) = \arg \min_{\tau} E_n[g(\tau, \hat{v})]$ and $\text{eig}_{\min}(E[\lambda_0(0)WW'])$ is the minimum eigenvalue of $E[\lambda_0(0)WW']$. Therefore, $r_n(\eta_n) - r_n(\sqrt{n}(\hat{\theta}_0 - \theta_0)) \geq \frac{1}{2}\|\sqrt{n}(\hat{\theta}_0 - \theta_0) - \eta_n\|^2 \text{eig}_{\min}(E[\lambda_0(0)WW'])$, so that Lemma B.5 implies $\sqrt{n}(\hat{\theta}_0 - \theta_0) - \eta_n = o_p(1)$. □

LEMMA B.7. $Pr(|Y - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|) = o(1)$ and $Pr(|Y - \alpha_0 - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|) = o(1)$.

Proof. By Lemma B.6 and Assumption 12(iii), the first part of this lemma follows from $Pr(|Y - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|) \leq Pr(\|(\hat{\theta}_0 - \theta_0)\| \leq n^{-1/2} \log(n)) + Pr(|Y - W'\theta_0| \leq n^{-1/2} \log(n) \|W\|) = o(1)$. The second part can be shown similarly. □

LEMMA B.8. $E_n[\hat{s}(\alpha_0; \pi_0, \pi_1)\hat{s}(\alpha_0; \pi_0, \pi_1)'] = E[s_0s_0'] + o_p(1)$.

Proof. The weak law of large numbers implies $E_n[s_0s_0'] = E[s_0s_0'] + o_p(1)$, and then it suffices to show $E_n[\hat{s}(\alpha_0; \pi_0, \pi_1)\hat{s}(\alpha_0; \pi_0, \pi_1)' - s_0s_0'] = o_p(1)$. Since

$$\begin{aligned} &E_n[(\hat{s}(\alpha_0; \pi_0, \pi_1) - s_0)(\hat{s}(\alpha_0; \pi_0, \pi_1) - s_0)'] \\ &= E_n[(\hat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0))WW'(\hat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0))'] \\ &\quad + E_n[(\hat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0))W\hat{\psi}_\delta']E_n[W(\hat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))'] \\ &\quad + E_n[(\hat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0))W\hat{\psi}_\delta'](E_n - E)[W\Xi_0(\delta_0)]' \\ &\quad + E_n[(\hat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0))W(\hat{\psi}_\delta - \psi_\delta)']E[W\Xi_0(\delta_0)]' \\ &\quad + E_n[W(\hat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))]E_n[\hat{\psi}_\delta W'(\hat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0))'] \\ &\quad + E_n[W(\hat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))]E_n[\hat{\psi}_\delta \hat{\psi}_\delta']E_n[W(\hat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))'] \\ &\quad + E_n[W(\hat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))]E_n[\hat{\psi}_\delta \hat{\psi}_\delta'](E_n - E)[W\Xi_0(\delta_0)]' \\ &\quad + E_n[W(\hat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))]E_n[\hat{\psi}_\delta(\hat{\psi}_\delta - \psi_\delta)']E[W\Xi_0(\delta_0)]' \\ &\quad + (E_n - E)[W\Xi_0(\delta_0)]E_n[\hat{\psi}_\delta W'(\hat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0))'] \\ &\quad + (E_n - E)[W\Xi_0(\delta_0)]E_n[\hat{\psi}_\delta \hat{\psi}_\delta']E_n[W(\hat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))'] \\ &\quad + (E_n - E)[W\Xi_0(\delta_0)]E_n[\hat{\psi}_\delta \hat{\psi}_\delta'](E_n - E)[W\Xi_0(\delta_0)]' \end{aligned}$$

$$\begin{aligned}
 &+ (E_n - E) [W \Xi_0(\delta_0)] E_n[\widehat{\psi}_\delta(\widehat{\psi}_\delta - \psi_\delta)'] E [W \Xi_0(\delta_0)]' \\
 &+ E [W \Xi_0(\delta_0)] E_n[(\widehat{\psi}_\delta - \psi_\delta) W'] (\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0))' \\
 &+ E [W \Xi_0(\delta_0)] E_n[(\widehat{\psi}_\delta - \psi_\delta) \widehat{\psi}_\delta'] E_n [W(\widehat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))]' \\
 &+ E [W \Xi_0(\delta_0)] E_n[(\widehat{\psi}_\delta - \psi_\delta) \widehat{\psi}_\delta'] (E_n - E) [W \Xi_0(\delta_0)]' \\
 &+ E [W \Xi_0(\delta_0)] E_n[(\widehat{\psi}_\delta - \psi_\delta)(\widehat{\psi}_\delta - \psi_\delta)'] E [W \Xi_0(\delta_0)]'
 \end{aligned}$$

and

$$\begin{aligned}
 &E_n[(\widehat{s}(\alpha_0; \pi_0, \pi_1) - s_0) s_0'] \\
 &= E_n[(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W W' \xi_0(\nu_0)'] \\
 &\quad + E_n[(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W \psi_\delta'] E [W \Xi_0(\delta_0)]' \\
 &\quad + E_n [W(\widehat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))] E_n[\widehat{\psi}_\delta W' \xi_0(\nu_0)'] \\
 &\quad + E_n [W(\widehat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))] E_n[\widehat{\psi}_\delta \psi_\delta'] E [W \Xi_0(\delta_0)]' \\
 &\quad + (E_n - E) [W \Xi_0(\delta_0)] E_n[\widehat{\psi}_\delta W' \xi_0(\nu_0)'] + (E_n - E) [W \Xi_0(\delta_0)] \\
 &\quad \quad E_n[\widehat{\psi}_\delta \psi_\delta'] E [W \Xi_0(\delta_0)]' + E [W \Xi_0(\delta_0)] \\
 &\quad E_n[(\widehat{\psi}_\delta - \psi_\delta) W' \xi_0(\nu_0)'] \\
 &\quad + E [W \Xi_0(\delta_0)] E_n[(\widehat{\psi}_\delta - \psi_\delta) \psi_\delta'] E [W \Xi_0(\delta_0)]',
 \end{aligned}$$

it suffices to show that $(E_n - E) [W \Xi_0(\delta_0)] = o_p(1)$, that $E_n[\widehat{\psi}_\delta W' \xi_0(\nu_0)']$, $E_n[\widehat{\psi}_\delta \psi_\delta']$, $E_n[\widehat{\psi}_\delta W' \xi_0(\nu_0)']$, and $E_n[\widehat{\psi}_\delta \psi_\delta']$ are $O_p(1)$, and that the sample averages of the following variables are $o_p(1)$: (1) $(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W \psi_\delta'$, (2) $(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W W' \xi_0(\nu_0)'$, (3) $(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W W' (\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0))'$, (4) $(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W \widehat{\psi}_\delta'$, (5) $(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W (\widehat{\psi}_\delta - \psi_\delta)'$, (6) $W(\widehat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))$, (7) $\widehat{\psi}_\delta W' (\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0))'$, (8) $\widehat{\psi}_\delta (\widehat{\psi}_\delta - \psi_\delta)'$, (9) $(\widehat{\psi}_\delta - \psi_\delta) W' \xi_0(\nu_0)'$, (10) $(\widehat{\psi}_\delta - \psi_\delta) W' (\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0))'$, and (11) $(\widehat{\psi}_\delta - \psi_\delta)(\widehat{\psi}_\delta - \psi_\delta)'$. The convergence $(E_n - E) [W \Xi_0(\delta_0)] = o_p(1)$ comes from the weak law of large numbers, and the other parts come from a combination of Assumption 10(iv)–(vii) and the following equalities:

$$E[\|\xi_0(\nu_0) W\|^2] < \infty \tag{B.7}$$

$$E[\|(\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)) W\|^2] = o(1) \tag{B.8}$$

$$E [W(\widehat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))] = o_p(1). \tag{B.9}$$

Note that equation (B.7) follows from $|\xi_0(\nu_0)| \leq 2$ and $E[\|W\|^2] < \infty$.

First, equation (B.8) is shown as follows. Since

$$\begin{aligned}
 |\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(\nu_0)| &\leq |1\{Y - W' \theta_0 \leq 0\} - 1\{Y - W' \hat{\theta}_0 \leq 0\}| \times |1 - \pi_1 - \hat{\nu}(Y, X, Z)| \\
 &\quad + |\tau - 1\{Y - W' \theta_0 \leq 0\}| \times |\nu_0(Y, X, Z) - \hat{\nu}(Y, X, Z)| \\
 &\quad + |1\{Y - \alpha_0 - W' \theta_0 \leq 0\} - 1\{Y - \alpha_0 - W' \hat{\theta}_0 \leq 0\}| \\
 &\quad \times |\hat{\nu}(Y, X, Z) - \pi_0| \\
 &\quad + |\tau - 1\{Y - \alpha_0 - W' \theta_0 \leq 0\}| \times |\hat{\nu}(Y, X, Z) - \nu_0(Y, X, Z)| \\
 &\leq 1\{Y - W' \theta_0 \leq 0\} \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|
 \end{aligned}$$

$$\begin{aligned}
 &+ 1\{|Y - \alpha_0 - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|\} \\
 &+ 2|\hat{v}(Y, X, Z) - v_0(Y, X, Z)|,
 \end{aligned}$$

it follows that

$$\begin{aligned}
 E[\|\widehat{\xi}(\alpha_0; \pi_0, \pi_1) - \xi_0(v_0)\|W\|^2]^{1/2} \\
 \leq Pr(|Y - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|\})^{1/4} E[\|W\|^4]^{1/4} \\
 + Pr(|Y - \alpha_0 - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|\})^{1/4} E[\|W\|^4]^{1/4} \\
 + 2E[|\hat{v}(Y, X, Z) - v_0(Y, X, Z)|^4]^{1/4} E[\|W\|^4]^{1/4}.
 \end{aligned}$$

By Assumption 10(ii) and Lemma B.7, equation (B.8) holds.

Next, equation (B.9) is shown as follows. Since

$$\begin{aligned}
 \widehat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0) &= 1\{Y - W'\hat{\theta}_0 \leq 0\} \left(\frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) - \frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) \right) \\
 &+ \left(1\{Y - W'\hat{\theta}_0 \leq 0\} - 1\{Y - W'\theta_0 \leq 0\} \right) \frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) \\
 &- 1\{Y - \alpha - W'\hat{\theta}_0 \leq 0\} \left(\frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) - \frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) \right) \\
 &- \left(1\{Y - \alpha - W'\hat{\theta}_0 \leq 0\} - 1\{Y - \alpha - W'\theta_0 \leq 0\} \right) \frac{\partial}{\partial \delta'} v_\delta(Y, X, Z),
 \end{aligned}$$

it follows that

$$\begin{aligned}
 E[\|W(\widehat{\Xi}(\alpha_0; \pi_0, \pi_1) - \Xi_0(\delta_0))\|] \\
 \leq 2E \left[\left\| W \left(\frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) - \frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) \right) \right\| \right] \\
 + Pr(|Y - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|\})^{1/2} E \left[\left\| W \frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) \right\|^2 \right]^{1/2} \\
 + Pr(|Y - \alpha_0 - W'\theta_0| \leq \|(\hat{\theta}_0 - \theta_0)\| \|W\|\})^{1/2} E \left[\left\| W \frac{\partial}{\partial \delta'} v_\delta(Y, X, Z) \right\|^2 \right]^{1/2}.
 \end{aligned}$$

By Lemma B.7 and Assumption 10(vi) and (vii), equation (B.9) holds. □

LEMMA B.9. $E_n[\widehat{\lambda}(\alpha_0; \pi_0, \pi_1) WW'] = E[\lambda_0(0) WW'] + o_p(1)$.

Proof. Since

$$\begin{aligned}
 E[\|\lambda_0(0) WW'\|^2] &\leq E[\|f_{Y|D,Z,X}(W'\theta_0)(1 - \pi_1 - D) \\
 &\quad + f_{Y|D,Z,X}(\alpha_0 + W'\theta_0)(D - \pi_0)\|^2 \cdot \|W\|^4] \\
 &\leq E[(\|f_{Y|D,Z,X}(W'\theta_0)\| + \|f_{Y|D,Z,X}(\alpha_0 + W'\theta_0)\|)^2 \cdot \|W\|^4] \\
 &\leq 4C^2 E[\|W\|^4] \\
 &< \infty,
 \end{aligned}$$

the weak law of large numbers implies $E_n[\lambda_0(0)WW'] = E[\lambda_0(0)WW'] + o_p(1)$. It suffices to show that $E_n[\hat{\lambda}(\alpha_0; \pi_0, \pi_1)WW'] = E_n[\lambda_0(0)WW'] + o_p(1)$. Using the mean value expansion,

$$\begin{aligned} \hat{\lambda}(\alpha_0; \pi_0, \pi_1) - \lambda_0(0) &= \left(K_h^{(1)}(\tilde{V})(1 - \pi_1 - D) + K_h^{(1)}(\tilde{V} - \alpha_0)(D - \pi_0) \right) W'(\hat{\theta}_0 - \theta_0) \\ &\quad + \left(K_h(Y - W'\theta_0) - f_{Y|D,Z,X}(W'\theta_0) \right) (1 - \pi_1 - D) \\ &\quad + \left(K_h(Y - \alpha_0 - W'\theta_0) - f_{Y|D,Z,X}(\alpha_0 + W'\theta_0) \right) (D - \pi_0), \end{aligned}$$

where $K_h(t) = K(t/h)/h$, and \tilde{V} is a value between $Y - W'\hat{\theta}_0$ and $Y - W'\theta_0$. Using the above mean value expansion,

$$\begin{aligned} &\|E_n[\hat{\lambda}(\alpha_0; \pi_0, \pi_1)WW'] - E_n[\lambda_0(0)WW']\| \\ &= \|E_n[(\hat{\lambda}(\alpha_0; \pi_0, \pi_1) - \lambda_0(0))WW']\| \\ &\leq \|E_n\left[\left(K_h^{(1)}(\tilde{V})(1 - \pi_1 - D) + K_h^{(1)}(\tilde{V} - \alpha_0)(D - \pi_0)\right)W'(\hat{\theta}_0 - \theta_0)WW'\right]\| \\ &\quad + \|E_n[(K_h(Y - W'\theta_0) - f_{Y|D,Z,X}(W'\theta_0))(1 - \pi_1 - D)WW']\| \\ &\quad + \|E_n[(K_h(Y - \alpha_0 - W'\theta_0) - f_{Y|D,Z,X}(\alpha_0 + W'\theta_0))(D - \pi_0)WW']\| \\ &\leq 2\sup_v |K_h^{(1)}(v)| \cdot \|\hat{\theta}_0 - \theta_0\| \cdot E_n[\|W\|^3] \\ &\quad + \|E_n[(K_h(Y - W'\theta_0) - f_{Y|D,Z,X}(W'\theta_0))(1 - \pi_1 - D)WW']\| \\ &\quad + \|E_n[(K_h(Y - \alpha_0 - W'\theta_0) - f_{Y|D,Z,X}(\alpha_0 + W'\theta_0))(D - \pi_0)WW']\| \\ &\leq O_p(n^{-1/2}) + \|E_n[(K_h(Y - W'\theta_0) - f_{Y|D,Z,X}(W'\theta_0))(1 - \pi_1 - D)WW']\| \\ &\quad + \|E_n[(K_h(Y - \alpha_0 - W'\theta_0) - f_{Y|D,Z,X}(\alpha_0 + W'\theta_0))(D - \pi_0)WW']\|. \end{aligned}$$

Since each entry in WW' has a finite variance, it suffices to show that

$$E_n[(K_h(Y - W'\theta_0) - f_{Y|D,Z,X}(W'\theta_0))(1 - \pi_1 - D)\omega] = o_p(1) \tag{B.10}$$

$$E_n[(K_h(Y - \alpha_0 - W'\theta_0) - f_{Y|D,Z,X}(\alpha_0 + W'\theta_0))(D - \pi_0)\omega] = o_p(1) \tag{B.11}$$

for a random variable ω such that ω is a function of (D, Z, X) and $E[\omega^2] < \infty$. The rest of the proof is going to focus on (B.10) because the proof for (B.11) is similar. The mean of the left-hand side of (B.10) is $O(h)$, because

$$\begin{aligned} |E[(K_h(Y - W'\theta_0) - f_{Y|D,Z,X}(W'\theta_0))\omega]| &\leq E\left[\int |K(v)| |f_{Y|X,Z,D}(W'\theta_0 + vh) \right. \\ &\quad \left. - f_{Y|D,Z,X}(W'\theta_0)| dv |\omega|\right] \\ &\leq ChE[|\omega|] \int |K(v)|v dv \\ &= O(h). \end{aligned}$$

The variance of the left-hand side of (B.10) is $O((nh)^{-1})$, because

$$\begin{aligned} &E[((K_h(Y - W'\theta_0) - f_{Y|D,Z,X}(W'\theta_0))\omega)^2] \\ &\leq E\left[\int K_h(y - W'\theta_0)^2 f_{Y|X,Z,D}(y) dy \omega^2\right] + E[f_{Y|D,Z,X}(W'\theta_0)^2 \omega^2] \end{aligned}$$

$$\begin{aligned}
 & -2E\left[\int K_h(y - W'\theta_0)f_{Y|X,Z,D}(y)dyf_{Y|D,Z,X}(W'\theta_0)\omega\right] \\
 & \leq Ch^{-1}\int K(v)^2dvE[\omega^2] + C^2E[\omega^2] + 2C^2E[|\omega|] \\
 & = O(h^{-1}).
 \end{aligned}$$

□

LEMMA B.10. $\hat{\Omega}(\alpha_0; \pi_0, \pi_1) = \Omega_0 + o_p(1)$.

Proof. Since $E[\lambda_0(0)WW'] = E[f_{Y-\alpha_0D^*|Z,X}(W'\theta_0)WW']$ is invertible, the statement of this lemma follows from Lemmas B.8 and B.9. □